IRE Transactions

on AUDIO

Volume AU-6 SEPTEMBER-OCTOBER, 1958 Number 5

Published Bi-Monthly

TABLE OF CONTENTS

PGA NEWS

PGA Chapter Statistics 91
Tapescripts Available 91
Announcements 92
Recent Chapter Meetings 93
Chapter News 94

FEATURES

With Other Acoustical and Audio Societies . . . . . *Benjamin B. Bauer* 95

CONTRIBUTIONS

Research and Development on the Piano . . . . . *J. P. Quitter* 96
A Survey of Speech Bandwidth Compression Techniques . . . . . *S. J. Campanella* 104

Contributors 116

PUBLISHED BY THE

Professional Group on Audio

World Radio History

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.

Administrative Committee for 1958-1959

F. H. Slaymaker, *Chairman*
Stromberg-Carlson Co., Rochester 21, N.Y.

M. S. Corrington, *Vice Chairman*
RCA Victor Division
Cherry Hill, Camden 8, N.J.

B. B. Bauer, *Secretary-Treasurer*
CBS Laboratories
Stamford, Conn.

S. J. Begun
Clevite-Brush Development
Cleveland 80, Ohio

J. K. Hilliard
Altec Lansing Corporation
Beverly Hills, Calif.

A. B. Bereskin
University of Cincinnati
Cincinnati 21, Ohio

H. F. Olson
RCA Laboratories
Princeton, N.J.

M. Camras
Armour Research Foundation
Chicago 16, Ill.

W. T. Selsted
Ampex Corp.
Redwood City, Calif.

P. B. Williams
Jensen Manufacturing Co.
Chicago, Ill.

IRE TRANSACTIONS® on AUDIO

Published by the Institute of Radio Engineers, Inc., for the Professional Group on Audio at 1 East 79th Street, New York 21, New York. Responsibility for the contents rests upon the authors, and not upon the IRE, the Group, or its members. Individual copies available for sale to IRE-PGA members at $0.65; to IRE members at $0.95, and to nonmembers at $1.95.

Editorial Committee

Marvin Camras, *Editor*
Armour Research Foundation, Chicago 16, Ill.

B. B. Bauer
CBS Laboratories
Stamford, Conn.

D. W. Martin
The Baldwin Piano Co.
1801 Gilbert Ave., Cincinnati 2, Ohio

A. A. Bereskin
University of Cincinnati
Cincinnati 21, Ohio

J. R. Macdonald
Texas Instruments, Inc.
6000 Lemmon Ave., Dallas 9, Texas

P. B. Williams
Jensen Manufacturing Co.
6601 S. Laramie Ave., Chicago 38, Ill.

Copyright © 1958—The Institute of Radio Engineers, Inc.
Printed in U.S.A.

All rights, including translations, are reserved by the IRE. Requests for republication privileges should be addressed to the Institute of Radio Engineers, Inc., 1 E. 79th St., New York 21, New York.

World Radio History

PGA News

PGA CHAPTER STATISTICS

Official Chapters

The following were official Chapters of IRE-PGA as of December 31, 1957. The number of paid members in each of these 17 chapters is indicated.

Albuquerque-Los Alamos	34	Houston	48
Baltimore	73	Milwaukee	58
Boston	252	Philadelphia	253
Chicago	233	Phoenix	34
Cincinnati	41	San Antonio	24
Cleveland	61	San Diego	55
Dayton	49	San Francisco	206
Hawaii	16	Syracuse	71
Washington, D. C.	157		

Chapter Officers

Where current officers of 1958–1959 have not been reported, the 1957–1958 officers are listed. Please notify IRE headquarters, and the Editor of any changes. Addresses may be found in the 1958 IRE DIRECTORY.

Albuquerque-Los Alamos 1958–1959	*Chairman:* Gordon R. Bachand *Vice-Chairman and Secretary:* Charles W. Remaley, Jr.
Baltimore 1958–1959	*Chairman:* John L. Makwalter *Vice-Chairman:* Wilbur Visher *Secretary:* George N. Webb
Boston 1958–1959	*Chairman:* Daniel Von Recklinghausen *Vice-Chairman:* Donald J. Fritch *Secretary-Treasurer:* H. S. Littleboy
Chicago 1958–1959	*Chairman:* Robert J. Larsen *Vice-Chairman:* Leonard G. Eckmann *Secretary:* William M. Idhe
Cincinnati 1958–1959	*Chairman:* William C. Wayne, Jr. *Vice-Chairman:* Clyde G. Haehnle *Secretary-Treasurer:* John P. Quitter
Cleveland 1958–1959	*Chairman:* Kenneth Hamann
Dayton 1958–1959	*Chairman:* J. S. Stanton *Vice-Chairman:* E. O. Valentine *Secretary:* R. W. Stanford
Hawaii 1958–1959	*Chairman:* D. H. DaShiell *Vice-Chairman:* Iwao Miyake *Secretary-Treasurer:* Daniel L. Pang
Houston	*Chairman:* W. C. Wrye, Jr.
Milwaukee 1958–1959	*Chairman:* Tim Houle
Philadelphia 1958–1959	*Chairman:* Robert W. Roop *Vice-Chairman:* R. M. Carrell *Secretary:* T. A. Benham
Phoenix 1957–1958	*Chairman:* Andrew B. Jacobsen
San Antonio 1958–1959	*Chairman:* Bill Case
San Diego 1958–1959	*Chairman:* Kenneth R. O'Neal *Vice-Chairman and Secretary:* Loman Tibbals
San Francisco 1958–1959	*Chairman:* Don L. Broderick *Vice-Chairman:* Lambert Dolphin *Secretary-Treasurer:* Lawrence W. Johnson
Syracuse 1958–1959	*Chairman:* William R. Chynoweth *Vice-Chairman:* Archie A. McGee, Jr. *Secretary-Treasurer:* G. Grenier
Washington, D. C. 1958–1959	*Chairman:* William B. Bernard *Vice-Chairman:* H. Peter Meissinger *Secretary:* Sachio Saito

TAPESCRIPTS AVAILABLE

Chairmen of PGA Chapters and Chairmen of IRE Sections will be interested in tapescripts either as primary or supplementary program material.

Tapescripts are tape-recorded talks with accompanying slides. They are loaned by the IRE-PGA Tapescripts Committee, and are available

Andrew B. Jacobsen, *Chairman*,
Motorola, Inc.,
8201 E. McDowell Road,
Phoenix, Ariz.

The only cost is the return postage on the material. It is very important that the program chairman request a tapescript in advance so he may review the material and prepare his program.

The best way to utilize tapescripts for local presentation is for someone to review the material and prepare a short discussion of the subject. This person should be prepared and qualified to answer questions that may arise. It is of the utmost importance that those presenting a tapescript scan the material from a purely mechanical standpoint to be sure they have the copies expected, and the technical equipment to reproduce the sound and picture.

Technical standards for tapescripts are as follows: sound on $7\frac{1}{2}$ inches per second, $\frac{1}{4}$-inch tape, full track on 7-inch reels. The slides are 2×2-inch card board mounts, double 35-mm slides. In a few cases where a limited number of copies are available, $3\frac{1}{4}\times 4$-inch slides are used.

The following tapescripts are available.

"Phonograph Reproduction," B. B. Bauer, Shure Brothers, Inc., Chicago, Ill. Grooves and needles, fidelity and efficiency, pickup arms, and recording-reproducing characteristics are some of the subjects discussed in this one-hour recorded paper.

"An Experiment Co-Channel Television Booster Station Using Crossed Polarization," John H. DeWitt, Jr., WSM, Nashville, Tenn. The special receiving and transmitting antennas and equipment developed to improve fringe and reception by co-channel booster methods are discussed. A report on the operation of an experimental installation at Lawrenceburg, Tenn., is given.

"Method for Time or Frequency Compression-Expansion of Speech," G. Fairbanks, W. L. Everitt, and R. P. Jaeger, University of Illinois.

"Magnetic Recording," Marvin Camras, Armour Research Foundation of Illinois Institute of Technology. Discusses fundamentals of wire and tapes; heads; bias; equalization; present problems; and future developments. Thirty minutes.

"Push Pull Single Ended Audio Amplifier," A. Peterson and D. B. Sinclair, General Radio Company, Cam-

World Radio History

bridge, Mass., A convention paper presented by Dr. Peterson. Twenty-five minutes. $3\frac{1}{4} \times 4$-inch slides.

"The Electrostatic Loudspeaker—An Objective Evaluation," R. J. Larson, Jensen Manufacturing Co., Chicago, Ill. The condenser loudspeaker, following development of new materials and methods, is now practical for high-frequency use in multichannel loudspeaker systems. Theoretical and mechanical design considerations illustrate the limitations at this stage of the art, including inherently high distortion at high output levels, and inability to withstand overloads. Principal advantages are low cost and efficient reproduction at extremely high frequencies. Demonstration will illustrate the operation of a typical electrostatic "tweeter" in combination with lower-frequency dynamic loudspeaker channels.

"Efficiency and Power Rating of Loudspeakers," R. W. Benson, Armour Research Foundation, Chicago, Ill. The specification of the performance of loudspeakers is a subject of international controversy at the present time. Several groups in this country are concerned with standardized methods of evaluating the performance of loudspeakers in order to have a basis for comparison. Several European countries are also concerned with similar problems.

The measurement of the response-frequency characteristic and directional characteristics of loudspeakers are routinely performed by various laboratories. Sufficient agreement can be attained by the various laboratories to standardize this measure of performance. It is important to have a measure of the efficiency, that is sound power output vs electrical power input, and also a method of specifying the power-handling capacity of a loudspeaker. Various methods are in use at the present time to indicate the characteristics of a loudspeaker concerning these two measures. The method of using a reverberation chamber in integrate acoustic power output and thus determine the efficiency is discussed in comparison to the more tedious method of analytical integration of measurements performed in free space. Power-handling capabilities of loudspeakers as determined by distortion measurements and mechanical or electrical failure are given. A summary is included of the various methods of specifying both efficiency and power-handling capacity as used in various laboratories in both the United States and Europe.

"Sound Survey Meter," Arnold Peterson, General Radio Company, Cambridge, Mass., A convention paper. Twenty minutes. $3\frac{1}{4} \times 4$-inch slides.

"Microphones for High Intensity and High Frequencies," John K. Hilliard, Altec Lansing Company. A convention paper. Twenty minutes. $3\frac{1}{4} \times 4$-inch slides.

"The Ideo-Synchronizer," J. M. Henry and E. R. Moore, Boston Bell. A humorous satire on technical writing, specifications, and engineering. Good for mixed audience. Twelve minutes.

"An Improved Optical Method for Calibrating Test Records," B. B. Bauer, Shure Brothers, Inc., Chi., Ill.

"Electronically-Controlled Audio Filters," L. O. Dolansky, Northeastern University, Boston, Mass.

"Energy Distribution in Music," J. P. Overlay, Radio Manufacturing Engineers, Inc., Peoria, Ill. A knowledge of the manner in which the acoustic power encountered in music varies with respect to frequency can be a useful tool in the decision of components to be used in audio reinforcement or reproduction systems. The author believes that energy distribution information available up to this time is not fully applicable for such a use because it concerns primarily the average energy distribution. This paper deals with the amplitude of fractional-second energy peaks, without reference to the rate of their occurrence. It is these peaks which must be considered when distortion is of primary consideration; average power is useful only in predicting temperature rise (where applicable) of signal-handling components. Throughout the discussion, emphasis is placed upon the difference between average and peak energy considerations.

The source material from which the distribution analysis is drawn consisted of recent commercial vinyl recordings played on a carefully equalized reproducing system. Ten various types of music are classified and a distribution curve for each is drawn. The methods used in arriving at a typical curve are shown by breaking the spectrum into octaves with a bandpass filter.

This distribution information mentioned above is applied to the design of a three-channel loudspeaker system as an example of use. Other possible applications are mentioned.

"Bells, Electronic Carillons and Chimes," F. H. Slaymaker, Stromberg-Carlson Co., Rochester, N. Y. Bells and chimes, unlike the more familiar string and wind instruments, produce tones in which the overtone structure cannot be expressed as a series of harmonics.

The accuracy of tuning of the various overtones varies widely but the better cast bell carillons, electronic carillons, and tubular chimes do have very accurately tuned overtones. The results of measurements on cast bells, electronic carillons, and tubular chimes are presented. The individual overtones of the three types of instruments will be demonstrated on a loudspeaker and data are given on relative amplitude and decay rates of the various overtones. The reaction of "out of tuneness" is discussed and explained. A new type of tone source for electronic carillons is described and demonstrated. With this new tone source, a rod of carefully controlled rectangular cross section is used, which can give in a single unified structure essentially the same overtones array as that of an accurately tuned cast carillon bell. The new rectangular bar tone source is demonstrated.

ANNOUNCEMENTS

A session on "Stereophonic Disk Recording and Audio" was held during the IRE Canadian Convention

World Radio History

and Exposition on October 9, 1958, in the Satellite Room of the Automotive Building, Exhibition Park, Toronto, Canada. Papers were as follows.

"Tracing Distortion in Stereophonic Disc Recording," M. S. Corrington and T. Marakami, RCA Victor Co. Ltd., Camden, N. J.

"Certain Problems and Solutions in the Recording and Reproduction of Stereophonic Discs," B. B. Bauer, Columbia Broadcasting System Laboratories, N. Y.

"Design Consideration in the Use of the Stereophonic Disc Cartridge," Howard Durbin, Electrovoice, Buchanan, Mich.

"A Double Cardioid Stereophonic Microphone," A. Jamroz and G. B. Thompson, Northern Electric Co., Belleville, Ont., Canada.

RECENT CHAPTER MEETINGS

Albuquerque-Los Alamos

January 16—A stereophonic demonstration was held to compare stereo with monaural and large and small speakers used stereophonically.

March 7—"Latest Developments in Stereophonic Recording (Demonstration of Stereo Equipment)," John McAllister, Viking.

Baltimore

January 16—"Recording and Reproducing Stereophonic Sound on Single Groove Disks," R. E. Warn, Westrex Corp.

April 17—"Transistorized Audio Amplifier," John Tewksbury, Bendix Radio.

"Design of a Horn-Loaded Speaker Enclosure," John Markwalter, Aircraft Armaments.

May 22—"The Dynamic Loudspeaker—Can We Improve It?" S. J. White, Racon Electric Co.

Boston

January 16—"The Ultrasonic 'Radar' of the Bat," D. R. Griffin, Harvard University.

March 20—"Multi-Channel Recording Demonstration," H. J. Bresson, Ampex Corp.

May 15—"Westrex Stereodisk," E. A. Dickinson, Westrex Corp.

"Practical Stereophonic Phonograph Records," R. E. Warn and E. A. Dickinson, Westrex Corp.

Chicago

March 14—"A New Stereo-Disk Pickup," J. F. Wood, Electro-Voice, Inc.

May 21—"RCA Compatible Color Television System," including tour of NBC color TV studios, W. C. Prather, WNBQ.

Cincinnati

January 21—"Development of Stereophonic Sound," J. Drews and B. C. Carr, Ampex Audio, Inc.

February 18—"Sources of Interference Inherent in Vehicular Electrical Systems," Brooks H. Short, Delco-Remy, Inc.

June 24—Inspection trip and demonstration of WLW's new Rockwell Cathanode Transmitter and Crosley's Voice of America Facilities. R. J. Rockwell, Crosley Broadcasting Division of Avco, Cincinnati, Ohio.

Cleveland

May 15—"Qualitative Discussion on the Present Day Techniques Used in Obtaining Low-Frequency Sound Reproduction in Speakers," Joseph Chernof, Jr., Bell Aircraft Corp.

Dayton

February 6—"A New Very High Frequency Driver Design," Howard Souther, Electro-Voice, Inc.

March 6—"Your Telephone 1965," R. T. Riefenstahl, Ohio Bell Telephone Co.

April 3—"A Review of Current High Intensity Sound System Developments," Richard E. Liebich, Stromberg-Carlson.

Hawaii

January 8—"A Practical Approach to Selecting a Loudspeaker Enclosure," Daniel Pang, U. S. Naval Shipyard.

Houston

February 26—"Some Recent Developments in Turntables and Tone Arms," W. C. Wrye, Wrye Co.

Milwaukee

March 18—"Magnetic Tape in Radio, TV, and Master Recording," Leonard Hase, Ampex Corp.

April 22—"High Fidelity Transistorized Audio Amplifier," Richard C. Janzow, RCA.

Philadelphia

January 15—"The Westrex Sterodisk System," R. E. Warn and E. A. Dickinson, Westrex Corp.

February 12—"Models of Auditory Processes," Dr. J. C. R. Licklider, Bolt, Beranek and Newman.

March 13—"Electrostatic Loudspeakers," Arthur A. Janszen, Janszen Labs.

April 9—"Pickups for Reproduction from Stereo Records," Paul Weathers, Weathers Industries.

World Radio History

May 14—"Stereo-Panoramic Sound Reproduction," John Volkmann, RCA.

San Francisco

January 14—"The Westrex Stereodisk System," Dr. John G. Frayne, Westrex Corp.

March 11—"Panel Discussion: Musical Reproduction-Viewpoints" Ross Snyder, Moderator, Ampex Corp.; Bernard N. Oliver, Hewlett-Packard; Roy A. Long, Stanford Research Institute; Leonard Ratner, Stanford University; Leonard Tressel, San Francisco Symphony and Ampex Corp.

Syracuse

April 10—"A Terminated Horn Enclosure, Electrostatic Speaker, Lightweight Phonograph Cartridge," Dr. W. E. Glenn, G. E. Res. Lab.

Washington

April 1—"Design Characteristics of a New FM Tuner," Gordon Gow, McIntosh Labs.

June 2—"Optimum Balance Between Performance and Costs of High Fidelity Components," F. Montgomery, P. Meisinger, F. McIntosh, J. McDonald, I. Levine, and A. Preisman.

CHAPTER NEWS

Chicago, Ill.

The main paper of the Chicago Section IRE meeting program of October 10, 1958, was sponsored by the Chicago PGA Chapter. It was entitled "Certain Problems and Solutions in Recording and Reproduction of Stereophonic Disks," by Benjamin B. Bauer, CBS Laboratories, Stanford, Conn. According to *Scanfax* of October, 1958:

> Stereo already has become an important entertainment medium; it is big business. In its present commercial stage, however, stereo is not realizing all the potential for acoustic realism inherent in the basic idea. Ben Bauer, Vice-President of CBS Laboratories, will tell what some of the remaining problems are and how to overcome them.
>
> We have not seen much of Ben since he moved to New York, but many of us remember him as one of the men that made the Chicago Section tick back in the 40's. Those of us who have heard him speak know what an excellent background he has on audio and acoustic subjects, and how clearly he can put across a complex idea.
>
> Ben is bringing out special reproducing equipment from his laboratory and will prove his points with audible demonstrations.

During his talk, Mr. Bauer pointed out that the new stereophonic disk recordings pose new problems to the recording engineer and playback equipment designer. He discussed some of the aspects of stereophonic recording, including the importance of adequate recorded level and the reduction of turntable rumble and noise. He also described methods for the preservation of perspective on disk and in playback systems.

World Radio History

With Other Acoustical and Audio Societies

BENJAMIN B. BAUER†

NOTEWORTHY in the May, 1958, issue of the *Journal of the Acoustical Society of America* is a collection of papers presented during a Symposium of Unsolved Problems in Acoustics at the 54th meeting of the Society in Ann Arbor, Mich. In the "Editor's Foreword," Dr. R. Bruce Lindsey of Brown University aptly states: "As an active branch of physics and as a subject fundamental to modern engineering, acoustics poses many unsolved problems of fundamental character. . . . Through the fertile interplay of theory and application, acoustics is in a particularly strategic position to suggest problems whose solution will cast new light on both the properties of matter and the interaction of sound and man."

A Part of a session concerning electroacoustics and transducers was handled by F. V. Hunt of Harvard University. Many of the problems dealing with sound radiation, practical boundry conditions, and reciprocity of transducers still challenge the theorist. Low efficiency of electroacoustic transduction for wide bands of frequency remains a major problem for the inventor and the innovator. An objective specification of what would constitute good performance and good acoustics in listening rooms is still lacking. Contemporary interest centers on the large-signal nonlinear acoustics, magnetohydrodynamics, and transducers for microwave acoustics. Dr. Hunt concludes that "the major unsolved problem in acoustics, as in other sciences, is still that of learning how to frame the right questions to address to nature." Speaking on sonic engineering, T. F. Hueter of Raytheon Manufacturing Company concentrated on those problems which stem from inadequacy of materials and performance of apparatus. There is the need, for example, for piezoelectric materials which are good heat conductors and at the same time good electrical insulators. Another need is for materials which are good pressure transducers and at the same time are resistant to hydrostatic pressure. A still further area of need in constructing transducers is the lack of cements which are strong but not brittle, cold curing waterproofing materials, low-frequency absorbing materials for anechoic test tanks, lightweight diaphragms impervious to cavitation damage, and permanent magnets of high coercivity coupled with high reversible permeability.

† CBS Labs., Stanford, Conn.

Among more basic problems is the understanding of the processes that promote cavity formation and that accompany cavity collapse. The manner in which the processing efficiency is related to cavitation, as a function of temperature, hydrostatic pressure, and viscosity, is still unknown. The role of the gas content, dust particles, and ion tracks in promoting nucleation still is not well understood. Another basic problem is that of generation of sonic power. Hydrodynamic, pneumatic, and spark discharge methods for generation of high energy require additional study and development. In the area of utilization, the principal need at present is for transducer materials with high coupling coefficient and low losses, which will provide satisfactory impedance match to the medium.

The subject of noise, shock and vibration was discussed by R. O. Fehr of the General Electric Company, Schenectady, N. Y. He dealt more nearly with technology of problems very close to us. Among them is the problem of evaluating the effect of a jet engine exhaust on the fuselage of an airplane. The vibration of the panels of the fuselage must not exceed those values which are determined by the fatigue strength of the material. In addition, the vibration of the panels is transmitted into the fuselage increasing the noise level inside the airplane. The question of the economical reduction of this vibration and noise is still in need of an answer. Additional work is required to develop the technique of measurement of impact noises and determining their sources.

Dr. Fehr stressed the importance of developing equipment for the measurement of sound power, and for developing methods of writing meaningful noise specifications. In the area of measurement of shock and vibration, the design of improved shock test equipment and measuring instruments is needed. A baffling and important problem is that of balancing rotors that are operating above their first resonant frequency.

Other papers presented at the symposium dealt with the unsolved problems connected with the properties of matter, biological and medical acoustics, psychological and physiological acoustics and speech communication.

World Radio History

Research and Development on the Piano*

J. P. QUITTER†

Summary—**Applications of scientific methods and electronic instrumentation techniques to the study of a nonelectronic musical instrument are discussed. It is shown how findings obtained have resulted in better understanding of the physical basis for tone generation and modification, leading to improvements and evaluations based upon engineering principles rather than upon empirical methods of the past.**

Physical and psychological aspects are presented in terms of communication engineering concepts. The nature and characteristics of piano tone are described, as well as other phases of this unique field of "musical engineering."

Introduction

In VIEW of the many current papers describing the latest electronic techniques and advanced products, a discussion of a 150 year old invention may seem somewhat prosaic. Nevertheless, the general prevalence of this instrument and its extensive usage justify the economics of further research and development. In any event, some unusual applications of electronic instrumentation to art and esthetics are interesting for their own sake.

The piano endures as the most popular, versatile, and expressive musical instrument, with a considerable background of experience and tradition; but still the industry strives to improve the quality and value of its product. Evidence of this is the current "basic" music research now being conducted at M.I.T. under a grant from the National Association of Music Merchants, as well as work being done in the laboratories of individual manufacturers.

It is interesting to learn that the Russian Academy of Science has also sponsored such research, having published several admirable papers in this field.[1-4] In their work as well as ours, electronic engineers play an important part.

Since the piano is an element in a communications link relaying esthetic information from composer to performer to listener, it seems appropriate to study the signal by means of electronic instrumentation, and to describe its characteristics in the language of communications engineers. It may be said, for instance, that this mechano-acoustical transducer system consists of a multiplicity of shock-excited transient oscillators, rich in harmonics, with high-Q transmission lines as tuning elements, mutually coupled to a relatively low-Q radiator with a complex frequency-response pattern, the entire system spanning a spectrum with a frequency ratio of about 200 to 1! Concepts and measurement techniques of communications engineering are useful in the study of this medium of expression.

Most hi-fi enthusiasts and many audio engineers are not aware that piano tone is one of the more complex musical sounds—transient, inharmonic, aperiodic waves of varying level and frequency spectrum, that never quite settle down to a steady-state condition. If the ear abhors monotony, as it has been said, then the tone of the piano is certainly free from this defect. Those who may have tried to generate such a tone by electronic means, in whole or in part, have indeed acquired an appreciation for the unique and subtle properties of the oscillating system described.

Description of Instrument

A brief review of the physical description of this mechano-acoustic transducer may aid in recalling the origin and nature of the tone. The standard instrument has 88 keys, (Fig. 1), each differing in frequency from

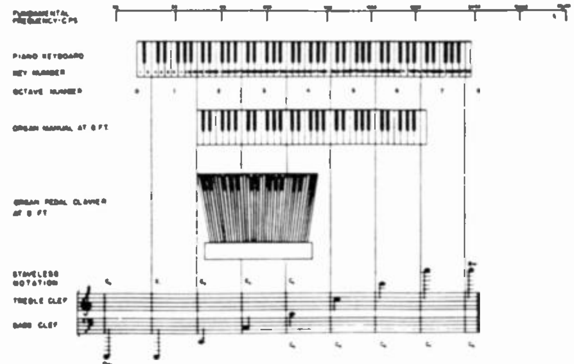


Fig. 1.

its neighbor by approximately 6 per cent, covering a fundamental frequency range of 27.5 to 4186 cps nominally, a range of $7\frac{1}{3}$ octaves (2:1 frequency ratio) from the lowest note designated A_0 to the highest note C_8. It is nominally tuned to the "equitempered" scale, which is based upon successive frequency ratios of the twelfth root of two, to one rather than upon the mathematically exact "just" scale which yields integral rating for important tonal intervals and chords, such as 3:2

* Manuscript received by the PGA, September 5, 1958; revised manuscript received, September 22, 1958. This paper is based on the paper presented by the author at WESCON, Los Angeles, Calif., August 20, 1958, entitled "Audio characteristics of piano tones."

† Baldwin Piano Co., Cincinnati, Ohio.

[1] N. I. Yakovlev, "The theory and practice of piano string excitation," *Zh. Tekhn. Fiz.*, Vol. 9, 321–328; 1939.

[2] A. I. Belov, "The reaction of the sounding board of a musical instrument on the vibration of a string," *Zh. Tekhn. Fiz.*, vol. 10, pp. 807–812; 1940.

[3] A. N. Rivin, "Experimental study of the forces acting on the string and sounding board of a piano when note is struck," *Trudy Komissii Akust.*, coll. no. 8, pp. 51–69; 1955.

[4] A. V. Rimskii Korsakov, "The dependence of frequency and wave form of a string on its supply of vibrational energy," *Trudy Komissii Akust.* coll. no. 5, pp. 83–94; 1950.

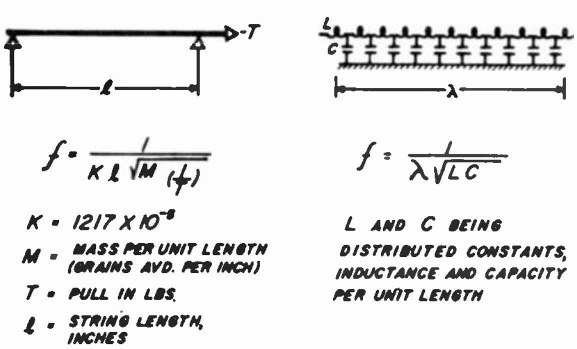
World Radio History

and 4:3. Thus it is possible to play scales of identical intervals beginning on any key, and so play in tune with a variety of instruments tuned to different keys. Actually, the equi-tempered scale itself is modified still further, as will be described later.

From about C_3 to C_8 (130 to 4186 cps), the sounds are generated by three high tensile steel strings for each tone. Below this to about G_1 (49 cps), the generator consists of two steel strings per note, copper-wrapped to obtain correct low frequencies within the available lengths. Below this, double wrapping may be required, depending upon available lengths, and one string per note is sufficient. The pull on all strings is nominally the same, varying uniformly from about 150 to 200 pounds. As might be expected, the equation for the resonant frequency of the vibrating string is of the same form as that of an electrical transmission line, viz., $f=1/\lambda\sqrt{LC}$ (Fig. 2) the terms being defined in the usual way, and in the customary units, losses being negligible. In the case of the piano string, the expression is:

$$f=\frac{1}{Kl\sqrt{M\left(\frac{1}{T}\right)}}$$

K being a proportionality constant equal to 0.01217.



(a) (b)

Fig. 2—(a) Mechanical. (b) Electrical.

In practical units, M is the mass per unit length or grains avoirdupois per lineal inch. T is the pull in pounds, l being string length in inches.

One end of the strings is rigidly terminated on metal stress-bearing structural members supporting a total load of some 40,000 pounds, and the other is compliantly connected to a relatively large diaphragm or soundboard, to improve coupling efficiency to the air for acoustical radiation. The strings are struck by felt-covered hammers at a given location over a finite distance and for a finite interval of time.

The purpose of the wooden mechanism or "action" is to transmit mechanical force from fingers to hammers with proportional dynamics and rapid repetition. Despite its somewhat archaic appearance, measurements shows this linkage to be quite efficient, and to date neither plastic moldings nor metal diecastings have proven superior or more economical. Let us proceed from this brief description to scientific considerations.

"Musical Engineering"

To establish the basic physics of the phenomena taking place in the sound of musical instruments, it is necessary to depart from the simple duplication of successful forms. To enhance existing musical instruments, or to develop new ones, the technological task force has three distinct functions.

1) The design phase, concerned with function, appearance, and the economics of manufacture. This work is based not only upon past experience, but is currently being modernized according to efficient industrial practice, and also the findings of the following two groups.

2) Acoustical investigation. As intended here, this pertains to the instrumentation techniques used in measuring physical phenomena associated with piano tone. Some of these quantities are: peak level, frequency rise time, duration, rate of decay, spectrum, dynamic response, and noise components. The findings of this group in turn must be evaluated in terms of 3).

3) Psychological testing, concerned with subjective responses such as loudness, pitch, timbre, tone color, and general acceptance of the esthetic message.

Manufacturers in general are concerned with the form and color of their products, but the musical instrument industry is unique in the variety and subtlety of subjective impressions created by its products. For this reason, instrumentation and physical data alone are not sufficient for complete evaluation of the product. In spite of sophisticated instrumentation, the human ear is the final judge. A correlation is sought between the two sets of attributes shown in Fig. 3.

BASIC CONCEPTS OF MUSICAL SOUND

SENSATION	WAVE
PITCH	FREQUENCY
LOUDNESS	INTENSITY
TIMBRE	WAVE FORM
TONE COLOR	SPECTRUM

Fig. 3.

As a member of a research and development team engaged in these activities, it is my privilege to report some of the results. Let us consider first a technical description of classical piano tone, and then means of generating and modifying it.

Description of Tone

The sound of the piano is not homogeneous throughout the scale. At least three definitive types may be

World Radio History

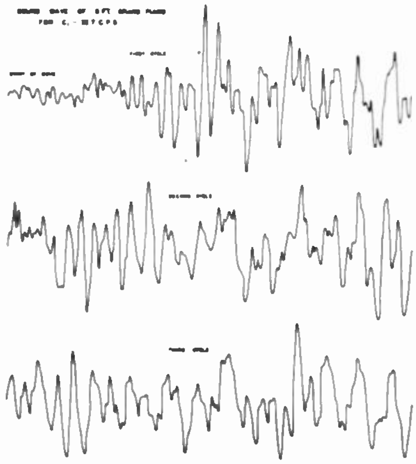


Fig. 4.

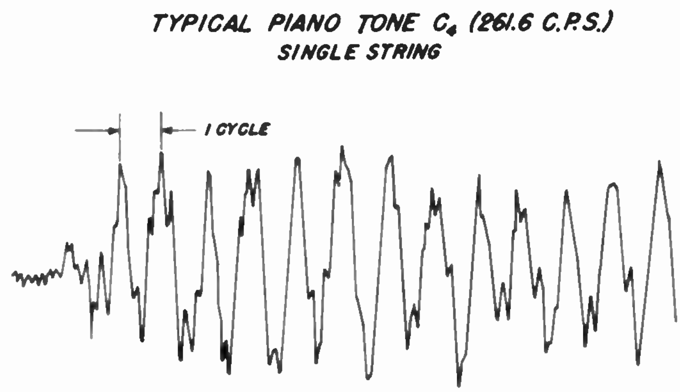


Fig. 6.

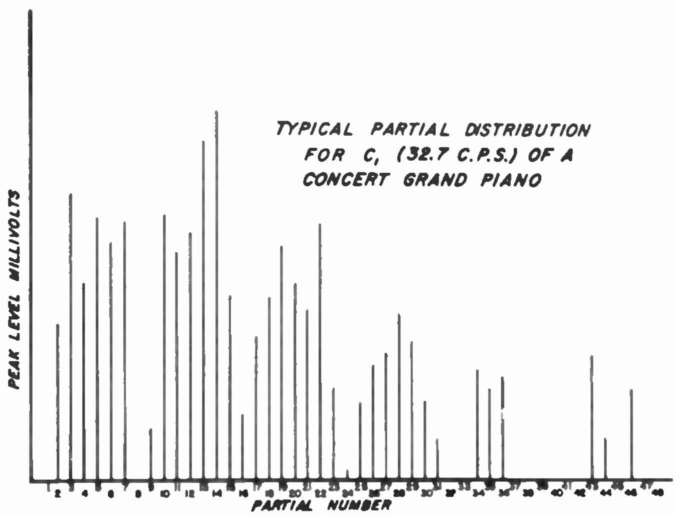


Fig. 5.

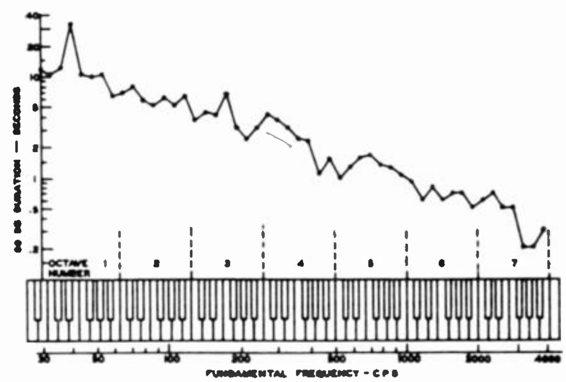


Fig. 7.

recognized, which gradually blend into each other as we proceed up the keyboard. Low tones are characterized by long durations and many partials. The word partial is used rather than harmonic, because of the progressive inharmonicity of the overtone series. Fig. 4 shows the first three cycles (about 1/10 second) of a typical low tone, C_1, 32.7 cps. Note the dissimilarity from cycle to cycle, and the noise components at the start of this wave, as viewed in the time domain. The periods were defined by extrapolating backwards from later waves which exhibited at least some periodicity.

In the frequency domain, the analysis of such a signal looks like Fig. 5, made during the course of a study on the effects of hammer felt variation. A close examination of this spectrum shows the conspicuous absence or diminution of certain partials—principally the eighth, sixteenth, and higher multiples thereof. This is because of the node caused by the hammer striking at this fraction of the string length. Curiously enough, even the first or second partial may be small, or entirely missing from the acoustic wave, although its presence may be detected upon the string itself. This absence is because the soundboard may not be large enough, depending upon the size of the piano, to radiate frequencies below 50 or 60 cps. The fact that the missing fundamental is apparently heard is a well known subjective phenomenon associated with the nonlinearity of the ear.

In the middle range of the piano, roughly from C_3 to C_6, the sound is of shorter duration, and fewer partials are present, owing to the increased string stiffness (Fig. 6). In the extreme treble, the sound is very short, on the order of a few tenths of a second, and partials are scarce.

Plotting the duration to the point 60 db. down from the peak level yields the graph of Fig. 7, first published by Martin.[5]

The sustained, yet decaying tone of the piano is one of its better known attributes. Many decay functions found in nature are exponential, as is the simple sine wave tone of a tuning fork for instance; here shown plotted against a logarithmic ordinate, Fig. 8. As might be expected from the complex nature of the source, the decay of a piano tone is not quite so simple. Fig. 9 shows an idealized version of the tone A_3, exhibiting at least two distinct slopes, also plotted logarithmically. An actual tone approaches this closely. The action of the

[5] D. W. Martin, "Decay rates of piano tones," *J. Acoust. Soc. Amer.*, vol. 19; 1947.

World Radio History

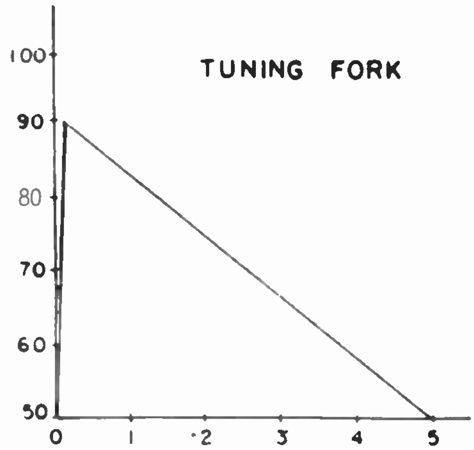


Fig. 8.

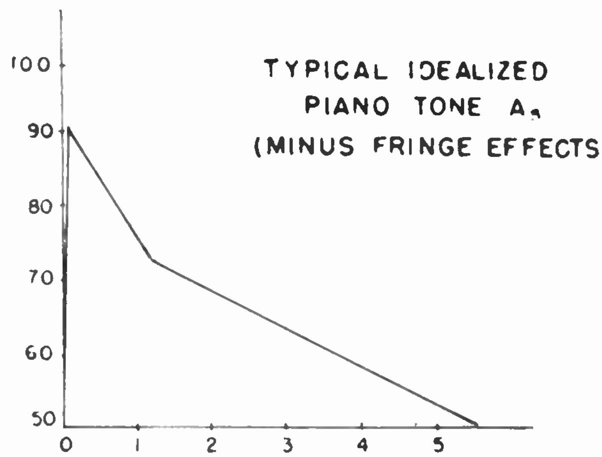


Fig. 9.

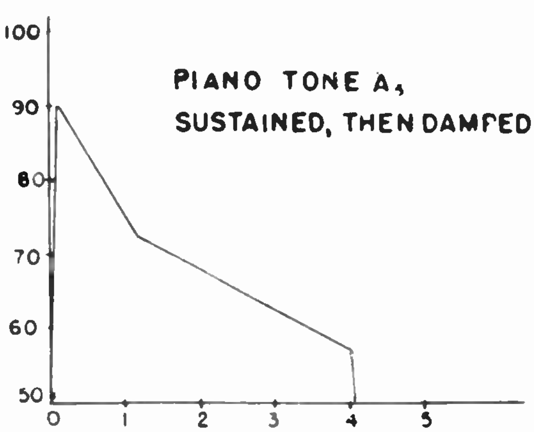


Fig. 10.

damper, operated by the foot pedal, is clearly seen in Fig. 10.[6]

Figs. 11 and 12 show the composite decay of two tones D_4 and C_5, the latter at several different force levels. Fig. 13, for F_4, shows the difference between the decay of the composite tone, including all partials, the fundamental alone, and the partials alone.

[6] T. C. Hundley, H. Benioff, and D. W. Martin, "Factors contributing to multiple decrement of piano tone envelope," *J. Acoust. Soc. Amer.*, vol. 29; 1957.

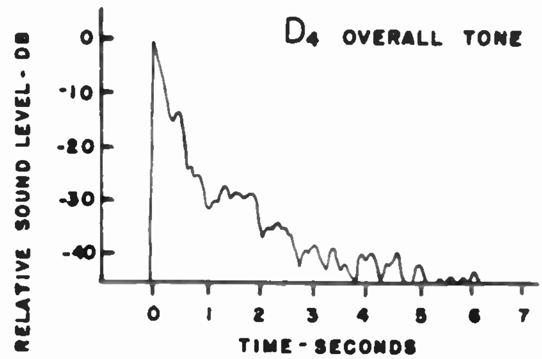


Fig. 11.

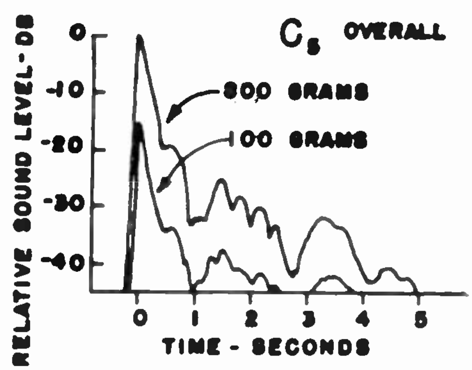


Fig. 12.

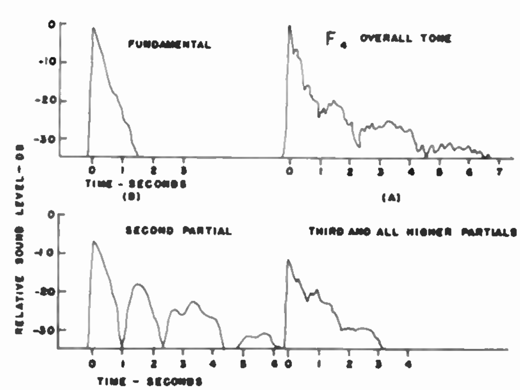


Fig. 13.

The duration of all partials is not the same. Fig. 14 makes this clear, showing that higher partials tend to die out a little sooner than the lower ones, changing the timbre of the tone during its brief history.

In connection with partials, their gross distribution over the keyboard is clearly seen in Fig. 15, which is a three dimensional plot of instantaneous peak levels. Here, too, the absence of certain partials is clearly evident. The keyboard is on the left, frequency on the right, and the ordinate is peak level.

Psychoacoustic Considerations

We have been dealing with readily measurable characteristics of piano tone. We come now to several as-

World Radio History

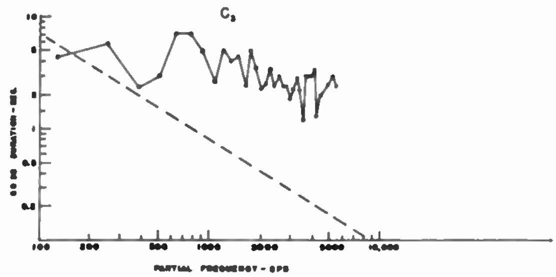


Fig. 14.

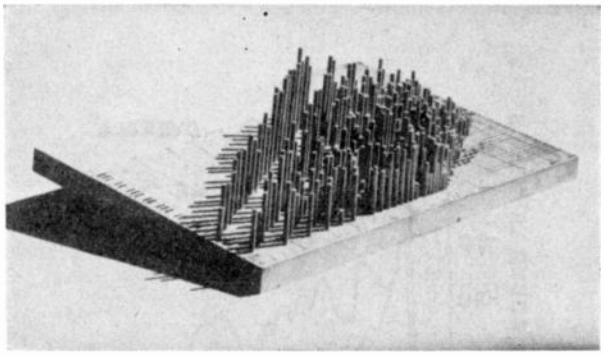


Fig. 15.

pects with definite psychological implications. First the physical evidence, dealing with multiple strings per note. It is obvious that added loudness is thus obtained; but less evident are subtle but important dispersion effects due to small phase differences between strings.

In Fig. 16 we see the wave envelope for C_4, with all three strings in exact tune. "Zero beat," communications engineers say. "Setting the unisons" say piano tuners. In the same illustration is shown the effect of mistuning. The practical difficulty of keeping several independent high-Q oscillators in synchronism during varying environmental conditions is well known to those skilled in the art, and the problem is equally difficult for the piano, typical detuned conditions also being shown. Yet such conditions are not necessarily undesirable, unless excessive. It has been shown by Kirk,[7] in studies made in company laboratories that a slightly detuned condition is preferred by most listeners. Optimum values of $\pm\frac{1}{2}$ to ± 2 cents have been found, (One cent being closely equal to 0.06 per cent or the twelve-hundreth root of two, to one) and these conditions are achieved by skilled tuners in actual practice. It is understood that a detailed paper on this subject is to be published shortly.

According to these findings, it seems possible that pianos tuned too perfectly may sound less interesting than those detuned by these slight amounts, whether intentionally, unintentionally, or as a result of natural drifts due to environment or usage. This may be why some artists feel that their instruments sound better after a recital intermission than before. Ward, Martin,

[7] R. E. Kirk, "Investigation of Tuning Preferences for Piano Three-String Unison Groups," Baldwin Piano Co., Cincinnati, Ohio, Acoust. Lab. Rep., no. A-131; 1956.

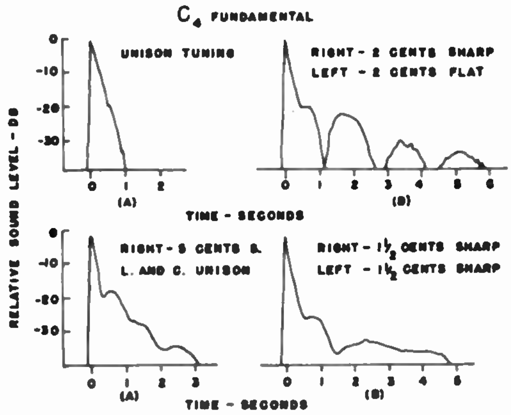


Fig. 16.

and Fay have found that such differences do occur during the playing of a concerto, as reported privately.[8]

Before leaving the matter of subjective response in actual performances, mention should be made of a paper by Martin[9] in which he shows that the stimulus presented to the audience is different in various parts of the auditorium, and also different from that heard by the performer!

Another interesting psychoacoustic phenomena is that associated with the inharmonicity of piano tone partials. Reference has been made to the "equitempered scale" to which the piano is nominally tuned. Actually, at the keyboard extremes, the frequencies depart considerably from the uniform series based upon the twelfth root of two. First observed by Railsback,[10] and reported upon later by Shuck and Young.[11,12]

This has come to be a well-known and well-documented phenomena, commonly called the "stretched scale," shown in Fig. 17. It is a consequence of tuning the piano extremes by zero-beating with partials from adjacent octaves. Since these are invariably sharp, the frequencies of the extreme treble are progressively sharp, and those in the bass flat. Ward and Martin[13] have shown that this stretch has not only become acceptable, but that "strict equal temperament was unequivocally rejected by Baldwin research engineers and music students." Similar conclusions were reached regarding the difference between the "just" and the equitempered scale. You will recall that the "just" scale is based upon intervals of exact integral ratios. It differs from the equitempered scale by as much as 14 and 16 cents at the points of greatest disparity (mi and la).

[8] D. Ward, D. W. Martin, and F. Fay, "Grand Piano Stability during Concerto Performance," Baldwin Piano Co., Cincinnati, Ohio, Acoust. Lab. Rep.; 1953.

[9] D. W. Martin, "The effect of music hall acoustics upon grand piano tone," *J. Acoust. Soc. Amer.*, vol. 25; 1953.

[10] O. L. Railsback, "Study of the tuning of pianos," (Abstract), *J. Acoust. Soc. Amer.*, vol. 10, p. 274; January, 1938.

[11] O. H. Schuck and R. W. Young, "Observations on the vibrations of piano strings," *J. Acoust. Soc. Amer.*, vol. 15; 1943.

[12] R. W. Young, "Inharmonicity of plain wire piano strings," *J. Acoust. Soc. Amer.*, vol. 24, pp. 267–273; May, 1952.

[13] D. Ward and D. W. Martin, "Psychophysical comparison of just tuning with equal temperament in sequences of individual tones," *J. Acoust. Soc. Amer.*, vol. 26; 1954.

World Radio History

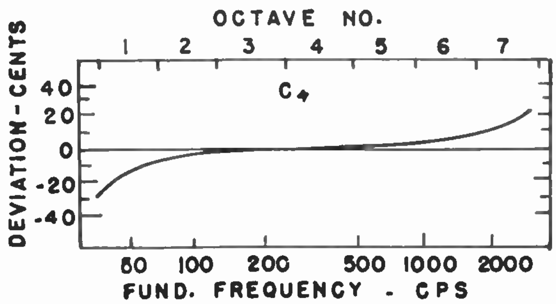


Fig. 17.

Musical purists have long held that the "just" scale is preferable, but this same paper also reports that "most people, even most musicians, cannot distinguish short melodic sequences in just intonation from those in equal temperament"; and further that "even when they can tell the difference, they will prefer equal temperament, the just scales often sounding flat."

Let us turn now to the actual mechanics of tone generation.

Tone Generation

Perhaps the most penetrating practical analysis of the mechanics of struck strings is Kock's classic paper.[14] The concept of analogy between electrical and mechanical quantities comes from the fact that mathematically the differential equations describing either system correspond one for one with the other. For example, a second-order system with one degree of freedom is represented mechanically by a single rigid body, and electrically by a single loop, or single node circuit. Differential equations describing them are given in Fig. 18. Analogy I is the conventional one, called the "old analogy" by Firestone. Analogy II was also proposed by Firestone[15] and Analogy III was used by Kock.

Of course, there is no reason why one cannot study the vibrating string without ever thinking about transmission lines, just as one can study transmission lines without knowing anything about vibrating strings. Nevertheless, it is something which may interest communications engineers.

Kock drew several important conclusions about practical aspects of piano manufacture. One point concerns hammer contact time. Referring to Fig. 19, we see that in striking the string, the hammer initiates a wave which travels away from the source—much as does the wave in a pond when a stone is tossed in.

The wave element traveling to the left is reversed in phase at the near end of the string, returns relatively quickly, and serves the function of reversing the direction of the hammer, sending it back on its return way. In the meantime, the other element sweeps over the bridge, actuates the soundboard with a pulse of energy,

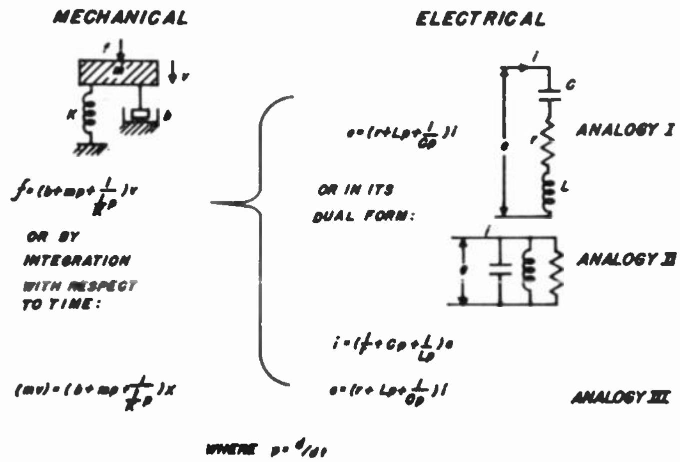


Fig. 18.

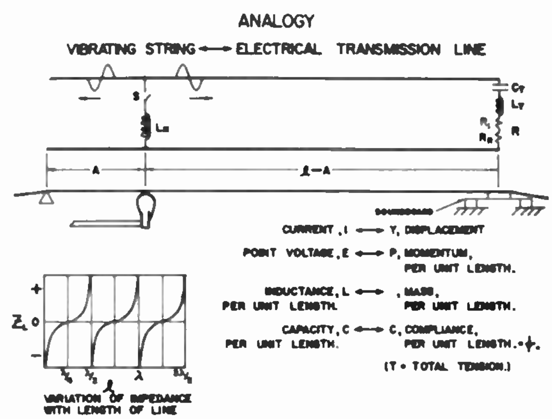


Fig. 19.

reverses its phase and heads back for the origin. Now if the hammer should still be in contact with the string, it will damp the wave, to the detriment of the tone.

A maximum permissible contact time may be computed as the product of the period and the percentage of the string defined by the striking point or $(l-A)/l$. This time is plotted against frequency in Fig. 20, together with some actual times observed during the course of an experiment with hammers. From such data, valuable clues are obtained as to the direction in which improvements should be sought.

Let us return again to the strings themselves, and consider the effect of the three sources upon the composite wave envelope. Fig. 21 shows that even though individual strings decay smoothly, the resultant is not simple.

Sometimes it is observed that even single strings do not exhibit smooth decay rates. One reason for such behavior are minute termination differences in horizontal and vertical planes, resulting in energy transfer from one mode to the other (Fig. 22). Motion perpendicular to the soundboard causes it to vibrate and radiate. Parallel motion does not. A string may oscillate in both planes at once, resulting in circular or elliptical motion, the axes of which may rotate slowly. This is why string termination is so important, and why such great care is taken in this respect by particular manufacturers.

[14] W. E. Kock, "The vibrating string considered as an electrical transmission line," *J. Acoust. Soc. Amer.*, vol. 8, pp. 227–233; 1937.

[15] F. A. Firestone, "The mobility method," *J. Appl. Phys.*, vol. 9; June, 1938.

World Radio History

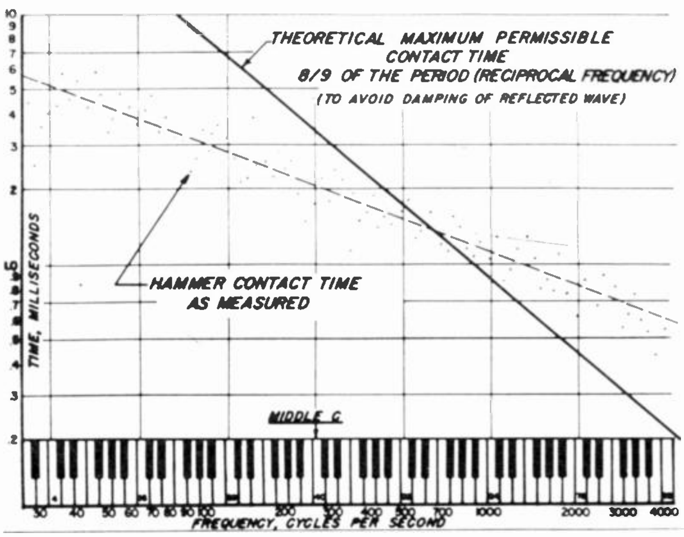


Fig. 20.

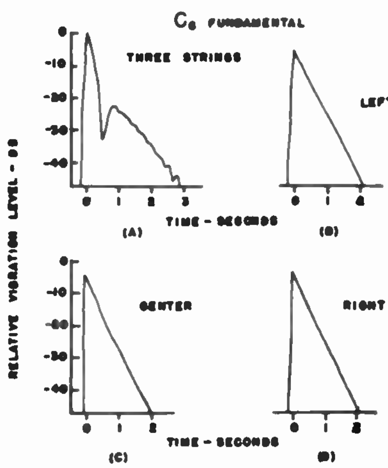


Fig. 21.

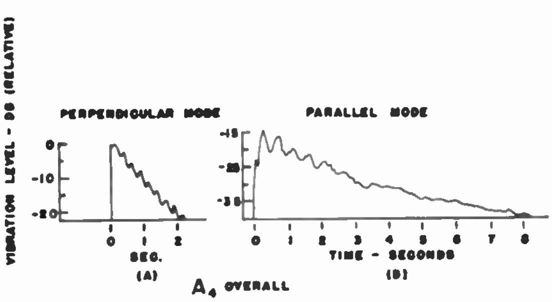


Fig. 22.

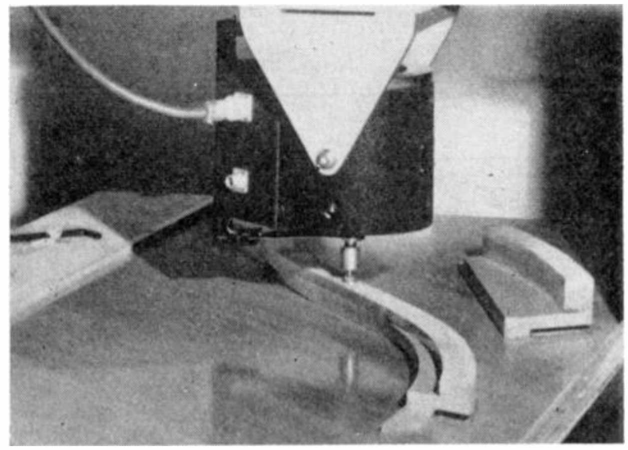


Fig. 23.

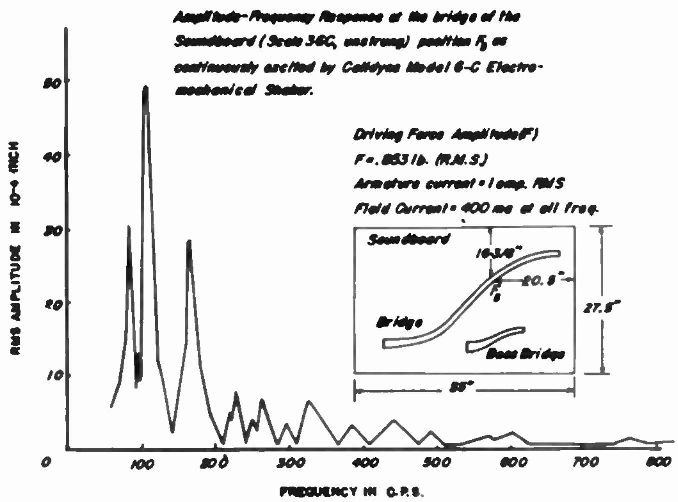


Fig. 24.

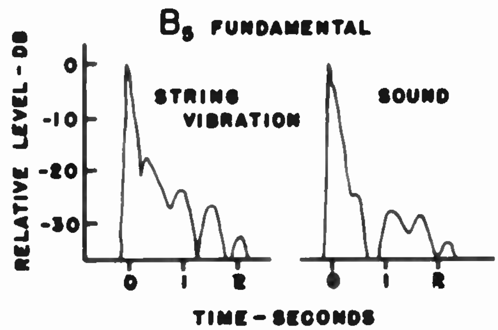


Fig. 25.

Tone Modification

The brief scope of this paper does not permit greater amplification upon results of the work described, but several other research tools may prove interesting. One is the electric vibrator or shaker. Just as aeronautical engineers on the West Coast gather useful data by vibrating aircraft parts, similar work with piano soundboards, Fig. 23) has shown the complex nature of their resonances and response patterns (Fig. 24). Work is also being done by means of step function excitation.[16] Results are not yet available. The effect of these resonances is seen in Fig. 25 which shows the difference between string vibrations and resultant sound.

Electric strain gauges are used in studying static stress distribution in piano plates, and also as transducers for dynamic electrical signals (Fig. 26).

Stroboscopic tuning devices have yielded valuable data about the tuning stability of pianos under extreme environmental conditions (Fig. 27), plotted here in three dimensions.

"Electro-Pianos"

Before closing, brief reference will be made to piano-like instruments involving electrical amplification, the

[16] A. B. Kaufman, "Measurement of damping and natural frequency," *Instruments and Automation*, vol. 28; December, 1955.

World Radio History

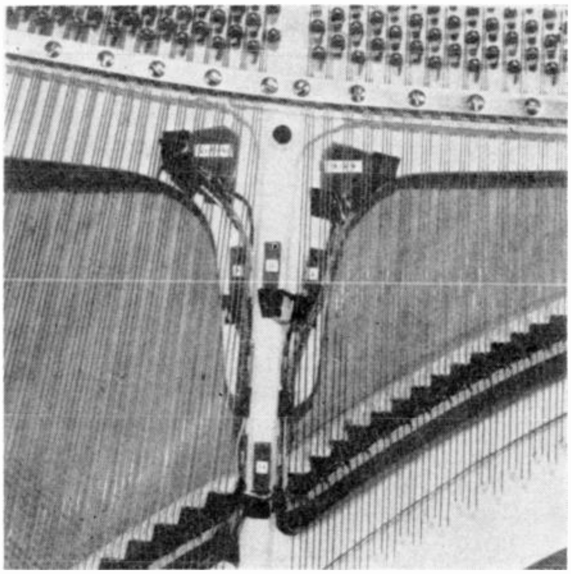


Fig. 26.

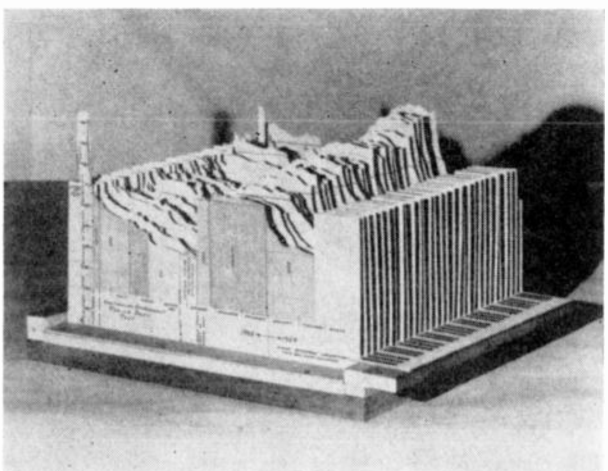


Fig. 27.

design objectives being portability, the possibility of wide sound distribution, and the potentiality for new effects. Brief consideration will reveal that all-electronic signal generation is not economical. In practice, either of two mechanical vibrators are used: struck reeds, or struck miniaturized strings (Fig. 28). In working with such instruments, quantitative data regarding duration is useful.

Duration to the 60-db down point has already been given, as well as evidence of the multiple decay rates actually obtained. To a first approximation however, single exponential rates are useful for empirical work. These have been worked out as shown in Fig. 29 for selected notes throughout the scale, and summarized in a general formula plotted in Fig. 30.

The agreement with actual observed data is fairly good.

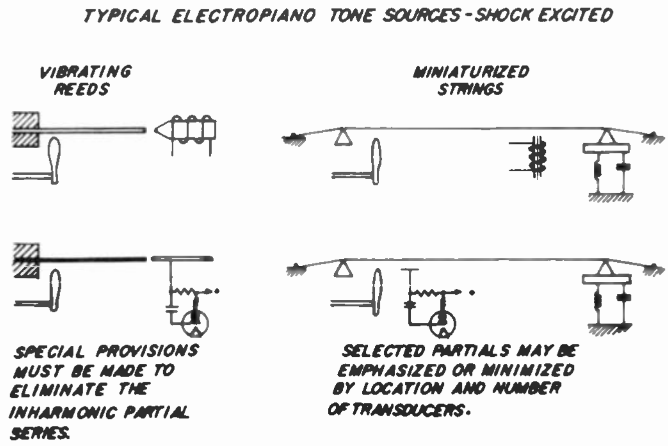


Fig. 28.

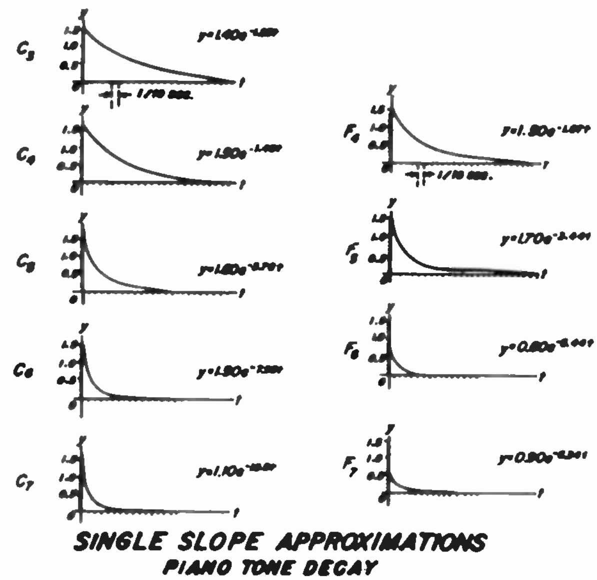


Fig. 29.

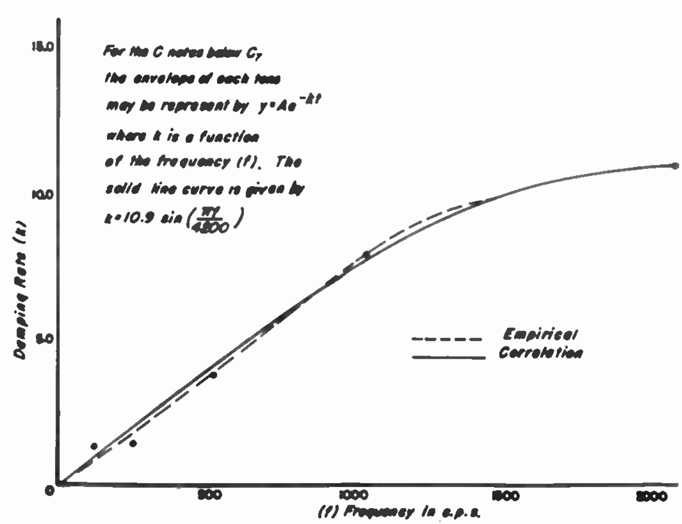


Fig. 30.

Conclusions

The results of these investigations make it possible to attach physical meaning to those purely empirical recommendations have been used to date in piano construction.

Acknowledgment

The author extends thanks to various associates at the Baldwin Piano Company in the Acoustical Laboratory and the Piano Research Laboratory for the use of certain data, as well as to the other sources cited.

World Radio History

A Survey of Speech Bandwidth Compression Techniques*

S. J. CAMPANELLA†

Summary—**Application of speech bandwidth compression techniques to voice communications provides the promise of more efficient utilization of the available radio spectrum and the possibility of improved performance of noisy, long distance communications links. That significant potential gain exists is evident from the fact that the information rate required for transmission of the conventional speech signal is approximately 24,000 bits per second, whereas that for transmission of the equivalent word-intelligence content by means of teletype is 75 bits per second. The paper presents brief descriptions of several speech bandwidth compression techniques which are currently being employed or investigated to achieve varying degrees of compression. Also a method of estimating the influence of signal-to-noise ratio on a communications link employing such compression techniques is presented.**

Introduction

FOR the purpose of conveying information from one human being to another, speech communication is the vehicle most readily accepted and rapidly comprehended. The human speech communication system may be considered to consist of the brain of the speaker as the source of information, the vocal mechanism of the speaker as the transmitter, the hearing mechanism of the listener as the sensory device, and the brain of the listener as the receiver of information. Throughout his evolution, man has naturally adapted himself to this communication system, using the speech signal as the medium of communication.

Unfortunately, from the point of view of the communication engineer, nature has not provided an efficient system. This fact is made evident when the information rate required to transmit the speech signal of 3000 cps bandwidth (24,000 to 50,000 bits per second) is compared with that theoretically required to transmit the basic information content of the speech signal (approximately 60 bits per second).

This inefficiency results in uneconomical use of channel capacity and radio spectrum space. The desire for more efficient use of our communication facilities for speech, coupled with the advantages of low information rate digital transmission for both purposes of privacy and of improved reliability, has led to the investigation of several speech bandwidth compression systems in this country and abroad.

Characteristics of the Speech Signal Spectrum

It is the idea-forming element of the brain acting on the vocal mechanism that gives rise to the speech signal. A simplified sketch of the vocal mechanism is shown in Fig. 1. It consists of a tube terminated at one end by the larynx and at the other end by the lips. Acoustic excitation may be applied either in the form of a periodic pressure wave generated at the larynx, or by turbulence generated at some point of constriction along the tube. Sounds produced by the periodic larynx excitation are usually referred to as being voiced, and sounds produced by turbulent excitation as unvoiced.

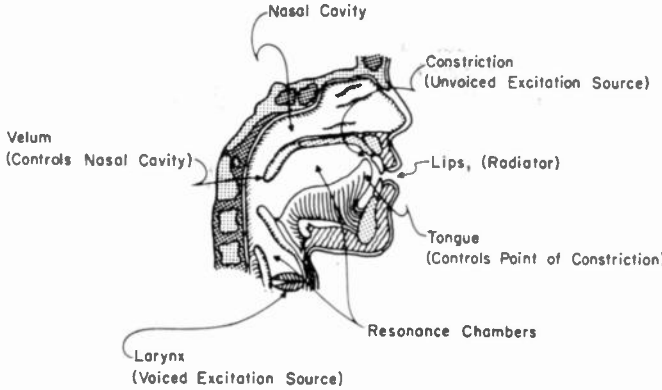


Fig. 1—The vocal mechanism.

The principal element entering into vocal tract variation is the tongue. It divides the vocal tract into two resonant cavities which to a large extent control the transmission characteristic of the vocal tract. The transmission characteristic can also be modified by coupling into the nasal cavity controlled by the velum. The influence of the vocal cavity on voiced excitation is illustrated in Fig. 2. The vocal cavity transmission characteristic acting on the larynx signal 2(a) causes certain frequency components to pass with less attenuation than others. The effect of a simple resonance transmission characteristic would produce a damped sinusoid, as shown in Fig. 2(b). For the more complicated transmission characteristic of the human vocal tract, the result is best illustrated by the spectral energy distribution shown in Fig. 2(c). The line structure appearing in the spectrum is produced by the harmonics of the periodic larynx excitation. The frequency of larynx vibration is generally referred to as the pitch frequency. The peaks of energy observed in the spectral distribution, resulting from the influence of the transmission characteristic of the vocal cavity, are referred to as formants. The formants are designated in their ascending order of appearance as F1, F2, etc. In general, sta-

* Manuscript received by the PGA, September 2, 1958; revised manuscript received October 27, 1958. This paper was presented at the IRE WESCON Convention, Los Angeles, Calif., August 20, 1958.

† Melpar, Inc., Falls Church, Va.

World Radio History

tistical studies have indicated that two formants suffice for specification of most of the voiced English vowels and consonants.[1] Similar formant structure is also observed for the unvoiced sounds of speech. Most of the significant information content of speech falls in the frequency range below 3000 cps.

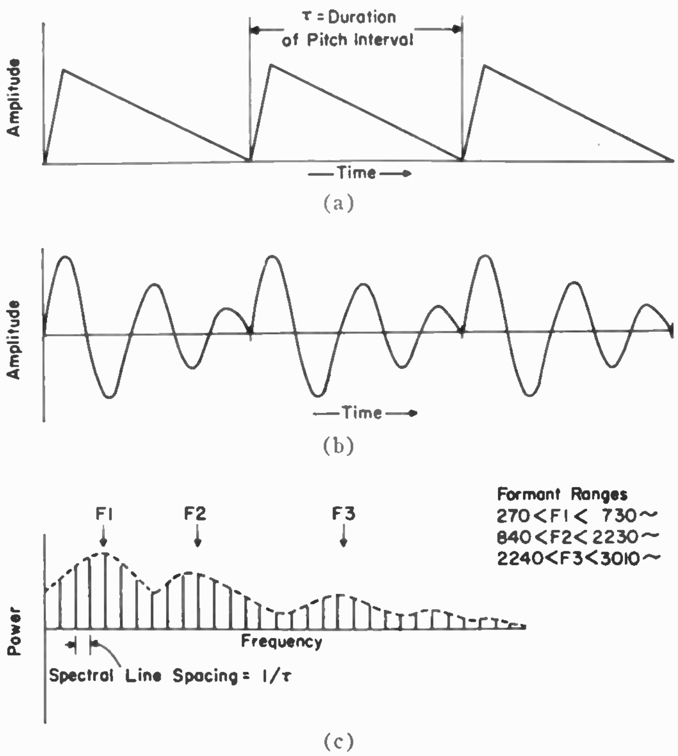


Fig. 2—Typical speech waveforms and spectral distribution (a) Pressure wave-shape at larynx. (b) Pressure waveshape at lips. (c) Spectral distribution of speech signal energy.

Bandwidth Compression Schemes

In conventional telephony, speech is transmitted by the more or less exact replica of the complicated acoustic waveforms that represent the sounds. In fact, if speech is regarded as a function of time, the process consists in principle of a mere change of the dependent variable from sound pressure to some electrical quantity at the sending end and vice-versa at the receiving end. It is possible to estimate the rate of information transmission for the speech signal by considering the normal telephone channel. The conventional telephone channel has approximately 3000 cps bandwidth and performs well at a signal-to-noise ratio of 20 db. Employing Shannon's relation for channel capacity,[2]

$$C = 2W \log_2 \left(1 + \frac{S}{N}\right) \qquad (1)$$

where W is the bandwidth of the channel and S/N is the signal-to-noise ratio, the capacity of such a channel is found to be equal to 6000 $\log_2(1+100)=40{,}000$ bits per second. Marginal performance might be achieved at a signal-to-noise ratio of 10 db, this giving a channel capacity equal to 6000 $\log_2(1+10)=20{,}000$ bits per second.

Consider the rate of speech information transmission now from the point of view of its source. In speech, a human utilizes approximately 40 basic sounds and produces these at an average rate of 10 per second. Using a six bit code, it is possible to uniquely specify $2^6=64$ sounds. Hence, the rate of information transmission of the speech channel is seen to be less than 60 bits per second.

So, there is a considerable difference between the information rate necessary to communicate the speech signal waveform and the actual rate at which information appears to be generated by the vocal mechanism. Several factors may be offered in explanation of the fact. It is rather obvious that in addition to the information contained in the principle articulatory positions of the vocal mechanism, each individual speaker also imposes information concerning his vocal identity and emotional status. This information is transmitted continuously, since at any instant in a conversation one can usually identify the speaker from the "sound" of his voice. Thus, not only does the speech signal identify the speaker initially, but it continues to do so throughout the event. This points out the fact that the speech signal is redundant. Speaker identity is not the only form of redundancy present in speech. Another significant form of redundancy is evident in the fact that most vowel sounds have a duration in excess of that required for the hearing mechanism to identify the sound, and furthermore, the typical vowel sound contains many fold repetitions of a basic vowel waveshape. A final factor influencing the speech channel information rate is inefficient use of the spectrum, *i.e.*, the speech signal does not occupy all of its spectrum space all of the time.

To summarize, there is a considerable difference between the information rate necessary to communicate the speech signal waveform in the conventional manner and that contained in the articulartory action of the vocal mechanism. The difference can be attributed to identity of the speaker and his emotional status, redundancy, and inefficient use of the spectrum. In general, all speech bandwidth compression systems attempt to exploit one or more of these factors to obtain compression of the bandwidth and consequently channel capacity (in bits per second).

The fact that communication of speech information by transmitting a replica of the speech waveform has been and remains the most widely used technique and is virtually considered as fundamental is undoubtedly due to the inherent simplicity and reliability of the equipment involved. Only recently the ever increasing demands on radio spectrum space and the requirement to achieve communication over long distant radio channels at reduced channel capacities have spurred con-

[1] G. E. Peterson and H. L. Barney, "Control methods used in a study of the vowels," *J. Acoust. Soc. Amer.*, vol. 24, pp. 175–184; March, 1952.

[2] W. G. Tuller, "Theoretical limitations on the rate of transmission of information," Proc. IRE, vol. 37, pp. 468–478; May, 1949.

World Radio History

siderable activity in achieving a means of voice communication which is more conservative in terms of channel capacity requirements.

In the following discussion several speech bandwidth or channel capacity reduction systems are described. In general these systems can be grouped into four principal categories:

1) Time or frequency compression methods. Such methods exploit the redundancy or regularities existing in the speech signal by sampling and frequency division techniques. These systems generally exhibit bandwidth compression in the order of 1:2 to 1:4 and can be transmitted in binary code form over channels of 5000 to 10,000 bits per second channel capacity.
2) Continuous analysis-synthesis methods. Such methods transmit in place of the speech signal spectrum a description of the spectrum in terms of a number of analog parametric control signals. As such, they exploit both the redundancy and inefficiency existing in the speech signal. In general, these systems exhibit bandwidth compression in the order of 1:10 to 1:20. A binary channel capacity of 1600 to 2000 bits per second has been demonstrated for vocoders, and it is expected that this channel capacity will be below 1000 bits per second for formant coding schemes.
3) Discrete sound analysis-synthesis methods. Such methods transmit in place of the speech signal code groups which identify the fundamental sounds that constitute the speech. As such, they exploit the redundancy and inefficiency of the speech signal, and, in addition, remove speaker identity and emotional status cues. These methods are best categorized in terms of the binary channel capacity required for transmission of the code groups. It is expected that such systems should be capable of transmitting speech at information rates as low as 60 bits per second.
4) Sound group analysis-synthesis methods. Such methods transmit only certain groups of sounds (particular words and phrases), each identified by a code group. Information rates in this case are, of course, a function of the size of the vocabulary. Such a system appears to be most useful at information rates in the order of 5–10 bits per second.

Time or Frequency Compression Methods

Doppler Frequency Compression

One of the principal forms of redundancy present in the speech signal is repetition of a waveshape characteristic of the sound generated. This is especially true for the prolonged vowel sounds. Consequently, it is possible to sample the speech signal periodically, and to ignore the remainder of the speech signal between the samples.[3] Each sampling interval should be long enough to obtain a sample representative of the sound (the pitch period would be the minimum). By the proper speech reduction apparatus, it is possible to simultaneously sample the speech and spread the sampled portion over the entire interval between samples. A method of accomplishing this is illustrated in Fig. 3. The sound signal to be processed is carried on magnetic tape. The tape is passed over a rotating playback head assembly with heads spaced 90° apart at positions *A*, *B*, *C*, and *D*.[4] The magnetic tape is in contact with the head over an arc of 90°; thus, one of the playback heads is always in contact with the surface of the magnetic tape.

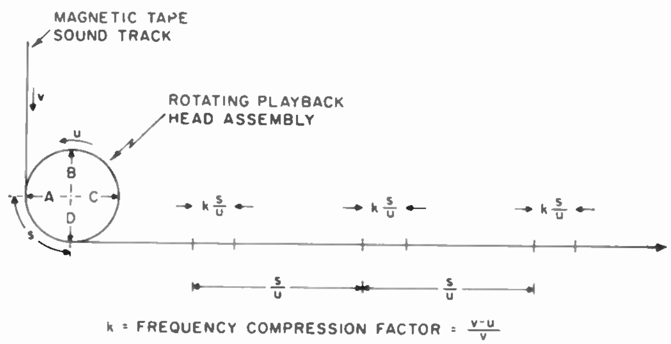


Fig. 3—Sampling action of Doppler frequency compressor.

Let s be the distance of contact along the arc, u the peripheral speed of the playback heads and v the speed of the magnetic tape. The time required for one of the playback heads to move through the distance s is s/u. This is the duration of the interval between samples. During this same time interval, the magnetic tape will move a distance $s(v/u)$, and the net slippage between the tape and the playback head will be $s(v-u)/u$ (length of the tape sampled by the playback head). Since the speed of the tape is v, the duration of the sample is $s(v-u)/uv$. The compression ratio is equal to the ratio of duration of sample to duration of interval between samples. Hence

$$k = \text{compression ratio} = \frac{v-u}{v}. \tag{2}$$

Since the sampled signal is spread over the entire interval, this is the ratio by which the frequencies of the speech signal are compressed. Also, it is identical to the relation for shift due to Doppler effect. So the name "Doppler frequency compression" may be applied to the technique.

Actually, the method may be used for both frequency compression and expansion. One unit, placed at the transmission end of the system, would be used to compress the signal by the factor k, and another unit, placed at the receiver, would expand the signal back to the

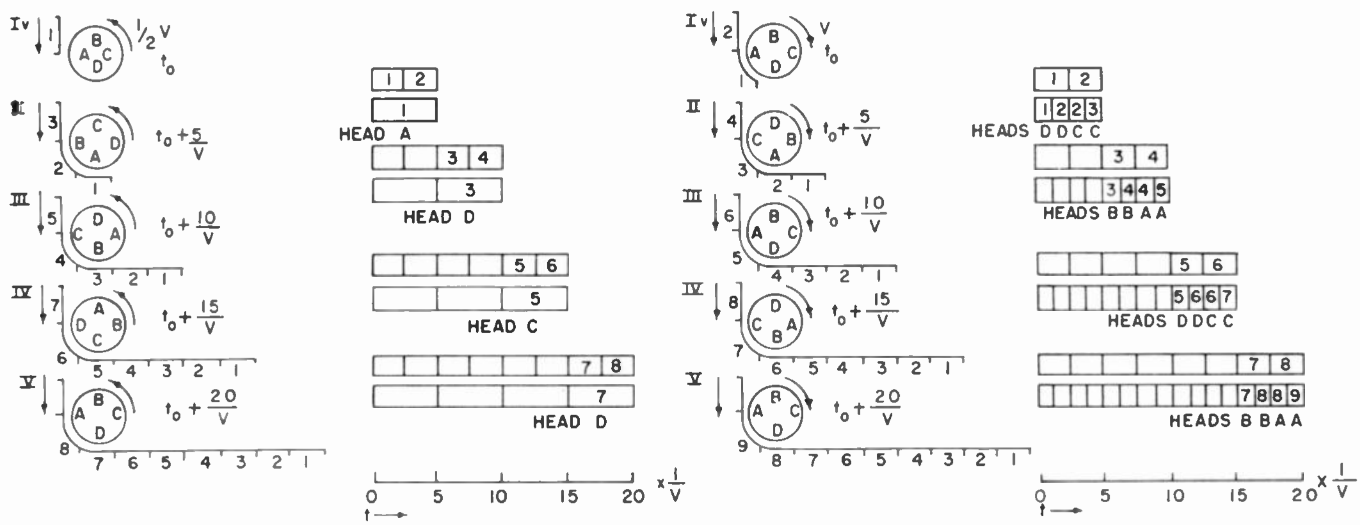
[3] G. E. Peterson, Ph.D. dissertation, Louisiana State Univ., Baton Rouge; 1939.

[4] G. Fairbanks, W. L. Everitt, and J. P. Jaeger, "Method for time or frequency compression-expansion of speech," 1953 IRE Convention Record, pt. 8, pp. 120–124.

World Radio History

normal frequency range by the factor $1/k$. Operation of such a compression-expansion system is illustrated in Fig. 4. The compressor has a compression ratio of one-half and the expandor a compression ratio of 2 (hence, expansion). Action of the compressor is shown in Fig. 4(a). For this discussion, the tape is considered to be divided into numbered segments, each equal to the peripheral distance between heads. At time I segment 1 is initially in contact with head *A*, as shown. As head *A* moves through a 90° arc to cover a distance s, the tape moves a distance $2s$, ending up in the position shown at time II. The total slippage is therefore a distance s; thus head *A* scans entirely tape segment 1 at half speed. So segment 1 is stretched over the time occur in synchronism with the pitch period of the speech, and the sampling interval is equal to the duration of an integral number of pitch periods. For this ideal case, half of the pitch waveshapes are played back at half speed, thus effectively reducing all spectral components by one-half, and consequently reducing the bandwidth by the same factor.

Systems employing the doppler frequency compression principle have been developed by Fairbanks *et al.*,[4] Gabor,[5] and Vilbig.[6] Experimental results indicate that compression factors in the order of 1:4 are possible, and that by observing proper precautions, it may be possible to obtain compression ratios of the order 1:6. Gabor points out that sampling noise is a very serious limita-



(a) (b)

Fig. 4—Illustration of frequency compression and expansion for $K=\frac{1}{2}$. (a) Compression process, $K=\frac{1}{2}$. (b) Expansion process, $K=2$.

interval occupied by segments 1 and 2. The end of the first scan and the start of the second is shown at time II. Note that head B is coming in contact with segment 3, and that segment 2 will be entirely skipped by both heads *A* and *B*. The scanning action continues for all heads as shown at times III and IV, and finally at time V the cycle is started over again. The compressed signal output, which is the sum of the outputs of all four heads, skips all even numbered segments and plays out all odd numbered segments at half speed.

The expansion process is shown in Fig. 4(b). At time I, head *A* is initially at the boundary between segments 1 and 2. Between time I and II, D will scan both segments 1 and 2, and *C* will scan segments 2 and 3. The same scanning action continues for the other heads as shown at times III, IV and V. The expanded output, which consists of the sum of the outputs of all four heads, is seen to consist of double playbacks of each segment at a speed-up factor of 2.

Action of a compressor with a compression ratio of one-half on a typical speech waveshape is shown in Fig. 5. This case is idealized since the sampling intervals

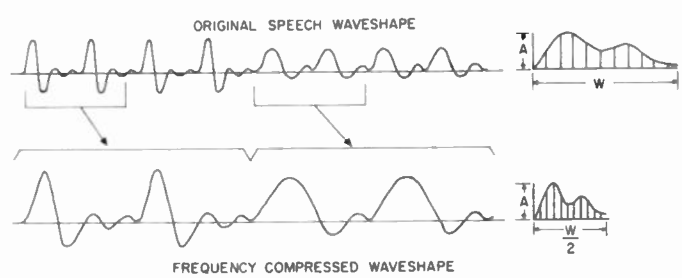


Fig. 5—Idealized frequency compressed speech waveform for $K=\frac{1}{2}$.

tion. He has constructed a device in which a multiplicity of overlapping samples, each modulated in amplitude by an error function, are summed to achieve a minimum transient condition in turning a sample on and off. Gabor also points out that the compressed and expanded speech contains not only spectral components

[5] D. Gabor, "New possibilities in speech transmission," *J. IEE*, vol. 94, pp. 369–390; November, 1948.

[6] F. Vilbig, "Frequency band multiplication or division and time expansion or compression by means of a string filter," *J. Acoust. Soc. Amer.*, vol. 24, pp. 33–39; January, 1952.

World Radio History

related to the original pitch frequency of the input speech, but also spurious components which are caused by the sampling action. He shows that these components may produce some roughness in the processed speech and proposes to eliminate this by synchronizing the sampling action with the pitch frequency of the speech signal.

Pitch-Synchronous Processing of Speech

The method of pitch synchronous processing[7] of speech proposes to exploit the pitch period regularity in speech to achieve information rate compression. In general, the generation of impulses by the larynx occurs at a rate considerably in excess of that exhibited by the relatively slowly moving vocal cavity. Literally, the impulses from the larynx appear to sample the size and shape of the vocal cavity; the resulting signal is radiated at the lips in the form of speech. The periodic impulse rates vary generally from as low as 70 cps for male speakers to as high as 500 cps for female speakers. Knowledge of the periodicity affords a means of reducing the channel capacity necessary to transmit a speech signal. The pitch synchronous method proposes to eliminate $N-1$ of every N pitch periods before transmission, and to restore the missing parts at the receiver by simply repeating each of the received periods $N-1$ times. Provided consonants can be similarily treated, the channel capacity in bits per second required to accommodate the chopped speech signal is reduced to $1/N$ of that required for the entire signal.

A block diagram of the transmitting and receiving terminal devices is shown in Fig. 6. The input speech signal is applied both to a gate and a pitch pulse generator. The pitch pulse generator produces a single impulse for each pitch period. These impulses are fed to a scale of N counter. The output of the scale of N counter opens the gate via a gate control circuit for a full pitch period. Hence the gate circuit passes a pitch period waveshape for every N pitch periods and remains closed for $N-1$ periods. The output may be referred to as the reduced representation. It can subsequently be coded to match a given transmission channel whose capacity need be $1/N$ of that required if the entire wave were to be transmitted. At the receiving terminal, the reduced representation is used to synthesize the original wave as closely as possible. Where a number of pitch periods are removed, their places are filled by repetitions of the previous retained period. As shown in Fig. 6, this is accomplished by supplying the reduced representation to a set of pitch-interval delay lines. For a removal of $N-1$ pitch intervals in the reduced representation, $N-1$ delays are used. The outputs of all delay lines are summed to provide the restored speech signal.

[7] E. E. David and H. S. McDonald, "Note on pitch-synchronous processing of speech," *J. Acoust. Soc. Amer.*, vol. 28, pp. 1261–1266; November, 1956.

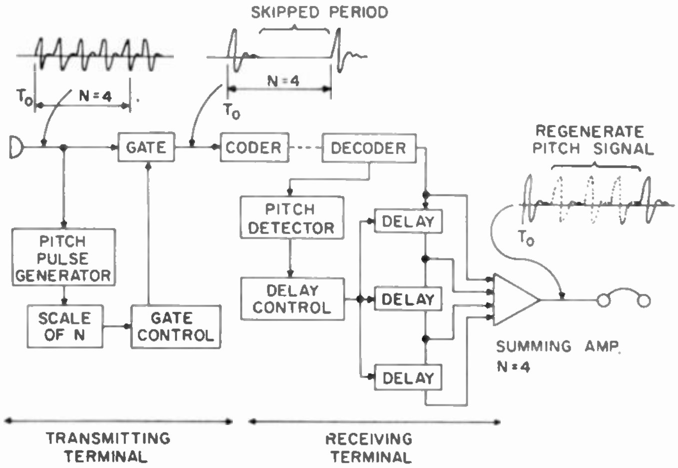


Fig. 6—Block diagram of a pitch synchronous chopping system. (Courtesy *J. Acoust. Soc. Amer.*)

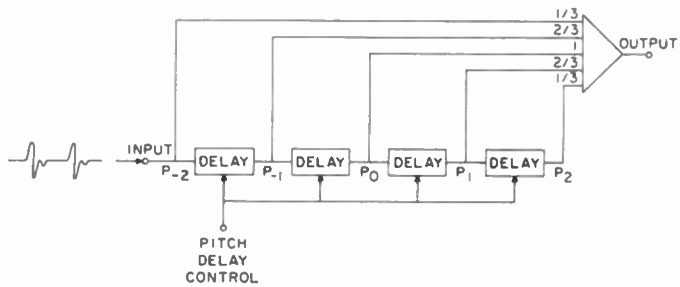


Fig. 7—Linear interpolator for $N=3$. (Courtesy *J. Acoust. Soc. Amer.*)

Provided that the speech intervals are periodic between samples, the restored speech signal may be quite accurate; however, should appreciable variation be exhibited between sampling intervals, considerable distortion in the form of low-frequency "gargle" or "burble" can occur. One method available for overcoming the difficulty is to perform a linear interpolation on the pitch period samples immediately preceding and following the zero interval in the reduced representation. By this means the intervening pitch intervals can be restored without creating sharp discontinuities. A circuit for accomplishing this for $N=3$ is shown in Fig. 7. A double set of pitch-length delay lines is used to supply both preceding and following periods to a weighted-summing network. Let the pitch period preceding the zero interval of the reduced representation be P_{10} and the one following be P_{20}; then, if $P_1, P_2, P_3 \cdots P_{N-1}$ are the periods to be inserted,

$$\begin{aligned} P_1 &= \frac{N-1}{N} P_{10} + \frac{1}{N} P_{20} \\ P_2 &= \frac{N-2}{N} P_{10} + \frac{2}{N} P_{20} \\ &\vdots \\ P_{N-1} &= \frac{1}{N} P_{10} + \frac{N-1}{N} P_{20}. \end{aligned} \tag{3}$$

World Radio History

This system of speech compression has been tested at the Bell Telephone Laboratory and reported by Davis and McDonald.[7] For this test, a channel vocoder was used as a convenient source of monotone speech. This procedure eliminated the necessity of making the pitch delay lines variable and was considered sufficient to demonstrate feasibility of the approach.

These results were reported. Elimination of $N-1$ of every N-pitch period does not destroy the fundamental phonemic information for values of N as great as 6 and a pitch frequency of 200 cps. The restored speech is highly articulate, although it is somewhat distorted by harmonic effects. The latter effect can be reduced by interpolating the missing parts as linear combinations of the adjacent terms rather than by simply repeating them. The unvoiced consonants appear to be relatively undisturbed by this processing, even though they are gated with the same duty cycle as the vowels.

Vobanc

The Vobanc,[8] (*Vo*ice *Ban*d *C*ompression) is a speech bandwidth compression system which provides a reduction of two in the transmission channel bandwidth required to accommodate the speech signal. The general principle is to divide the speech band into three parts—0.2–1 kc, 1–2, kc and 2–3.2 kc—by filters located at carrier frequency. Each of these bands contains one of the three principal vowel formants. The signal in each band is passed through a regenerative modulator, which halves the frequency of the strongest components of the formant and translates the neighboring frequency components downward correspondingly. The output of the regenerative modulator[9] is filtered to a bandwidth one-half that of the original. At the receiving end, each of the component bands is frequency doubled and recombined to occupy the normal speech spectrum range.

A block diagram of the Vobanc is shown in Fig. 8. The input speech signal is modulated by a 108-kc oscillator. The difference frequency components are selected by the A filters in three separate channels. Transmission characteristics of the A filters are shown in Fig. 9(a). The A1 filter transmits the band from 107.8 to 107 kc, which corresponds to the difference components resulting from the product of the local oscillator and the 0.2 to 1-kc first formant range. Second and third formant ranges are transmitted by filters A2 (107 to 106 kc) and A3 (106 to 104.8 kc) respectively.

The output of each of the A filters is supplied to a regenerative modulator. A block diagram of the modulator is shown in Fig. 10. The circuit consists of a balanced modulator for which the local oscillator input is fed back from the modulator output via a filter which

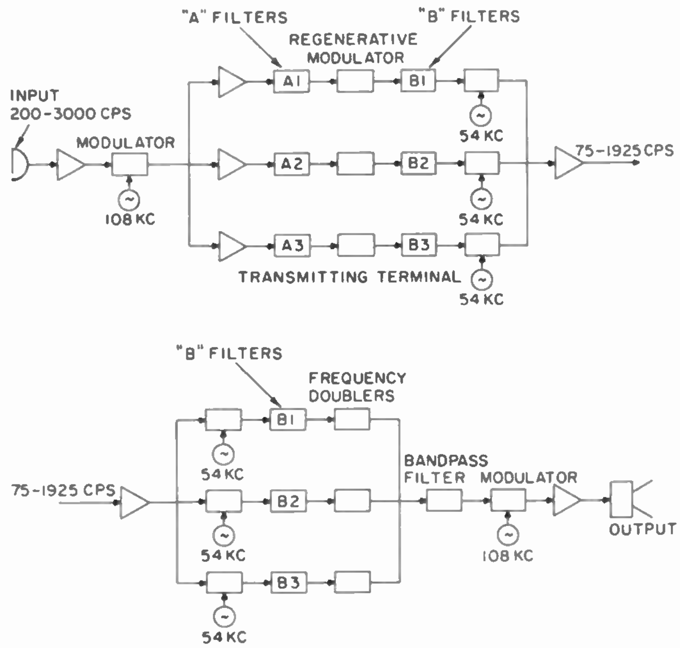


Fig. 8—Block diagram of the vobanc. (Courtesy *J. Acoust. Soc. Amer.*)

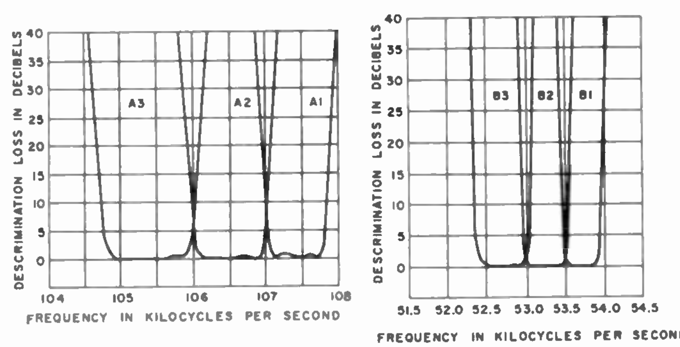


Fig. 9—Vobanc filter response characteristics. (Courtesy *J. Acoust. Soc. Amer.*)

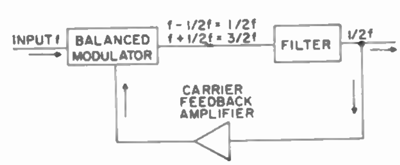


Fig. 10—Block diagram of the regenerative modulator. (Courtesy *J. Acoust. Soc. Amer.*)

selects only difference frequency components. Because of balance, no feedback develops unless an input signal is applied, and the output frequency is one-half the input frequency. In practice, a dynamic range of 35 db can be obtained with the circuit. In order that the circuit be stable, it is necessary to tune the carrier feedback amplifier broadly to the output frequency.

If two closely spaced frequencies are impressed on the regenerative modulator, the average frequency of the input signal is halved, while the difference frequency between the two components is preserved. For more com-

[8] B. P. Bogert, "The vobanc—a two-to-one speech bandwidth reduction system," *J. Acoust. Soc. Amer.*, vol. 28, pp. 399–404; 1956.
[9] R. L. Miller, "Fractional-frequency generators utilizing regenerative modulation," Proc. IRE, vol. 27, pp. 446–457; July, 1939.

World Radio History

plicated speech signals, the frequency of the strongest components appears to be halved with the surrounding components displaced downward by the same amount. The spacing between harmonic components of the speech signal remain the same in the process, but the range of formant variation is halved. Thus, at the output of the regenerative modulator the speech formant range can be included within a bandwidth one-half that of the *A* filters. The filters which select the half frequency components are designated the *B* filters. Their transmission characteristics are shown in Fig. 9(b). The frequencies passed by the *B* filters are now modulated down to the frequency range 75 to 1925 cps by the output mixers of the transmitting terminal.

nant articulation scores for the system range between 79 and 91 per cent, depending on the degree of learning involved. Compared to this, the score on the conventional 3500-cps bandwidth telephone is 80 to 95 per cent.

Continuous Analysis-Synthesis Methods

Vocoder

The vocoder (*Voice Coder*) is very likely the most widely used speech bandwidth compression system in use today. It was first proposed and successfully demonstrated by Dudley[10] in 1939. Except for changes in the circuitry details, the functional concept of the device has changed little from that originally conceived. A

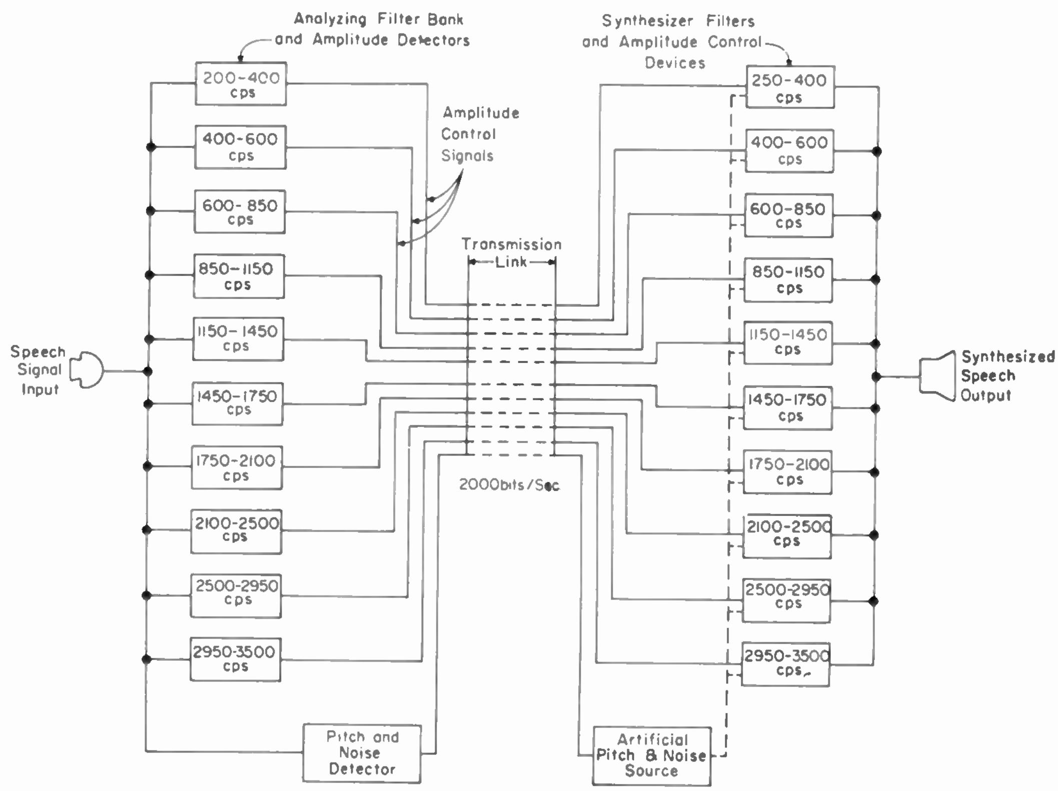


Fig. 11—Block diagram of the vocoder.

At the receiving terminal the compressed bandwidth speech signal is again modulated up to the frequency range of the *B* filters and supplied to a set of *B* filters. Each of the *B* filters outputs is then doubled in frequency, summed, and filtered to restore the 108-kc carrier range. The resulting signal is then mixed with a 108-kc local oscillator to restore the signal to the original audio frequency range

In actual use, the frequency bandwidth is not quite halved since some guard-band must be allowed between the three channels. The most noticeable distortion introduced by the Vobanc is the "burbles," which may be produced on fricative speech sounds, breathing and other background noises. This distortion is more of an annoyance than a restriction on intelligibility. Conso-

block diagram of the vocoder is shown in Fig. 11. It consists basically of a speech analyzer located at the transmitting terminal and a synthesizer at the receiving terminal. In the analyzer, the speech spectrum is divided into a number of contiguous bands by an analyzing filter bank. It has been a general practice to identify a vocoder in terms of the number of anlayzing filters. Thus, the device shown in Fig. 11 is a 10-channel vocoder. The signals transmitted by each filter are detected to determine the amplitude of the signal level in each band. These signals are then transmitted to the synthesizer, and there are used to enable artificial voice

[10] H. Dudley, "Remaking speech," *J. Acoust. Soc. Amer.*, vol. 11, pp. 169–177; October, 1939.

World Radio History

and noise (unvoiced) excitation falling in a band corresponding to that from which they were originally detected.

A pitch control channel is also included to control the artificial voice excitation in the synthesizer. The pitch information is derived from the speech signal in the analyzer by a pitch extractor circuit. This signal not only controls the frequency of pitch of the synthesized speech, but also controls the selection of voice excitation for the vowel sounds and noise excitation for the fricative sounds. The pitch control signal can be used to control voice-unvoice selection because of its its nature. It takes on values above a certain threshold for voiced sounds, but remains at a steady state value below the threshold for silence and fricative sounds.

The method of allocating the analyzing filters is of interest. It is generally the practice to select the bandwidths of the filters so that they increase logarithmically with increasing frequency. This arrangement, called a Koenig scale, keeps the amount of voice energy intercepted by each filter roughly equal.

It is possible to transmit each of the amplitude control signals and the pitch control signal in analog form over a channel of approximately 25 cps bandwidth. Hence, the total bandwidth required for the 10-channel vocoder is 275 cps. For binary coding it is possible to encode each channel in a 4 bit code employing a sampling rate of 50 cps. The entire information output of the 10 channel vocoder can therefore be transmitted at approximately 2200 bits per second.

For a 10 channel vocoder, Swaffield[11] has reported syllable articulation in the order of 83 to 85 per cent, as compared to a value of 90 to 91 per cent for a high quality voice circuit degraded by restricting its bandwidth to 250 cps to 3000 cps. It can be expected that a vocoder using a larger number of analyzing filters (up to 18) will exhibit improved performance over that achieved for the 10 channel unit reported above; however the performance will be improved at the expense of increased bandwidth and information rate.

Formant Tracking Systems

Formant tracking speech compression systems exploit the fact that all of the vowel sounds occurring in speech can be specified uniquely in terms of the relative frequency positions of the first two principal formants encountered in the speech signal spectrum, and that the specification is improved by including the third formant.[1,12] Typical formant structure of speech is illustrated in the sonagrams shown in Fig. 12. For male spoken vowels, the first formant can range from $200 < F1 < 1000$

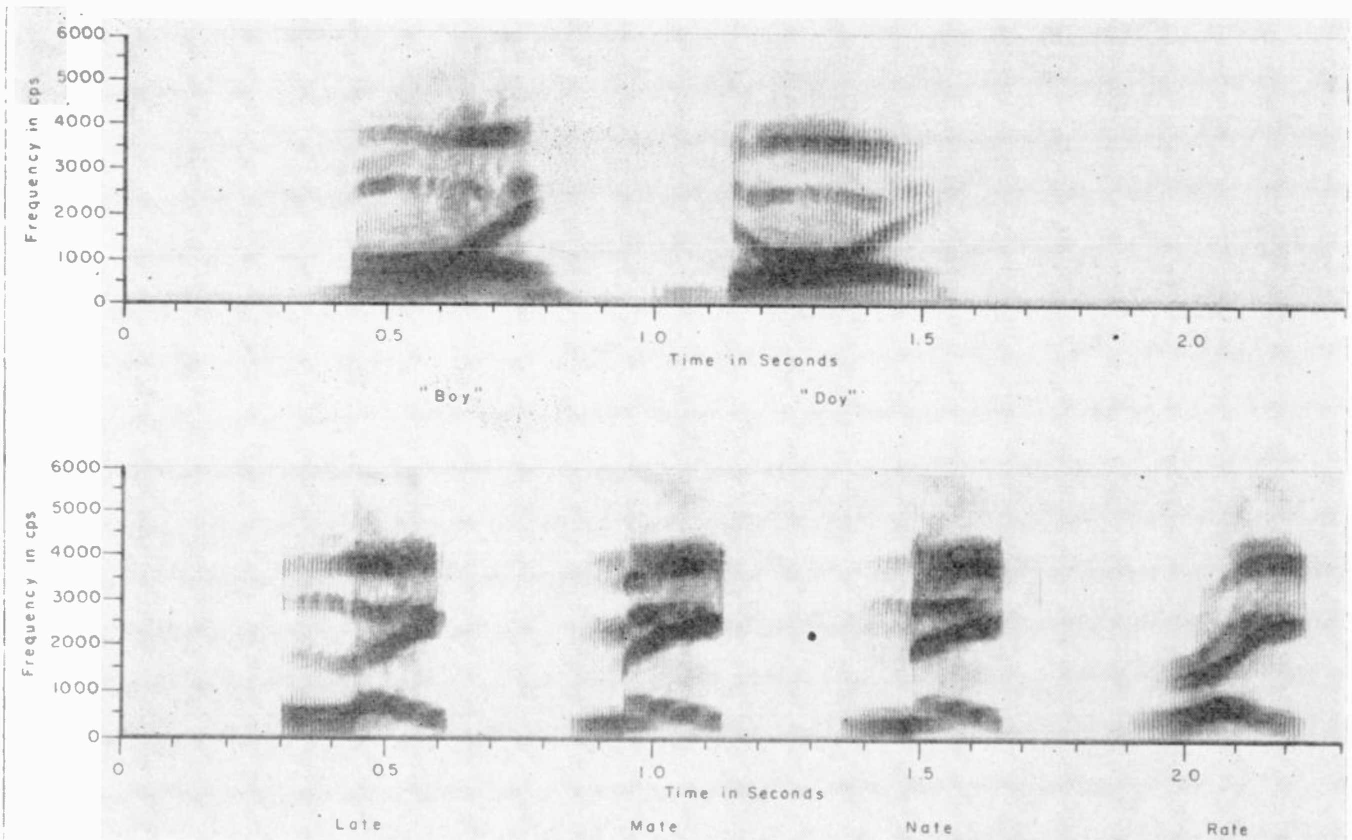


Fig. 12—Typical speech sonagrams.

[11] J. Swaffield, "The potentialities of the vocoder for telephone over very long distance," *P.O. Elect. Eng. J.*, vol. 41, pt. 1, pp. 22–28; April, 1948.

[12] P. C. Delattre, A. M. Liberman, and F. S. Cooper, "Two-formant synthetic vowels and cardinal vowels," *Le Maitre Phonetique;* July–December, 1951.

World Radio History

cps, the second, $800 < F2 < 2300$ cps, and the third, $2300 < F3 < 3800$ cps. Spectral distributions for unvoiced sounds also show similar formant structure. Although, the third formant is not required to successfully specify and synthesize vowel sounds, it appears to be essential to the specification and synthesis of the unvoiced sounds. This fact is apparent from the study of sonagrams of fricative sounds. As for the synthesis of consonants, both bursts of noise—occupying definite frequency positions—and formant transitions constitute important identification cues.[13]

Several speech bandwidth compression systems based on this principle have been constructed and successfully demonstrated by Campanella and Bayston,[15] Chang,[16] and Lawrence.[17] One typical system is illustrated in Fig. 13. The analyzing device contains three formant trackers, a low-frequency energy detector (less than 1000 cps), a high-frequency energy detector (greater than 1000 cps), and a pitch extractor. For the purpose of demonstrating the system, these six signals are transmitted in analog form to a synthesizer via 25 cps low-pass filters. In a communication system these signals would be either time division or frequency division multiplexed on a carrier, or converted to binary code for transmission.

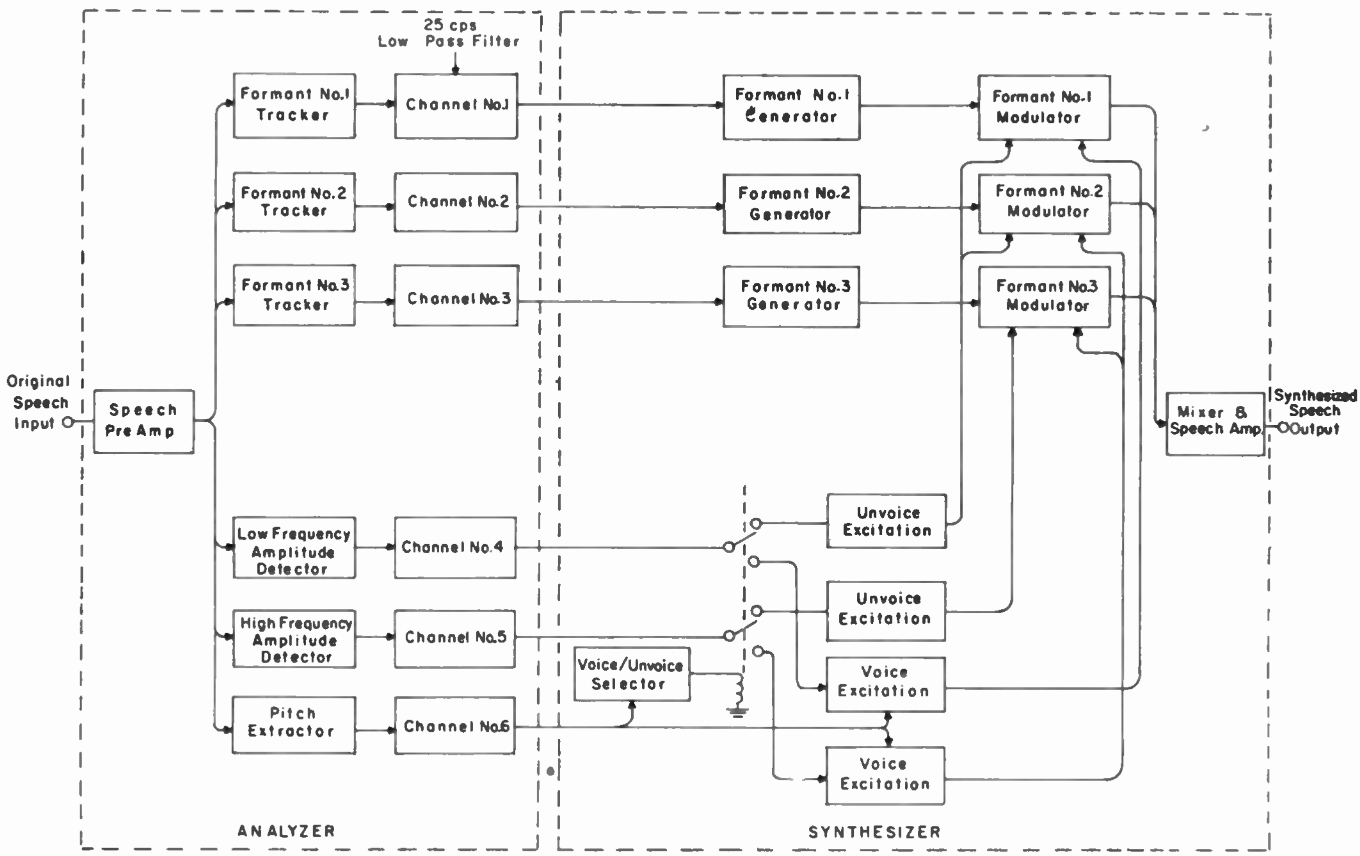


Fig. 13—Parametric speech bandwidth compression system.

From these facts it appears reasonable to expect that the speech signal can be specified in terms of a small number of control signals or parameters and accurately reconstructred from such information. Lawrence demonstrated the reconstruction of speech from a set of six parameters consisting of three formants—voice amplitude, unvoiced amplitude and pitch.[14] Each of these signals can be conservatively passed through low-pass filters of 25 cps bandwidth, thus making it possible to transmit speech in a total bandwidth of 150 cps.

At the synthesizer a heterodyne process is used to generate the artificial speech. Each of the formant control signals displaces a formant generator oscillator from a carrier of 18 kc by an amount proportional to the formant frequency. The first formant generator has a range from 18.2 to 19.0 kc, the second from 18.8 to 20.4 kc, and the third from 20.3 to 21.8 kc. These im-

[13] A. M. Liberman, P. C. Delattre, F. S. Cooper, and L. J. Gerstman, "The role of consonant-vowel transitions in the perception of the stop and nasal consonants," *Psyc. Monographs General and Appl.*, vol. 68, pp. 1–13; 1954.

[14] W. Lawrence, "The synthesis of speech from signals which have low information rate," in "Communication Theory," W. Jackson, ed., Butterworth Scientific Publications, London, Eng., ch. 34; 1953.

[15] S. J. Campanella and T. E. Bayston, "A continuous analysis speech bandwidth compression system," *Third Annual Aero-Com Symposium Abstracts*, pp. 10–12; November, 1957.

[16] S. Chang, "Two schemes of speech compression system," *J. Acoust. Soc. Amer.*, vol. 28, pp. 565–572; July, 1956.

[17] Signals Research and Development Establishment Technical Summary No. 24, SRDE Speech Bandwidth Compression Project; January 20, 1958.

World Radio History

pulses are supplied to balanced modulators, where they are mixed with sources of voiced or unvoiced excitation centered at 18 kc. The bandwidth of the sources of excitation determines the bandwidth of the synthesized formants and is in the range of 150 to 200 cps. The outputs of all the modulators are summed and the difference frequency components selected to produce the synthesized speech. The selection of voiced or unvoiced excitation is controlled by the pitch signal. The repetition frequency of the source of voiced excitation is of course controlled by the pitch signal. To permit independent control of the amplitude of excitation in the low and high frequency ranges of the synthesized speech, two sources are supplied for each type of excitation. Such independent control was found necessary to produce the nasal and fricative consonants.

Articulation tests conducted on an experimental model of the system described above gave an average score of 67 per cent for PB (phonetically balanced) word lists.[18] It is expected that considerably improved performance will be achieved in future models of the system.

The total information rate required to accommodate the six control signals has been estimated to lie between 640 and 1200 bits per second. These estimates are based on quantizing each parametric control signal with three or four bits (*i.e.*, with 8 or 16 quantized levels) and sampling at a rate of 40 or 50 cps. The resulting quantized levels of each control signal produce changes in the parameter under control which are in the neighborhood of those just discernible by the human ear.[19] The upper estimate in information rate of 1200 bits per second assumes four-bit quantization at a sampling rate of 50 cps. The lower estimate of 640 bits per second is obtained by considering three-bit quantization for all parameters with the exception of the pitch, using one bit for the pitch (this will give a monotone quality to the synthesized speech) and sampling at 40 cps.

Discrete Analysis-Synthesis Methods

Spectrum Pattern Quantization

For all of the sounds of speech, it is well known that each is characterized by a given shape of its spectrum. If one could specify a set of shapes which uniquely define all of the basic speech spectrum patterns both for steady state and transient sounds, it would be possible, provided this set is of reasonable size, to transmit the speech spectrum at relatively low information rates.

A technique to exploit this possibility has been proposed by Smith.[20] A block diagram of the method is shown in Fig. 14. At the transmitting terminal, the output channels of a conventional vocoder are supplied to a pattern correlation matrix. It is the function of this matrix to select the stored spectrum pattern most closely correlated with the input spectrum and to indicate this selection at its output. A 4×5 pattern correlation matrix for performing the pattern selection is shown in Fig. 15(a). The device consists of a matrix of cells. The columns of cells are connected to the spectrum amplitude outputs of a vocoder analyzing filter bank. The rows are connected to an amplitude quantizing resistance divider. When the inputs to any cell are within plus or minus one-half the difference between the quantized amplitude levels, the cell produces a unit of output voltage. Otherwise its output is zero. In order to normalize the amplitude levels, the voltage reference for

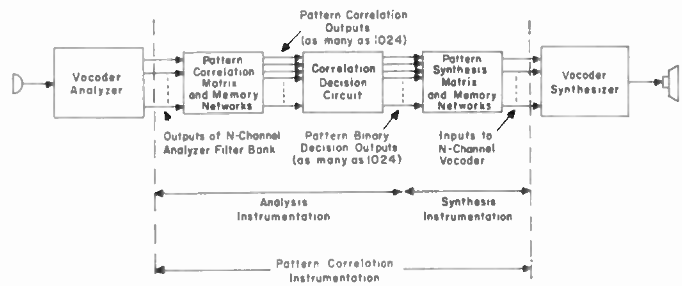


Fig. 14—Correlation instrumentation for the vocoder.

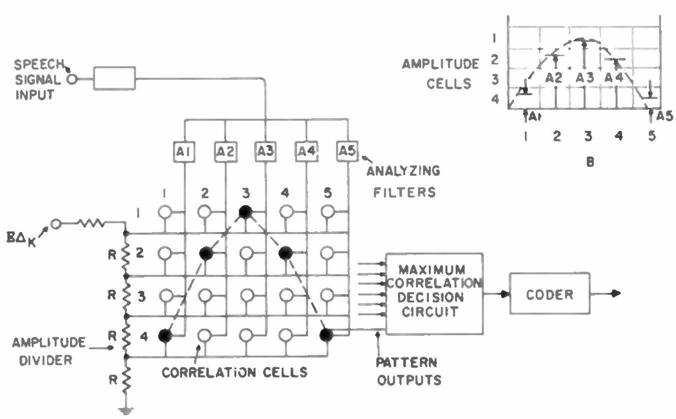


Fig. 15—Pattern correlation matrix.

the amplitude quantizing is determined by the average of all the vocoder outputs. The result of the amplitude and frequency quantization in a sample spectrum is illustrated in Fig. 15(b). In this case, cells giving unit output are 41, 22, 13, 24, and 45. If these cells are interconnected so that the output from each is added, the output exceeds that of all other possible interconnections by at least one unit. Thus, if many pattern outputs are laced through the matrix, the most probable pattern existing at any instant will be that corresponding to the interconnection of cells having the greatest output voltage. For each pattern selection a code can be assigned and transmitted to the receiving terminal. At the receiving terminal a corresponding pattern can be se-

[18] "Development of a Continuous Analysis Speech Compression System," Final Eng. Rep., Contract AF33(600)-32293; July, 1952.

[19] J. L. Flanagan, "A Speech Analyzer for a Formant-Coding Compression System," Sci. Rep. No. 4, USAF Contract No. AF19(604)-626; May, 1955.

[20] C. P. Smith, "Speech Data Reduction," AFCRC-TR-57-111, ASTIA Document No. AD117290; May, 1957.

World Radio History

lected from a set of stored patterns and used to control a conventional vocoder synthesizer.

It has been estimated that the total information rate necessary to transmit speech information in this form would be less than 1000 bits per second.[20] An 11-bit code could be used to specify a set of 2048 spectrum patterns. If the sampling rate were 50 per second, then 550 bits per second would be required to transmit these patterns. An additional 450 bits per second would suffice for pitch and amplitude, thus giving a total of 1000 bits per second.

Phoneme Quantizing

As far as speech compression is concerned, phoneme quantizing would most probably provide the least possible information rate for transmitting the information content of speech. Using such a technique, speech information could be transmitted at teletype code rates. Approximately 40 phonemes constitute the sounds of the English language, and they occur at rates of approximately 10 per second.[21] A 6-bit code would suffice to specify 64 different phonemes. Thus, the phoneme content of speech can theoretically be transmitted at an information rate of 60 bits per second. However, this information rate does not allow for inflections in pitch or changes in intensity of the synthesized speech.

A block diagram of a possible phoneme coded system is shown in Fig. 16. To this date no one has successfully demonstrated a speech compression device based on phoneme code, although a synthesizer has been demonstrated which, when properly programmed, could produce recognizable speech from a set of phoneme sounds recorded on magnetic drums. The greatest difficulty in realizing the device is in the analysis of the original speech. Although only 40 phonemes may be required to specify the speech, each individual speaker imposes his own unique characteristics on the spectrum, and it becomes difficult to produce a device which indicates the correct phoneme selection invariant of the characteristics of the individual speaker. Perhaps with improved analysis devices associated with appropriate spectrum normalization to remove speaker-to-speaker variances, a phoneme coded system may become a reality.

Sound Group Analysis-Synthesis Methods

For a limited vocabulary, a system which automatically recognizes entire words, transmits a unique code for each word, and at the receiver converts the word code to synthetic speech may offer a means of transmission at extremely low information rates. For example, a vocabulary of 32 words can be handled at information rates of 10 bits per second. No attempt has been made to produce such a system; however, both analyzers and synthesizers are being separately investigated.

The telephone company is working on an analyzer which will automatically recognize spoken digits. This device has been given the name of Audrey[22] (*Au*tomatic *D*igit *Re*cognizer). The machine analyzes the speech input to determine which sound in its memory it is most like. It first breaks the spoken digit into a series of sound identifications and then determines by comparison which of the ten digits in its memory has the same sequence. The device is able to recognize the digits as spoken by certain voices; however, it cannot as yet perform satisfactorily for all voices. It is pointed out that the sound recognition portion of the Audrey system is operating quite successfully, and may constitute a useful analyzing device for phoneme coded speech compression system.

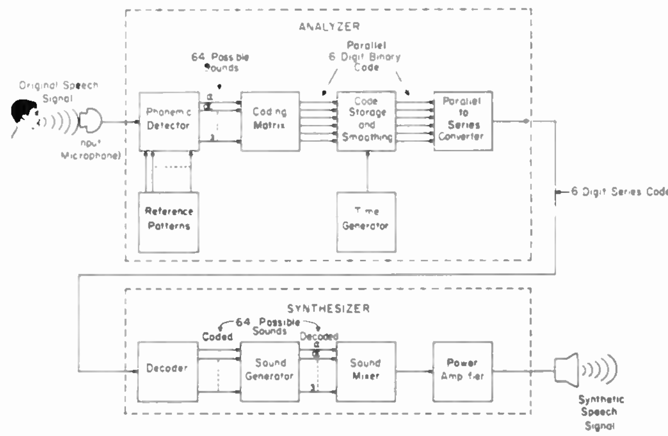


Fig. 16—Discrete phoneme coding system.

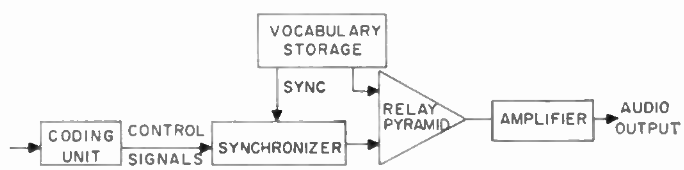


Fig. 17—Discrete word coding system.

For the synthesis of complete words, magnetic tape playbacks operated from a source of digital information provides a satisfactory device. The Automatic Voice Readout System (AVRS)[23] shown in Fig. 17 is such a device. The AVRS is intended for use as a digital code to voice converter to read out commands from a digital computer vocally. A present model of the device has a capacity of 35 words and has a potential capacity of 100 words. Five digit control signals required to control a relay pyramid are supplied from a coding unit and synthesizer. The output of the relay pyramid drives an audio amplifier. The synchronizer serves as the basic timing unit within the system. Synchronization pulses are received every drum revolution or one-half second,

[21] R. J. Halsey and J. Swaffield, "Analysis-synthesis telephony with special reference to the vocoder," *J. IEE*, vol. 95, pp. 391–406; September, 1948.

[22] W. E. Kock, "Speech bandwidth compression," *Bell Labs. Reps.*, pp. 81–85; March, 1956.

[23] C. W. Poppe and P. J. Suhr, "An automatic voice readout system," *Proc. Eastern Joint Computer Conf.*, pp. 219–221; December, 1957.

World Radio History

and are used to advance the synchronizer to the next word interval. The function of the coding unit is to convert the input code to a five digit parallel code required to drive the relay pyramid.

A system consisting of a limited vocabulary coding analyzer such as Audrey, and limited vocabulary decoding synthesizer such as the AVRS would, when operated with both parts, constitute an operable voice communication system of extremely low information rate. Although such a system possesses a highly limited vocabulary, there are certain situations such as aircraft traffic control, where such a limited vocabulary would be entirely satisfactory.

Signal-To-Noise Ratio in Compressed Speech Channels

The amount of information which can be transmitted over a channel per unit time is given by (1) in terms of bandwidth and signal-to-noise ratio. By use of this expression, it is possible to relate the signal-to-noise in the compressed speech channel to the ratio of bandwidth reduction, the ratio of information reduction and the signal-to-noise ratio in the original speech channel. If $S1/N1$, $W1$, and $C1$ are the signal-to-noise ratio, bandwidth, and channel capacity, respectively, of the original speech channel, and if the corresponding parameters for the compressed speech channel are indicated by a subscript 2, then the signal-to-noise ratio in the compressed speech channel is given by the relation

$$S2/N2 = [1 + (S1/N1)]^{(W1/W2)(C2/C1)} - 1; \qquad (4)$$

and for $S2/N2 \gg 1$ and $S1/N1 \gg 1$ the following approximation is valid:

$$S2/N2 = (S1/N1)^{(W1/W2)(C2/C1)}. \qquad (5)$$

In Fig. 18 the compressed speech channel signal-to-noise ratio is plotted as a function of the signal-to-noise ratio in the noncompressed speech channel for values of the exponent $(W1/W2)(C2/C1)$ of 1 and 10. The value of unity corresponds to the case where, by elimination of redundancy, inefficiency and/or speaker identity, the channel capacity is reduced by the same factor as the channel bandwidth. To obtain comparable performance in this case the signal-to-noise ratio in the compressed speech channel will be the same as that in the noncompressed speech channel. Furthermore, since the bandwidth required to accommodate the compressed speech channel is reduced by the factor $(W1/W2)$, the interferring noise energy intercepted is reduced by the same factor (assuming that the noise is uniformly distributed in the frequency spectrum) and the immunity of the compressed speech channel to noise interference is improved by $10 \log (W1/W2)$ db.

Values of the exponent $(W1/W2)(C2/C1) > 1$ result when the channel capacity is not reduced by a factor as great as that for the reduction in bandwidth. In this case, the signal-to-noise ratio in the compressed speech channel will always be greater than that in the noncompressed speech channel, if comparable performance is to be maintained. Consider for example a case in which the value of the exponent is 10. This could result if an attempt were made to reduce the bandwidth by a factor $W1/W2 = 10$, but no attempt were made to eliminate redundancy, inefficiency and/or speaker identity to reduce channel capacity, *i.e.*, $C1/C2 = 1$. As shown in Fig. 18, the signal-to-noise ratio in the compressed speech channel must exceed that in the noncompressed speech channel by a considerable amount if comparable performance is to be achieved. For example, to obtain performance comparable to that achieved with a 20 db signal-to-noise ratio in the non-compressed speech channel, the signal-to-noise ratio in the compressed speech channel would have to be 200 db.

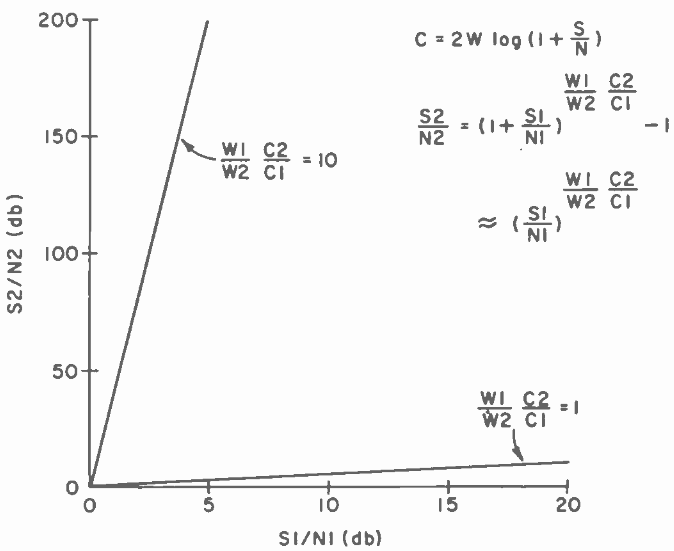


Fig. 18—Compressed speech channel signal to noise ratio as a function of original channel signal to noise ratio.

The preceding discussion points out the rather interesting fact that speech bandwidth compression without corresponding elimination of information can be accomplished only at the cost of considerable transmitted power. In this respect it is pointed out that the true measure of the effectiveness of a bandwidth compression system cannot be measured by the bandwidth reduction factor alone; the influence of information reduction must also enter the picture in terms of the signal-to-noise ratio, or the information rate that must exist in the compressed speech channel to obtain speech reproduction with good signal-to-noise ratio.

Conclusions

There are several methods of speech bandwidth compression which are currently being investigated or used today. These methods may be divided into four categories:

1) frequency or time compression,
2) continuous analysis-synthesis,
3) discrete sound analysis-synthesis, and
4) sound group analysis-synthesis.

World Radio History

Systems in the first category are the most simple to implement and can provide bandwidth compression ratios of 1:4 to 1:6 with corresponding reductions in information rate. Systems in the second category are considerably more complicated and can provide bandwidth compression ratios in the order of 1:10 to 1:20, using information rates in the range from 800 to 3000 bits per second. Systems in the third category have not as yet been successfully demonstrated. They promise to provide speech communication at information rates below 1,000 bits per second, and perhaps as low as 60 bits per second. The lower bit rates would be achieved by use of a phoneme coding and would possess no speaker identity cues. Systems in the fourth category may prove useful with a limited vocabulary of perhaps less than 100 words. Such systems would employ extremely low bit rates and would probably be used only in highly specialized applications. It is not expected that they would be useful for general speech communication.

It is important to note that bandwidth reduction without elimination of unnecessary information can be accomplished only by an excessive increase in the signal-to-noise ratio of the compressed bandwidth channel. Hence, in specifying the performance of a speech bandwidth compression system, either the signal-to-noise ratio or information rate required to provide satisfactory speech reproduction should be indicated.

Contributors

John P. Quitter (S'40–A'43–M'44–SM'54) was born in Budapest, Hungary, July 24, 1919. He received the E.E. degree from the University of Cincinnati, Ohio, in 1941, and has done graduate work in physics.

J. P. Quitter

In 1941 and 1942 he was associated with the Westinghouse Electronics Division in the development of magnetrons and high-frequency triodes. From 1943 to 1948 he was with Sperti, Inc., of Cincinnati in the development and manufacture of various instruments and armament devices.

In 1948 he joined the Baldwin Piano Company as research engineer engaged in acoustical study of the piano, and is now supervisor of the Piano Laboratory. His work is primarily in the fields of vibration problems, audio instrumentation, electronic counterparts of the piano, and application of modern engineering principles to the design and manufacture of the piano. Mr. Quitter is also mathematics instructor at the University of Cincinnati Evening College.

He is a registered professional engineer and a member of AIEE, Eta Kappa Nu, and the Engineering Society of Cincinnati. Mr. Quitter has served as Cincinnati Chairman of IRE and AIEE, and of the Technical and Scientific Societies Council of Cincinnati. He is presently secretary of the Cincinnati PGA.

❖

S. Joseph Campanella (A'52–M'57) was born in Washington, D. C., on December 26, 1926. He attended the Catholic University of America in Washington, where he received the B.E.E. degree (magna cum laude) in 1950, and the University of Maryland, College Park, Md., where he received the M.S.E.E. degree in 1956.

In 1950, he joined the Naval Research Laboratory in Washington as an electronic scientist. He participated in development of various experimental sonar systems and designed and developed sonar receiving systems and correlation devices. Since 1953 he has been at Melpar, Inc., where he has performed fire control systems study, investigated computer techniques for statistical data processing, designed and developed special purpose power spectrum analysis equipment, conducted research and development in speech bandwidth compression techniques for use in communications systems, and performed communications system analysis and design.

S. J. Campanella

He is a member of Sigma Xi, Phi Eta Sigma, and AOA.

World Radio History

INSTITUTIONAL LISTINGS

The IRE Professional Group on Audio is grateful for the assistance given by the firms listed below, and invites application for Institutional Listing from other firms interested in Audio Technology.

ALLIED RADIO CORPORATION, 100 N. Western Ave., Chicago 80, Illinois
Everything in Radio, Television, and Industrial Electronics

BALLANTINE LABORATORIES, INC., Fanny Rd., Boonton, New Jersey
Electronic Voltmeters, Decade Amplifiers, Voltage Calibrators, Multipliers, Shunts

JENSEN MANUFACTURING COMPANY, 6601 South Laramie Ave., Chicago 38, Illinois
Loudspeakers, Reproducer Systems, Enclosures

KNOWLES ELECTRONICS, INC., 9400 Belmont Ave., Franklin Park, Illinois
Miniature Microphones and Receivers, Special Recorder and Audio Devices

JAMES B. LANSING SOUND, INC., 3249 Casitas Ave., Los Angeles 39, California
Loudspeakers and Transducers of All Types

STROMBERG-CARLSON CO., SPECIAL PRODUCTS DIV., 1400 N. Goodman St., Rochester 3, N.Y.
High-Fidelity Equipment, Intercommunication, Public Address and Sound Systems

UNITED TRANSFORMER COMPANY, 150 Varick St., New York, New York
Manufacturers of Transformers, Filters, Chokes, Reactors

UNIVERSITY LOUDSPEAKERS, INC., 80 South Kensico Ave., White Plains, New York
High Fidelity, Commercial-Industrial, and Military Speakers and Accessories

Charge for listing in six consecutive issues of the TRANSACTIONS—$75.00.
Application for listing may be made to the Technical Secretary, Institute of Radio Engineers, Inc., 1 East 79th Street, New York 21, N.Y.

World Radio History