UNIVERSAL
LIBRARY

Speech and Hearing

By

HARVEY FLETCHER, PH.D.

Acoustical Research Director

BELL TELEPHONE LABORATORIES, INC.

WITH AN INTRODUCTION

BY

H. D. ARNOLD, PH.D.

Director of Research

BELL TELEPHONE LABORATORIES, INC.

MACMILLAN AND CO., LIMITED

ST. MARTIN'S STREET, LONDON

1929

PRINTED IN U. S. A.

PRESS OF
BRAUNWORTH & CO., INC.
BOOK MANUFACTURERS
BROOKLYN, NEW YORK

PREFACE

Some fifteen years ago the Research Laboratories of the Bell Telephone System undertook a comprehensive survey of speech and hearing to obtain the fundamental facts on which to base the design of apparatus and systems for telephone use. The art of analyzing electrical currents into their component frequencies and accurately measuring them, and the correlated art of describing and measuring the characteristics of mechanically vibrating systems, had reached an advanced state of development. It was apparent that great advantages would come from similarly analyzing speech and hearing, for if we could accurately describe every part of the system from the voice through the telephone instruments to and including the ear, we could engineer the parts at our disposal with greater intelligence.

A greater plan for a prolonged laboratory investigation was evolved and has been in operation since that time. The attack was first launched most vigorously on the constitution of speech in an effort to establish a reasonable description of average speech, and to find to what extent small imperfections and variations in speech affected intelligibility. It is obvious that this work could not go far without entailing a study of the organs of speech and the organs and mechanisms of hearing, and its scope came to be extended as well to some of the abnormalities in these faculties.

As the work progressed it became apparent that better and more precise instruments must be developed than were available, and a considerable part of the effort has been devoted to the matter of securing devices which would convert sound waves into electrical form and reconvert them again to sound with the least possible distortion. Out of this have come unexpected rewards to the telephone and phonograph arts, for as these devices were perfected they found very immediate application to the great advantage of those industries.

One of the most difficult phases of the investigation has been that relating to the degree of precision with which the mind can differentiate and interpret sounds that are very nearly alike. This does not lend itself so readily to analysis and measurement as does the purely mechanical operation of the ear itself. The approach to this problem has been through the use of essentially perfect reproduction systems which could be deteriorated step by step until their faults became noticeable to the observer. This set a limit to the degree of perfection which could ever be demanded in the apparatus. When the deterioration was carried somewhat further an estimate could be obtained of the degree of dissatisfaction presented by certain measured imperfections, and hence a practical basis of choice of a reasonably perfect system could be established.

As this large program was undertaken with definite idea of meeting and solving whatever difficulties might arise in its progress, it is not surprising that it has brought us a wealth of new and useful information on the fundamentals of speech and hearing. Although of course the nature of this information could not be predicted in advance, it is a foregone conclusion that no such persistent and thorough-going study can be carried through without large additions to the philosophy of the subject. The principal unforeseen and unexpected return has been found in the application of the unusually perfect devices developed in the course of the experimentation to practical commercial problems. Some such unlooked for applications will naturally arise from any thorough-going research, but the magnitude of the results that were obtained by adapting the devices of these experiments to the phonograph, to radio, and to the problems of the deafened could hardly have been visioned by the most ardent protagonist of research as a speculative enterprise.

The work as originally planned is far from finished; and as it has progressed new problems have arisen that bid fair to occupy us for years to come. Meanwhile, however, part of it has already been the inspiration of a book by Dr. I. B. Crandall on "Vibrating Systems and Sound." Now another part is crystallized in the present volume, taking the general form in which it was presented in Bell Laboratories' program of "out-of-hour courses."

The author desires to acknowledge the brilliant work of the large number of the staff of Bell Telephone Laboratories who have contributed to the work which he reports, to recognize the helpful criticisms and suggestions offered by the members of the staff of the American Telephone and Telegraph Company who have been following it; to thank many of his colleagues who have greatly aided the preparation of the material for this book; and to express his sincere gratitude to Dr. Arnold for his constant inspiration and sympathetic understanding of the many intricate problems that have arisen during the progress of the research work.

HARVEY FLETCHER.

NEW YORK,
December 1, 1928.

CONTENTS

Introduction xi

CHAP. PAGE

PART ONE—SPEECH

I. Mechanism of Speaking 3
II. Characteristics of Speech Waves 14
III. Speech Power 64
IV. Frequency of Occurrence of the Different Speech Sounds 81

PART TWO—MUSIC AND NOISE

I. Physical Properties of Musical Sounds 87
II. Noise 99

PART THREE—HEARING

I. Mechanism of Hearing 111
II. Limits of Audition 132
III. Minimum Perceptible Differences in Sound 145
IV. Masking Effects 167
V. Binaural Beats 188
VI. Methods of Testing the Acuity of Hearing 189

PART FOUR—THE PERCEPTION OF SPEECH AND MUSIC

I. The Loudness of Sounds 225
II. The Recognition of the Pitch of Musical Tones 245
III. Methods of Measuring the Recognition of Speech Sounds 255
IV. Effect of Changes in the Received Intensity of Speech Sounds upon Their Recognition 270
V. Effect of Frequency Distortion upon the Recognition of Speech Sounds 279
VI. Effect of Other Types of Distortion upon the Recognition of Speech Sounds 290
VII. Effect of Noise and Deafness upon the Recognition of Speech Sounds 297

INTRODUCTION

The atmosphere of sounds in which we live ministers so constantly to our knowledge and enjoyment of our surroundings that through long familiarity we have come to feel, if not contempt, at least indifference toward the marvelous mechanism through which it works. Hearing, we are inclined to consider as little a matter for concern as breathing; and so long as our own faculty remains unimpaired we feel little curiosity concerning the provisions of nature either for ourselves or for others. When we hear too faintly or indistinctly we know we need only trace the sound to its source to hear its perfect form, for that is the method we have used from childhood in investigating the sounds of our immediate neighborhood.

Now with one broad sweep the barriers of time and space are gone and all the world becomes our vocal neighborhood. No longer can we transport ourselves to the origin of a sound and thus become convinced that we are hearing it aright, for that origin may be thousands of miles away or may have vanished years before; and so we must establish a new method to measure the accuracy of the copy which reaches our ears. We must also find a clearer index to our satisfaction in it, for we are no longer concerned with the immutable provisions of nature but may approach at corresponding expense whatever perfection we may demand in our instruments of translation and reproduction. Thus the telephone and the phonograph should excite a keener interest in how we hear and in what measures our satisfaction in the speech and music which they provide.

Our ears are only machines to translate air waves into a form suited to stimulate the auditory nerve; and as machines we may measure and describe them in the same terms that apply to devices we ourselves construct. We may compare them as to performance, and may accommodate our devices to their requirements. But, to understand the mechanism of the ear is by no means to understand the act of hearing, for we have not heard until the brain has perceived the message sent by the auditory nerve. We cannot explain in precise mechanical terms how this is done, nor indeed have we any very clear comprehension of the process at present. Some important factors relating to the process of hearing we can, however, determine by measuring the least changes in sound which can be detected under a variety of conditions of pitch, loudness, and accompanying noise. Thus we may obtain a quantitative means of comparing individuals in this respect, and establish a standard of average hearing.

There is a most important factor in hearing, however, which is much more difficult of analysis and measurement. This is the individual's ability to recognize small defects in those sounds with which he has become especially familiar. We all know how quickly we note a slight change in a friend's voice, and with what uncanny skill a trained musician will detect minute imperfections in very complex sounds. Our approach to a quantitative understanding of the importance of this must be by an indirect method. First we must construct devices so perfect that even the keenest ear cannot find a flaw in their rendition, and then step by step we may introduce measured imperfections until an observer can detect a fault. In the response of individuals to this test there will, of course, be great differences; but when we have collated opinions from a wide variety of observers we may forecast in a reasonable way the degree of mechanical perfection that may be demanded of our instruments.

This, then, has been the philosophy of the investigation of hearing which has been carried on in Bell Telephone Laboratories during the past fifteen years: to get an accurate physical description and a measure of the mechanical operation of human ears in such terms that we may relate them directly to our electrical and acoustical instruments; to test the keenness of the sound-discriminating sense and find what is the smallest distortion which the mind can perceive and how it reacts to somewhat larger distortions; and thus to reach a reasonable basis of design both for separate instruments and for systems, as a whole, to give a proper balance between cost and performance.

With hearing, speech and music are linked inseparably for they only bring a meaning through our aural sense. It is an instinctive first thought that they must be heard to be criticized. They can, nevertheless, be investigated by mechanical means and be described in the same physical terms that we use in describing hearing; and thus to an extent we may consider them both objectively. But if we attempt to divide the study between speech and music we come at once upon the difficulty that speech conveys information by intonation as well as by articulate syllables; and this makes it infeasible to set a definite boundary between them. A division, however, between vocal sounds and instrumental sounds proves more useful, for in the one case we are limited by our vocal organs which we must take as they are, while in the other we have a definite control and can adapt the nature and complexity of the sounds produced to conform to our sense of hearing and our musical appreciation. The investigation of speech and music has been governed by these general considerations. An attempt has been made to establish in definite terms the performance and limitation of the voice and, although so far in considerably less detail, to find the corresponding factors in instrumental music.

With a clear knowledge of the nature of the sounds that we must produce and the accuracy with which we must maintain their form, there remains the probiem of securing instruments which are sufficiently refined for the purpose. Instruments of remarkable precision are required in the conduct of the investigation, since if we are to measure the smallest detectable variations in sounds we must obviously use equipment which is capable of a degree of exactness beyond these small quantities. Such instruments would appear at first sight not to have much utility outside the laboratory, since they are costly and often complicated and difficult of adjustment.

It is interesting to note, however, that some of the instruments, in essentially their original laboratory form, have found other important uses. Indeed, a surprising number of modern acoustical accomplishments have come about through the use of slightly modified forms of the apparatus which was originally developed for these investigations. Modern phonographic records are produced with an electrical transmitter which was developed in the very early stages of these studies; and radio broadcasting has grown up around this same "microphone." The reproducing equipment of the modern phonograph and of the radio were predicated directly upon these investigations; and talking motion-pictures owe their success and much of their apparatus to this same source.

Although the results which relate to normal speech and hearing are naturally the most familiar and widely known, there have also been important outgrowths in the way of aids to those handicapped in one or the other of these faculties. In establishing the functioning of the average ear it was obviously necessary to investigate a large number of cases and among them some which departed rather widely from the average. For this study an instrument was devised, now known as the audiometer, which has put within the reach of all who need it the possibility of an accurate measure of their hearing. In quite analogous fashion there grew out of the investigation of the limits of hearing a better knowledge of ways to provide aids for those partially deaf; and it has even become possible to provide means of speech for some persons whose vocal chords are gone.

Valuable as these results are, economically the most important outcome of the work has been the increase of exact knowledge as to the requirements and limitations to be placed upon the transmission of speech in the telephone system. As time goes on there must be an evolution toward even greater perfection in those particular elements which are most important to intelligibility. The system is so large that the cost of such an evolution is immense and changes undertaken without an accurate knowledge of their value might lead to burdensome expenditures for disproportionate results; but, with the facts established by this investigation in hand, we can weigh any contemplated change and judge whether it is the one that offers most improvement at the moment and what its ultimate effect will be in its joint operation with other elements of the system.

The work that Doctor Fletcher discusses drew at the start on all the acoustic knowledge available in the literature and during its progress every effort has been made to use to the best advantage the information found by other experimenters. For the most part, however, he describes experiments performed and conclusions reached in Bell Telephone Laboratories during investigations, captained in their early stage by Doctor Crandall and himself, for which since Doctor Crandall's death he has had the full responsibility. No one can speak with better knowledge of the facts or with more complete authority for the opinions which he expresses.

The work is not complete—indeed some parts of it are hardly more than started; yet its results have been so great, both for the original purpose which was planned and for the many issues which have since arisen, that it presents a unique exemplification of the worth of systematic and sustained research; and Doctor Fletcher is to be congratulated that he has seen it through with such clear vision as permits its presentation in its present form.

H. D. ARNOLD.

Part One

Speech

SPEECH AND HEARING

CHAPTER I

Mechanism of Speaking

Origin and Evolution of Language

When beginning a study of language some of the first questions which naturally arise are: "How is it that people everywhere do not speak the same language?", "How were words first created?", "Why is such-and-such a person and such-and-such a thing called this and not that?", and similar questions. According to some anthropologists, 150 separate languages which seemed to have no common origin were spoken among the American Indians. For this reason the Indians of one tribe communicated with those of another by means of signs and gestures. Thus, the so-called Indian sign language grew up. It is very probable that such signs, gestures, and expressions of the face were used before the evolution of the spoken language had progressed very far. According to some philologists, the vocal sounds of very primitive people were exclamatory and song-like and used mainly to express emotion. Sounds mimicking nature came to designate certain things connected with the thing imitated. As man's power of analysis developed, the sounds gradually developed into spoken words having definite meanings.

According to Sir Richard Paget,[1] human speech began by the performance of sequences of simple pantomimic gestures of the tongue, lips, etc., comparable with the natural gestures (of hands, etc.) which are still made by deaf mutes, and that these gestures were made audible by breathing or grunting.

[1] "The Origin of Speech, Sir Richard Paget, *Proc. Royal Soc.*, May, 1928, p. 157.

For example, consider the word "hither." The tongue makes the same beckoning gesture while speaking this word as is made with the hand.

Although there are a great many different languages spoken in different parts of the earth and each language has a system of speech sounds of its own, there is a great similarity among these fundamental speech sounds. This is necessarily true since there is only a limited range of distinct sounds that can be made by the organs of speech. Although the mechanism of producing particular speech sounds in the various languages is somewhat different, the general mechanism of producing speech is similar for all people.

Description of the Organs of Speech

The organs of speech are the lungs which by their bellows-like action supply the streams of air which pass in and out the

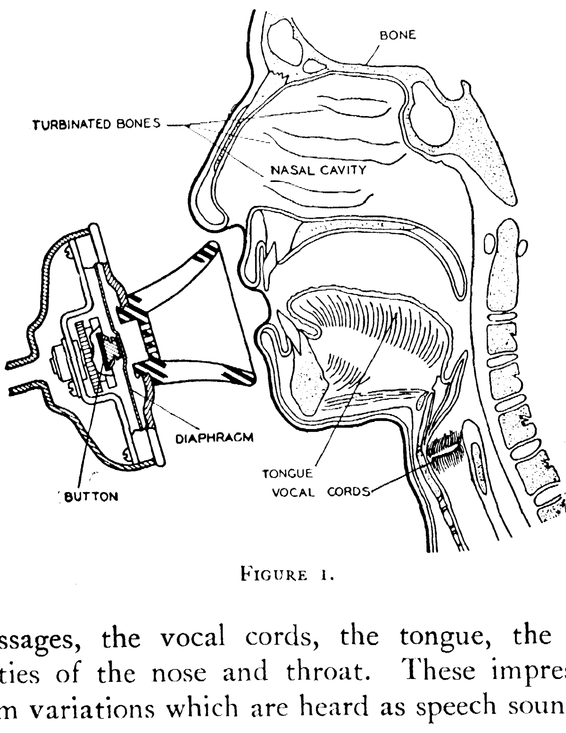


FIGURE 1.

vocal passages, the vocal cords, the tongue, the lips, and the cavities of the nose and throat. These impress on the air stream variations which are heard as speech sounds.

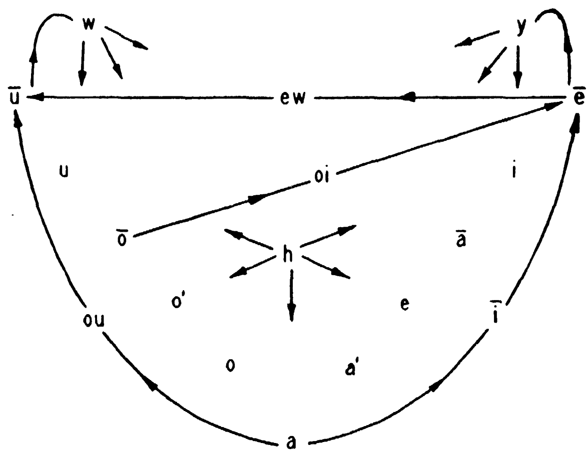
In Fig. 1 is shown the cross-section of the human head in the position to speak into a telephone transmitter. This shows the relative positions of the various organs of speech. The vocal cords are a pair of muscular ledges on both sides of the larynx forming a straight slit through which the breath passes. The vibration of these vocal cords starts a train of sound waves which pass through the vocal passages which impress on it certain resonant characteristics so that the vibrations finally emerge from the mouth as speech sounds. The pure vowels, the diphthongs, the transitionals, and the semivowels are produced in this manner. Other sounds called the unvoiced consonants are produced without using the vocal cords at all. They are produced by passing the air through small openings or over sharp edges in the mouth. There is a third class of speech sounds called the voiced consonants which are produced by a combination of the two processes just mentioned.

Description of the Formation of the English Speech Sounds

Different classifications of the spoken sounds of English may be made, depending upon the purpose one has in mind. The International Phonetic Association uses a basic alphabet of 65 different letters and also uses numerous modifiers which serve to distinguish several hundred different sounds. Such a system, of course, is altogether too complex for use in engineering work. The revised scientific alphabet, sometimes called the N. E. A. alphabet, uses 48 sounds. It is difficult for the average person to differentiate between some of these sounds. After considering the manner of formation of the speech sounds and studying their physical characteristics and the interpretation given by the average person, 39 speech sounds which can be readily distinguished by an average English-speaking person were chosen. These are the same sounds as were selected by the Simplified Spelling Board except that the sounds in the word t*o*n and the word nat*ion* were considered near enough alike to be designated by a single letter "o" and the sounds in the words p*a*rt, n*o*t, and f*a*ther were sufficiently similar to be designated by a single letter "a." These fundamental speech sounds are divided into six classes, namely, pure vowels, diphthongs, transitionals, semi-

TABLE I

CLASSIFICATION OF THE SPEECH SOUNDS



1. Pure Vowels—11
   Long —ū (tool), ō (tone), ó (talk), a (far), ā (tape), ē (team)
   Short—u (took), o (ton), á (tap), e (ten), i (tip)
2. Diphthongs—4
   ī, ou, oi, ew
3. Transitionals—3
   w, y, h
4. Semi-vowels—5
   l, r, m, n, ng
5. Fricative Consonants—8

Voiced	*Unvoiced*	*Formation of Air Outlet*
v	f	lip to teeth
z	s	teeth to teeth
th (then)	th (thin)	tongue to teeth
zh (azure)	sh	tongue to hard palate

6. Stop Consonants—8

Voiced	*Unvoiced*	*Formation of the Stop*
b	p	lip against lip
d	t	tongue against teeth
j	ch	tongue against hard palate
g	k	tongue against soft palate

vowels, fricative consonants, and stop consonants. The two classes of consonants are further divided into the groups designated voiced and unvoiced. The complete list is tabulated in Table I. The diacritical marks usually used are entirely too complicated for use in engineering work, so it will be noticed that only vertical and horizontal lines above the letters are used to indicate how the sounds should be pronounced. Such a simplification certainly makes it very much easier to write these sounds. At the top of Table I the pure vowels, the diphthongs and the transitionals are shown in a diagram which helps to illustrate the manner in which the sounds are formed. Starting with the sound ū, the lips are rounded and there is formed a large resonating cavity in the front part of the mouth and a smaller and less important one in the throat cavity. Passing down the left side of the triangle in the diagram from ū to a, the mouth is gradually opened with the tongue lowered to form the successive vowels. In all these vowels, the throat resonance is playing only a minor part. Going up the right side of the triangle from a to ē, the tongue is gradually raised to the front part of the mouth, thus forming two resonance chambers, both of which produce marked effects upon the sounds from the vocal cords.

In Fig. 2 are shown the tongue and the lip positions for forming these vowel sounds.[1] An infinite number of different shadings of these vowels may be produced by placing the mouth in the various intermediate positions, but the ones shown were chosen as being the most distinct. It is very difficult to define differences between vowels and consonants which are satisfactory in all cases. In general, however, pure vowels are characterized by continuous wave trains formed in the throat and passing through opened passages. Sounds at the lower end of the consonant list, on the other hand, represented by the unvoiced stop consonants, are characterized by a short group of waves formed mostly by the mouth and released suddenly by opening the lips. Between these two extremes is the graduated series of diphthongs, transitionals,

[1] For a further discussion see "Sounds of Spoken English," by Walter Ripman.

semi-vowels, voiced fricatives, unvoiced fricatives, and voiced stop consonants.

The diphthongs ī, ou, oi, and ew are really combinations of two of the pure vowels. If the mouth is placed in the position to say a and then changed over without interrupting

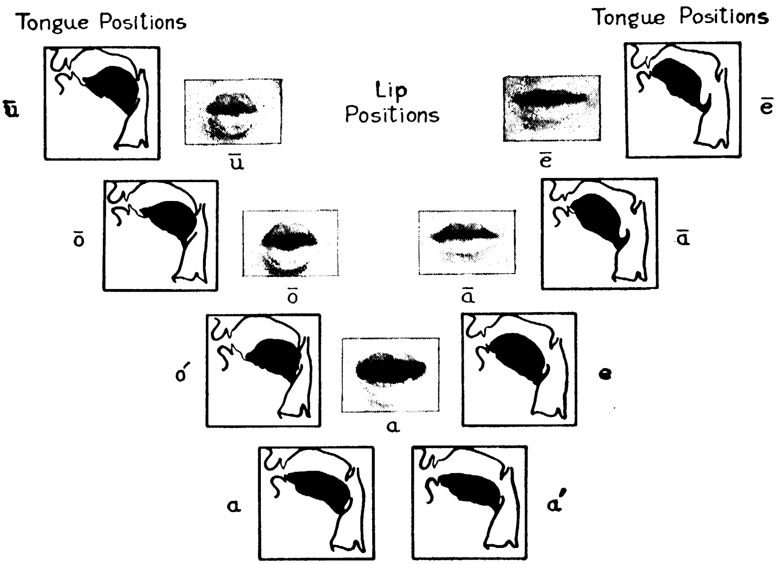


FIGURE 2.

the sound to the position to say ē, the result is signified in writing by the letter ī. Similarly, as illustrated in Table I, the diphthong ou is a combination of a and ū, the diphthong oi the combination of ō and ē, and the diphthong ew the combination of ē and ū.

The three sounds, w, y, and h, are called transitionals since they represent a particular way of beginning the vowel sounds. If the mouth is placed in the position to say ū and then suddenly changed so as to form any other vowel in the diagram, the result obtained is signified in writing by placing w before the vowel. In a similar way we obtain the effect designated by y if the position of the vowel suddenly changes from ē to any other vowel. The diphthong ew is different from the sound represented by yū only in that when the y is used the transition from the ē to the ū sound is much more rapid than when the diphthong is formed. An infinite variety of diphthongs and transitionals can be formed by varying both the rate of change and the size of the vocal cavities necessary to form the two vowels. The most distinct and principal ones used in our language are those shown in Table I. When a vowel begins a syllable, it is formed by suddenly opening the glottis and thus permitting the air which has been held in the lungs to escape into the mouth formed for the proper vowel. If the glottis is originally open the vowel is started by the sudden contraction of the lungs. Under these conditions the effect would be represented in writing by placing h before the vowel.

The sounds l, r, m, n, and ng are classified as semi-vowels since for these sounds the passage from the vocal cords to the outside air is partially blocked. In the case of l and r the sound is allowed to flow around the tongue which is placed in a particular positon in the mouth. For the sounds m, n, and ng the usual path through the mouth is interrupted so that the sound and the air accompanying it flow through the nasal cavities. For this reason they are sometimes called nasalized stop consonants.

The fricative consonants are characterized by the rushing sound of the breath through the characteristic air outlet which is usually of very small dimensions. The manner in which these sounds are formed is evident from Table I. For producing the sound f the outlet is produced by holding the lower lip to the upper teeth. If while the f sound is being thus produced, a tone from the vocal cord is also sounded, the speech sound v is formed. Similarly, other unvoiced and voiced fricative sounds are produced by changing the nature of the air outlet as shown in the table. The stop and fricative consonants are classified in a similar way, being both the voiced and unvoiced consonants. For example, b and p are both characterized by a stop formed with the lip against the lip, d and t with the tongue against the teeth, j and ch with the tongue against the hard palate, and g and k with the tongue against the soft palate.

The voiced sounds may be divided into two classes; those produced by a continuous flow of air which may be called the continuants and those produced by stopping the sound flow in certain ways which may be called the stops. The former group, including vowels and voiced consonants, is the one used to carry the pitch in singing. It is seen from Fig. 1 that such sounds pass through two variable resonating cavities, namely, the mouth and the throat. For this reason, all voiced sounds are characterized by having component frequencies magnified in two particular regions. It is largely the characteristics of these regions of resonance that distinguish one of these sounds from another, especially when they are sung. In speaking, however, the way the sounds are started and ended has considerable to do with the ability to recognize them. In any of the voiced sounds it is important to notice that it is the modulation of the cord tone that gives the distinctive sound rather than the characteristic of the vocal cords. The latter determines the type of voice and identifies the person who is speaking but has little to do with the characteristics of the speech sounds which determine their recognition.

Artificial Production of Speech Sounds

Once having a clear picture of the mechanism of speaking, one can easily see how an electrical apparatus can be made which will produce some of these vowel sounds. A convenient form of such apparatus is one in which the generator of the electrical vibrations is an overloaded vacuum tube oscillator which corresponds to the vocal cords. The complex wave generated consists of a fundamental and a large series of harmonics. These electrical vibrations are conducted through two electrically resonant circuits and then to a loud speaker. A schematic of the circuit for producing artificially the simple speech sounds is shown in Fig. 3. An arrangement similar to this was first used by Stewart.[1] If the resistance, capacity, and inductance of the two resonant circuits are adjusted to have resonant frequencies and dampings corresponding to those existing in the mouth and throat when one of the vowel sounds is being produced, then sound waves

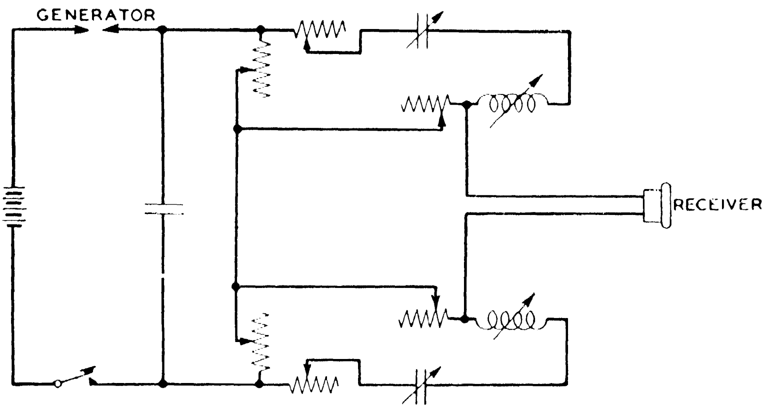


FIG. 3.—SCHEMATIC OF CIRCUIT FOR PRODUCING ARTIFICIALLY THE SIMPLEST SPEECH SOUNDS.

issuing from the receiver will have characteristics similar to this vowel sound. In producing the diphthongs (and in English speech most of the vowels are spoken as diphthongs) the resonant properties of these circuits must be varied from one condition to another in a definite manner.[2]

Sir Richard Paget has recently developed some acoustic apparatus for artificially producing speech. A device for making a sound similar to that produced by the vocal cords is attached to a large bellows which is operated by the foot. This is attached to resonating air chambers having the proper resonant frequencies and damping characteristics for representing the various vowel sounds. A good representation of

[1] Stewart, J. Q., "An Electrical Analogue of the Vocal Organs," *Nature*, Vol. 110, September 2, 1922, pp. 311-312.

[2] An apparatus similar to that described was used by the author before the New York Electrical Society, February, 1924, to demonstrate the production of the sounds, a, e, i, o, u, ya, mamma and papa.

some of the stop consonants is produced by properly interrupting the sound at an artificial mouth which is provided.

Artificial Larynx

Some recent experiences with persons who have lost their larynx through an operation has emphasized the fact that the

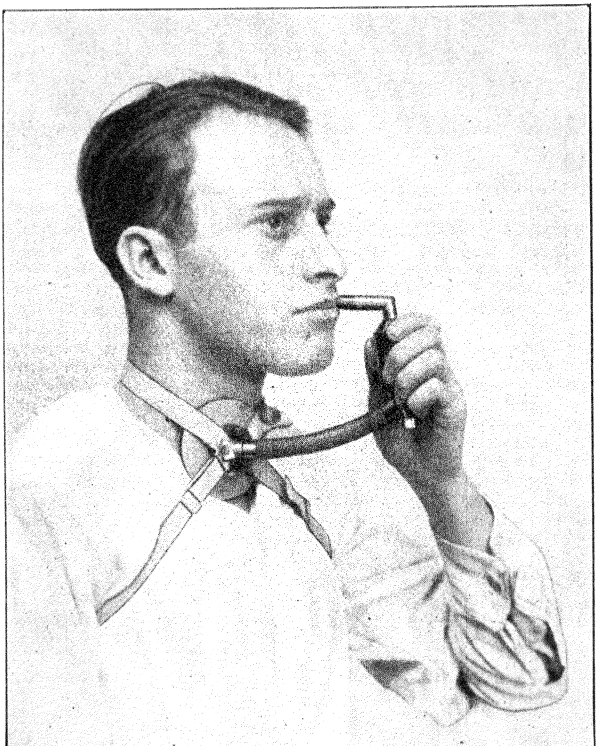


FIGURE 4.

differentiation of the speech sounds is practically all accomplished by the mouth and lip positions and that the sounds from the vocal cords act only as a sort of carrier.

The surgical operation known as a tracheotomy leaves no connection between the lungs and the mouth. It is performed usually in an emergency to prevent the patient from dying of suffocation. When a patient recovers from such an operation, the process of breathing is carried on by drawing the air in and out through a small opening in the neck. Because of this by-passing of the larynx, the patient can make no vocal sounds. However, if an attachment is made to the opening of the wind pipe so that the patient can blow a whistle, something similar to that used in toy balloons, and the sound directed into the corner of the mouth, the patient can learn to talk again.

Through the cooperation of Dr. J. E. Mackenty of New York City and engineers of Bell Telephone Laboratories, a device called an artificial larynx was developed and is now being used. Figure 4 shows a photograph of this device. More than one hundred persons in the United States who have undergone this operation are now using it successfully.

CHAPTER II

Characteristics of Speech Waves

Speech sounds radiating from the mouth are transmitted through the air by means of pressure waves, successions of condensations and rarefactions of the air. The magnitudes of the pressure changes making up these vibrations are exceedingly small and the wave form is complicated as the cavities of the mouth and the throat are continually varying in size and the stream of air is being constantly interrupted. The physical characteristics of these waves which carry typical speech sounds will be discussed in this chapter.

Methods of Recording Speech Waves

Inasmuch as the time of passage of such waves is very short, it is desirable for scientific study to have permanent records of the movements of the air particles as the wave traverses them. The examination of such records will reveal the physical characteristics of the sound wave.

Helmholtz and some of the other earlier investigators in this field analyzed the intoned vowels by listening to the sound after it passed through acoustical resonators. Since his time, these well-known devices have been called Helmholtz resonators. However, due to uncertainties in an observer's judgment only a very rough analysis can be made by using them. By these means, however, Helmholtz and some of his contemporaries found that the vowels on the left side of the vowel triangle in Table I were characterized by a single resonant reinforcement, and those on the right side by a double resonant reinforcement and for each they gave values for the characteristic resonant frequencies.

The sound wave corresponding to a word or syllable, assuming that it is free to travel without reflection or refraction, can be specified in either of two ways: (1) by giving for every air particle over the entire length of the disturbance the displacement at any instant from its position of equilibrium along a line perpendicular to the wave front or (2) by giving at every instant of time while the wave disturbance is passing the displacement of a single particle. Recording a syllable having a duration of one-fifth of a second requires a free air space about 200 feet long. Although it is not impossible to think of some scheme of photographing the rarefactions and condensations along such a space, no practical method has yet been devised for doing it. The nearest approach to it is probably that obtained by Sabine [1] in photographing the condensations and rarefactions produced in a large theater by the sound of an electric spark.

The second method has been used in several ways for making permanent records of speech sounds. Usually a light diaphragm is used to indicate the movements of the air particles. If such a diaphragm is very light and unconstrained except by the layers of air on either side of it, it will execute practically the same motions during the passage of the sound wave as the air particles would execute if the diaphragm were absent.[2]

Probably the nearest approach to such a diaphragm is a soap film. An instrument provided with such a film and called a Weiss [3] phonoscope, has been used with some success although considerable difficulties are encountered in keeping it in adjustment. A beam of light reflected from the film surface indicates its movement. Most such indicating devices use diaphragms which are comparatively heavy and are con-

[1] "American Architect," 104, pp. 257–279.

[2] The diaphragm must be lighter than a quantity of air having an equal cross-section which is vibrating approximately in phase, that is, which is about one-tenth of the wave length in width. Such a quantity of air weighs about one milligram per square centimeter of cross-section for a 5000-cycle sound wave.

[3] *Mediz. naturw. Arch.*, Band I, Heft 2, December 15, 1907. Otto Weiss, *Das Phonoskop, eine Vorrichtung zur analyse und registrierung schwacher Schallqualitaten.*

strained by being clamped at the edge. These diaphragms usually have several pronounced resonant frequencies but serve to give a rough indication of the type of motion of the air particles as the sound wave passes over them.

Koenig's Phonautograph and Manometric Capsule

One of the first instruments using a diaphragm of this description, called the "phonautograph," was designed by Koenig and Scott.[1] A photograph of this instrument[2] is shown in Fig. 5. The membrane is clamped at the end of a horn. A stylus attached to this membrane makes a trace on the smoked paper carried on the revolving drum.

Koenig also devised the manometric capsule for producing the manometric flame which is still used as a demonstration instrument in many physical laboratories. This apparatus[3] is shown in Fig. 6. It may be seen from the cross-section (*A*)

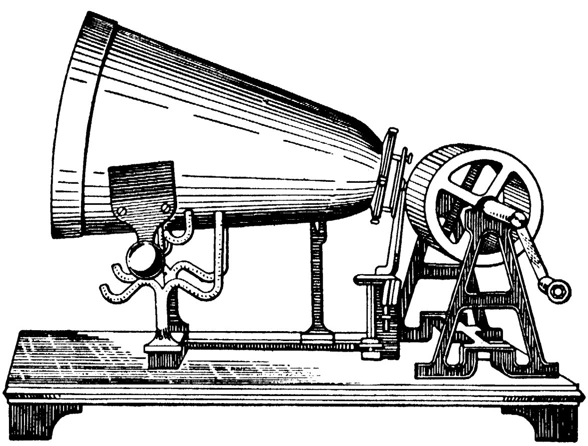


FIGURE 5.

that the capsule is divided into two compartments separated by a membrane usually made of thin rubber. The gas supplied for the flame enters one compartment and the sound

[1] Koenig and Scott, "Cosmos," 14, p. 314, 1859.
[2] Rousselot, P., "Principes de Phonetique Experimentale," p. 110.
[3] Rousselot, P., "Principes de Phonetique Experimentale," p. 111.

waves are sent into a tube connected with the other. The variations produced in the gas pressure are indicated by a rise and fall of the gas flame. By reflecting the gas light from a revolving mirror as shown in the figure, these variations

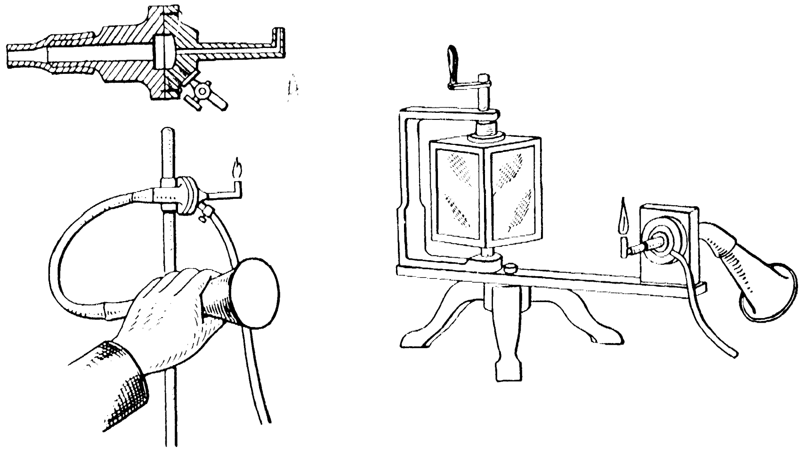


FIGURE 6.

are made visible or may be photographed. Professors Nichols and Merritt [1] perfected this method and obtained records of vowels and spoken words. Figures 7 and 8 show some of their records.[2] There have been variations of these two methods since Koenig's time but no experimenters have been successful in obtaining a diaphragm which would execute vibrations even approximately proportional to the pressures produced in the speech wave. For this reason, any speech wave pictures obtained with such apparatus give only a rough indication of what is taking place as the wave passes through the air.

The Phonodeik

The most highly developed successor of the phonautograph is an instrument devised by D. C. Miller and called by him the "phonodeik." In this instrument the stylus of the

[1] *Physical Review*, Vol. 7, p. 93, 1898.
[2] *Physical Review*, Vol. 7, p. 92, 1898.

phonautograph is replaced by a light-weight mirror system which reflects rays to a moving film. The following descrip-

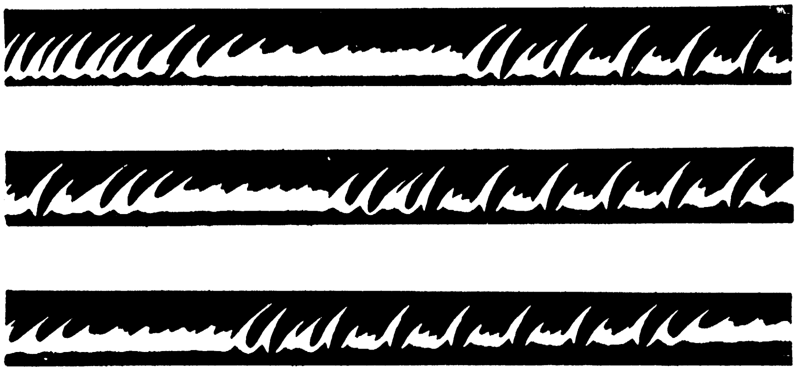


FIG. 7.—RECORD OF "R" (ROLLING). OBTAINED BY PROFESSORS NICHOLS AND MERRITT.

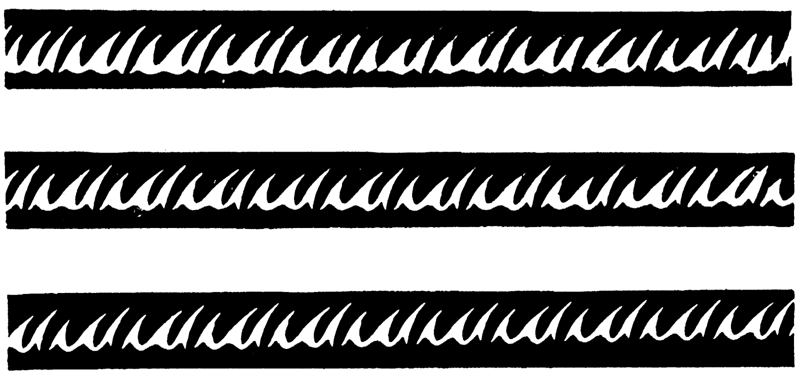


FIG. 8.—RECORD OF "A" ("A" FLAT). OBTAINED BY PROFESSORS NICHOLS AND MERRITT.

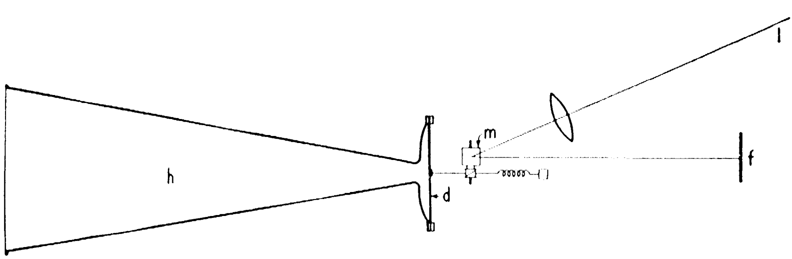


FIG. 9.—PRINCIPLE OF THE PHONODEIK.

tion of this instrument is taken from his book, "The Science of Musical Sounds."

"The sensitive receiver of the phonodeik is a diaphragm, *d*, Fig. 9, of thin glass placed at the end of a resonator horn *h*; behind the diaphragm is a minute steel spindle mounted in jeweled bearings, to which is attached a tiny mirror *m*; one part of the spindle is fashioned into a small pulley; a few silk fibers, or a platinum wire 0.0005 inch in diameter is attached to the center of the diaphragm and being wrapped once around the pulley is fastened to a spring tension piece; light from a pinhole *l* is focused by a lens and reflected by the mirror to a moving film *f* in a special camera. If the diaphragm moves under the action of a sound wave, the mirror is rotated by an amount proportional to the motion, and the spot of light traces the record of the sound wave on the film."

Figure 10 shows a photograph [1] of the phonodeik ready for use. Professor Miller did a large amount of work in an endeavor to correct the speech records obtained by the phonodeik for the distortions introduced by the resonance characteristics of the horn and diaphragm. These corrections were for the amplitude distortion only, no attempt being made to correct for phase distortion since he was interested mainly in the relative amplitudes of the component frequencies and not in the relative phases.

Numerous oscillograms taken by the phonodeik and then analyzed by a harmonic analyzer verified the conclusion of Helmholtz that the vowels on the left side of the triangle in Table I were singly resonant while those on the right side were doubly resonant. Although other apparatus is now available which gives a very much more accurate copy of speech waves, the phonodeik is still a very satisfactory instrument for demonstration purposes.

The Phonograph

Another method of studying the forms of sound waves uses the phonograph, invented in its original form by Edison in 1877, which makes a permanent record of speech waves.

[1] Taken from D. C. Miller's book, "Science of Musical Sounds."

The trace on the phonograph record can be made visible in a number of ways. Herman[1] in 1890 and Bevier[2] in 1900 used a delicate tracing point carrying a mirror mounted so that as

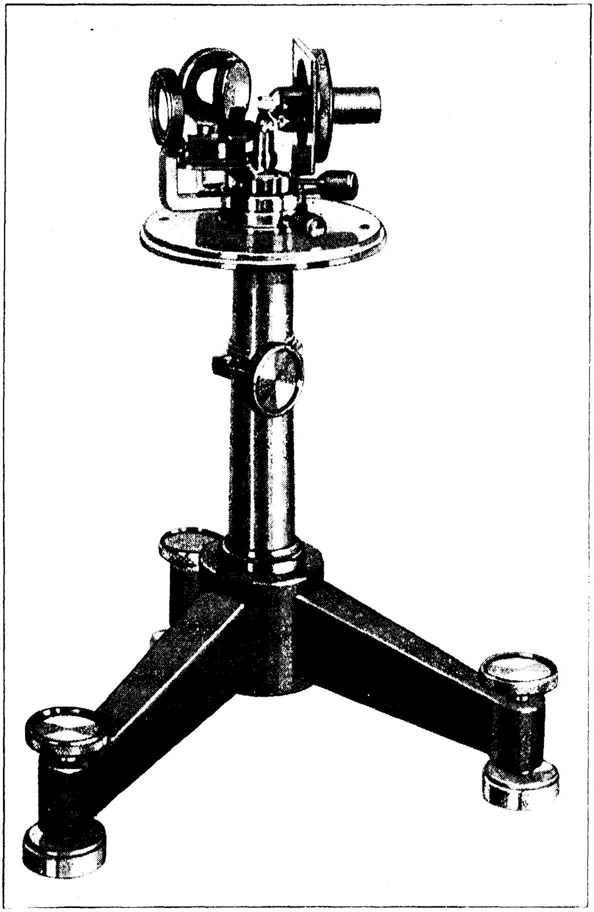


FIG. 10.—THE PHONODEIK.

the point traversed the undulations in the record, a beam of light, reflected from the mirror, fell upon a moving photographic plate. The record can be turned so slowly that the

[1] Herman, L., *Pflügers Archiv.*, 45, 282 (1889); 47, 42, 44, 347 (1890); and others.
[2] Bevier, L., *Physical Review*, 10, 193 (1900).

mass and elasticity of the lever system do not cause any appreciable distortion. Scripture[1] did a large amount of work using a similar method. In place of the mirror system he substituted a long lever which carried a stylus at its end. As it vibrated it traced an undulating curve on smoked paper mounted on a revolving drum. The arrangement of Scripture's apparatus[2] is shown in Fig. 11. Using this apparatus,

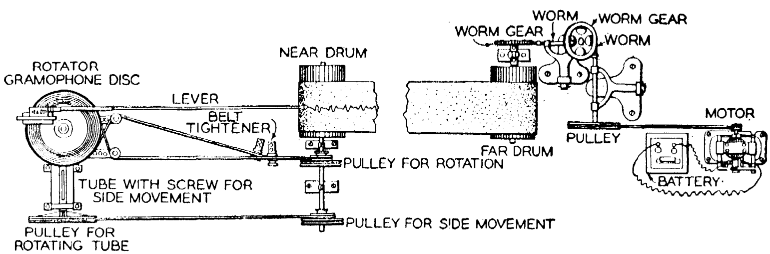


FIGURE 11.

Scripture obtained tracings[3] shown in Fig. 12 which give the wave form for the vowels spoken by Joseph Jefferson and recorded on a disc record using facilities available at that time (1903).

Just recently there has been considerable improvement in the technic of recording phonograph records. In the old method the power of the sound being recorded was used to operate the recording instrument. The old phonautograph with a sharp needle for a stylus illustrates the principle of the phonograph method of recording used until just recently. In order to obtain good records by this method it is necessary to obtain large intensities close to the horn of the recorder. Figure 13 shows some of the difficulties encountered when using this method. This and the next figure were taken from

[1] Scripture, E. W., "Researches in Experimental Phonetics," Carnegie Institution of Washington Publication No. 44.

[2] Scripture, E. W., "Researches in Experimental Phonetics," Carnegie Institution of Washington Publication No. 44, p. 24.

[3] Scripture, E. W., "Researches in Experimental Phonetics," Carnegie Institution of Washington Publication No. 44, p. 51.

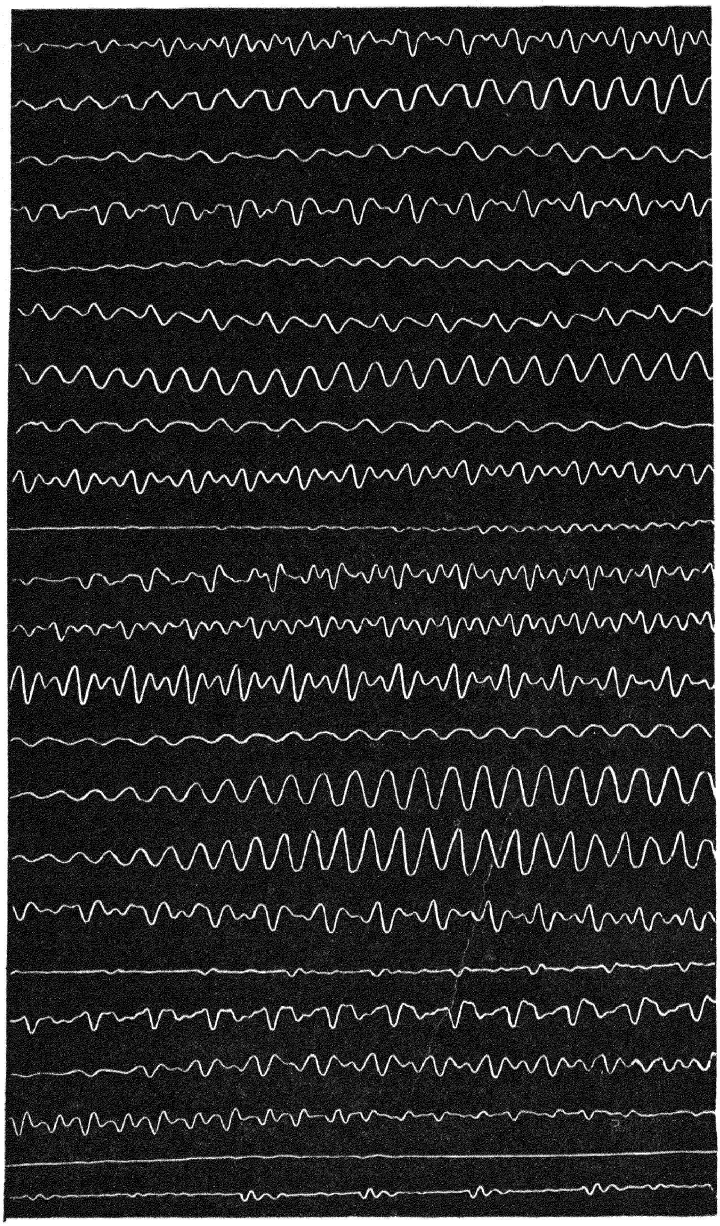


Fig. 12.—Waves from Vowels by Joseph Jefferson.

a paper [1] by J. P. Maxfield and H. C. Harrison, two engineers of Bell Telephone Laboratories who were principally responsible for the development of the new electrical method of recording. In the new method high-quality telephone apparatus with vacuum tube amplifiers is used, which gives more freedom to the artists and much better control of the cutting

FIGURE 13.

needle. The amount of power available to operate the cutting needle is not dependent upon the original acoustic energy in the sound wave, but may be made any convenient value by changing the amplification in the vacuum tube amplifiers. Figure 14 shows the same orchestra as shown in Fig. 13 but recording by the new electrical process. The greater flexibility of the new system of recording makes possible very much better records of speech and such records naturally yield more reliable results when analyzed by the method above mentioned.

The phonograph method has a distinct advantage that the speech may be reproduced and compared with the original.

[1] Methods of High Quality Recording and Reproducing of Music and Speech Based on Telephone Research, presented at A. I. E. E. Convention at New York City, February 8–11, 1926.

If it is a sufficiently good reproduction it may be safely assumed that the wave on the record faithfully represents the original sound. The analysis of the speech wave into component frequencies can be greatly facilitated when it is on a phonograph record for all the component frequencies can be located by changing the speed of the record during the reproduction and noting the response of a single resonator, either acoustic or electric.

Another method of recording speech which is now assuming considerable commercial importance, makes use of a motion-picture film. The record consists of variations in the density of the film corresponding to pressure variations in the sound wave. There are two principal methods used. In the first

FIGURE 14.

a slit of fixed width is placed near the moving film and illuminated by an electric lamp whose intensity is controlled by variations in the pressure of the sound wave. For this purpose the apparatus for electrical recording which has been described can be used, but in place of the cutting tool a lamp is substituted. A circuit is arranged so that variations in the voltage impressed upon the lamp cause similar variations in the intensity of the light emitted. Various types of lamps have been proposed, the requirement for obtaining faithful record-

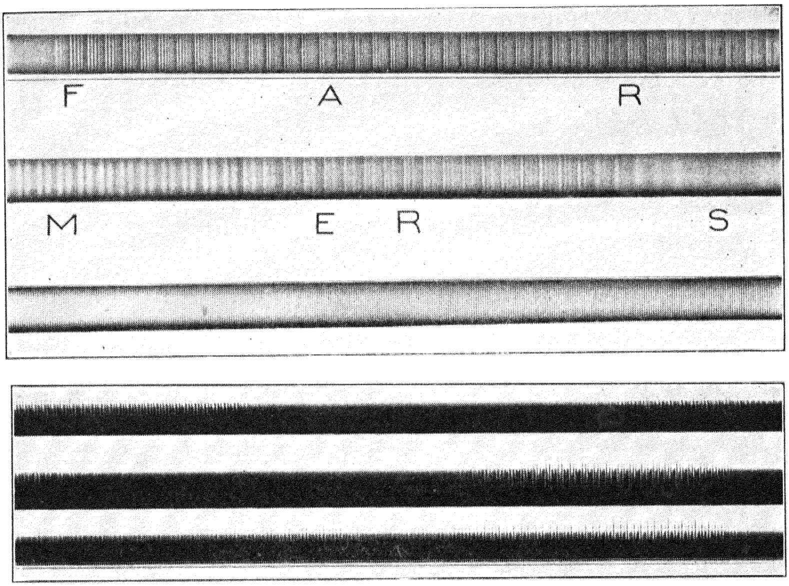


FIG. 15.—SOUND FILM RECORDS. ABOVE, LIGHT VALVE RECORDS, WITH 1000 CYCLE TONE; BELOW, PROFILE RECORDS.

ing being that the intensity of the light emitted be proportional to the voltage impressed upon its terminals. It is difficult to obtain an ideal lamp of this sort although some fairly good records have been produced in this manner.

In the second method the intensity of the light is held constant while the width of the slit is made to vary in accordance with changes in the amplitude of the sound wave. A device for doing this is called a light valve. It consists essentially of a shutter or shutters which are controlled electromagnetically. If this device is properly constructed, it can be made to give very accurate records. The first two charts of Fig. 15 give sample records taken by the latter method. The first is for the word "farmers"; the second is for a pure tone having a frequency of 1000 cycles per second. It is interesting to note that such a record gives a true picture of the variations in the density of the air at any instant along the train of waves passing through it. The original sounds can be reproduced from such records by means of suitable apparatus. For this purpose a narrow beam of light is transmitted through the moving film and permitted to fall upon a photo-electric cell. The variations in the intensity of the light which are caused by variations in the density of the film cause similar variations in the photo-electric currents. These currents are then magnified by means of vacuum tube amplifiers and finally reproduced as sound waves by a loud speaker. Although such records can now be made very accurately and are useful for purposes of analysis, they do not give a picture of the wave form that is as readily interpreted by the eye as an oscillogram.

The High-quality Oscillograph

On account of the importance to the telephone industry of obtaining accurate pictures of speech waves, considerable research work has been done in Bell Telephone Laboratories to perfect the method of recording speech sounds by means of an oscillograph. This work was directed by the late Dr. Crandall, most of the details of the designing of the apparatus and the making of the records being done by C. F. Sacia.

Briefly stated, the principles involved in this method are as follows: Speech waves are picked up by a telephone transmitter and converted into electrical waves. They are then magnified by means of a special amplifier and sent into an oscillograph, where they cause a tiny ribbon to vibrate. The motion of this ribbon is then photographed on a moving film. The perfection of this instrument for recording speech sounds, as well as many other instruments used in acoustic measurements, some of which will be described later, depends upon the development of three important devices, namely: (1) the condenser transmitter and the means of calibrating it, (2) the vacuum tube with the circuit arrangements for producing amplification and electrical oscillations, and (3) the oscillograph. By means of these devices, an instrument for recording speech sounds was finally obtained which was more nearly free from distortion than any which had yet been used. For this reason a somewhat detailed description of it will be given.[1]

The element for converting the sound waves into electrical waves is the condenser transmitter which has been thoroughly investigated by E. C. Wente. It has an approximate uniform-response characteristic from 0 to 8000 cycles per second, giving about 3×10^{-4} volts at its terminals per bar (dyne per square centimeter) pressure on its diaphragm. The vacuum tube amplifier consisted of 7 stages, the last stage containing 8 tubes in parallel, making 14 tubes in all. This provides a voltage amplification of about 40,000 and the 8 tubes in the last stage make it possible to work into the low impedance oscillograph vibrator without using a coil transformer. The coupling between the tubes was entirely free from coils. This was necessary in order to preserve a uniform characteristic for various frequencies. The vibrator of the oscillograph was specially constructed, having small mass, high tension, and a moderate amount of damping. This was necessary to obtain a uniform response. These three elements of the recording system were connected together so as to produce as little distortion as possible. An ideal system would reproduce all frequencies with the same efficiency and produce phase lags which are proportional to the frequency. The energy output should also be proportional to the energy input. Such a system would record sounds without any distortion. In Fig. 16 are shown the amplitude and phase characteristics of the system as it was finally arranged. It will be seen that the first two requirements mentioned are well fulfilled. The measurements indicated that throughout the range of intensities used, the last condition was also well fulfilled.

The apparatus was sufficiently powerful to record sounds

[1] For a more complete description see "Sounds of Speech," by I. B. Crandall, published in *The Bell System Technical Journal*, October, 1925.

spoken in an ordinary tone of voice when the speaker's lips were about 3 inches from the transmitter. A key was pressed by the speaker just before the sound was spoken, which released a shutter placed before a rotating film drum on which the record from the oscillograph vibrator was traced. With the size of the drum and the speed at which it was rotated each one-hundredth of a second corresponded to 2 inches or more on the time scale. The apparatus was arranged so that the helical trace made on the record was

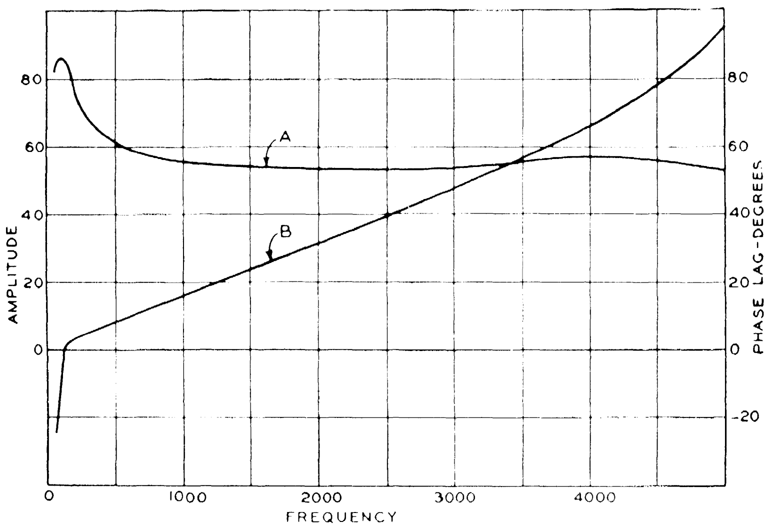


FIG. 16.—OVERALL FREQUENCY CHARACTERISTICS OF AMPLITUDE AND PHASE OF THE RECORDING SYSTEM. CURVE A: OSCILLOGRAPHIC AMPLITUDE PER UNIT OF PRESSURE ON TRANSMITTER DIAPHRAGM. CURVE B: PHASE LAG OF OSCILLOGRAPHIC AMPLITUDE BEHIND PRESSURE ON DIAPHRAGM.

200 inches in length for one second of time. As is indicated from the records obtained, special care was taken with the optical system to insure fine definition, and in the development of the films to obtain the proper contrast.

Typical Speech Waves

The wave forms of the words "farmers," "seems," "poor," and "alters" taken by this method are shown in Figs. 17, 18, 19, and 20. These serve to illustrate the complicated structure of speech waves, and the effects of starting and stopping the sound. It will be seen in Fig. 17, which gives the oscillogram of the word "farmers," that the first letter sound "f" is

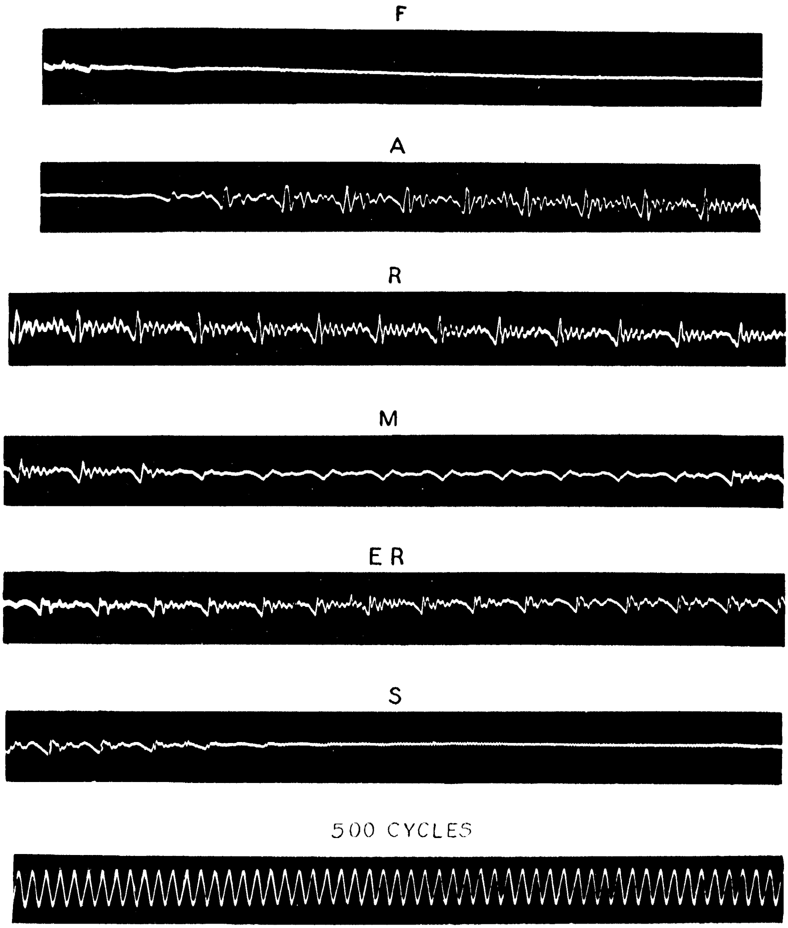


FIG. 17.—WAVE FORM OF THE WORD "FARMERS."

characterized by very high frequencies. After these high frequencies the "a" sound is produced by only five complete waves having fundamental frequencies corresponding to approximately 120 cycles per second. The "a" sound is followed by about twenty complete waves of the "r" sound having this same fundamental frequency, followed by about nine complete waves of the "m" sound also with the same frequency. As the "er" sound was reached the pitch of the voice was slightly raised to a pitch corresponding to a fundamental frequency of about 130 cycles per second. This was followed

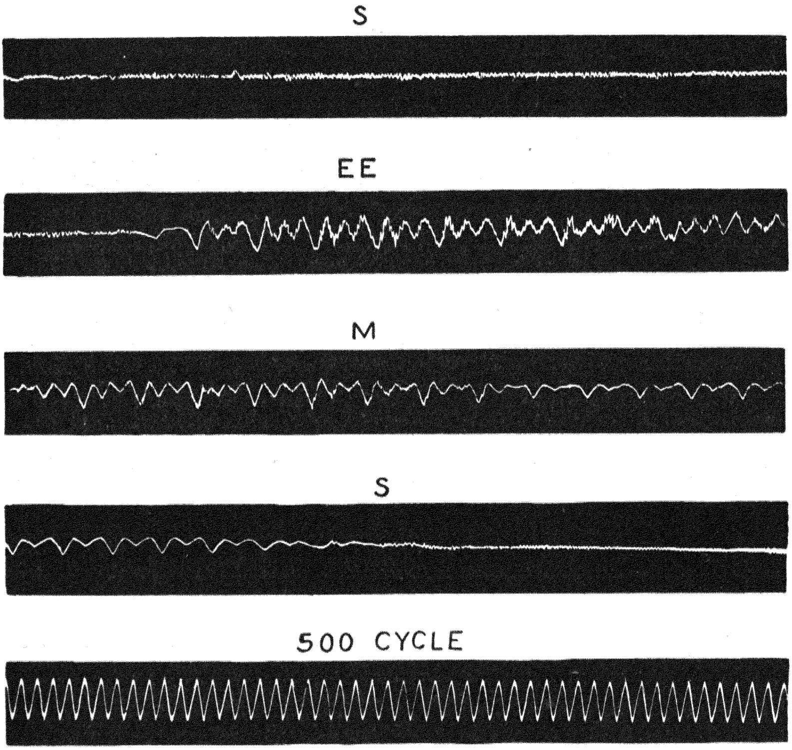


FIG. 18.—WAVE FORM OF THE WORD "SEEMS."

by the "s" sound, again characterized by very high frequencies.

Charts for all the speech sounds both for male and female voices were obtained by means of this apparatus. In Figs. 21 to 32 complete charts are given for twelve of these records, the ones selected by Crandall as being typical. In Figs. 33, 34, and 35 are shown the records taken by the same voice for all the long vowels, the short vowels, and the semi-vowels. Only the typical part of the wave is given for each case. These pictures show the forms of the waves as they emerge from the mouth. In a room with reflecting walls, the sound which finally reaches the ear of a person three or four feet away from the speaker is a combination of the original wave and several reflected waves. The amplitudes and phases of the components in the reflected waves which reach the ear are very

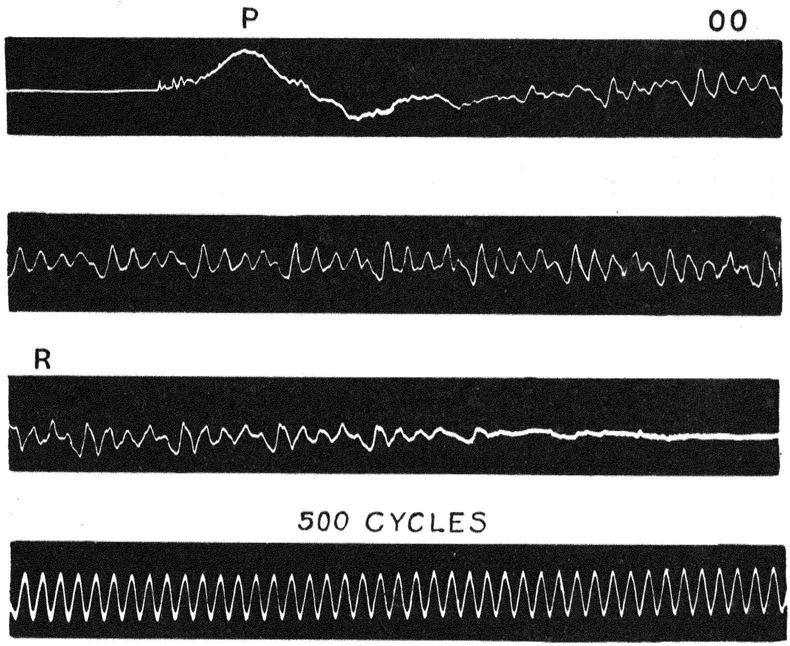


FIG. 19.—WAVE FORM OF THE WORD "POOR."

different from those of the original and also from each other, so that when they combine at the ear they form a wave with a shape entirely different from that of the original wave emerging from the mouth. If the phases only of the components are changed and the relative amplitudes remain the same, the ear usually recognizes no change; in other words, the ear does not ordinarily recognize phase differences. To illustrate this change in wave form with phase shift, two graphs representing the vowel sound "ah" are shown in Fig. 36. The amplitudes of the component frequencies were experimentally determined and the graphs were then calculated. The first wave form represents the wave picture when the component frequencies have no phase displacements. The component frequencies of the second wave form have the same amplitudes as the

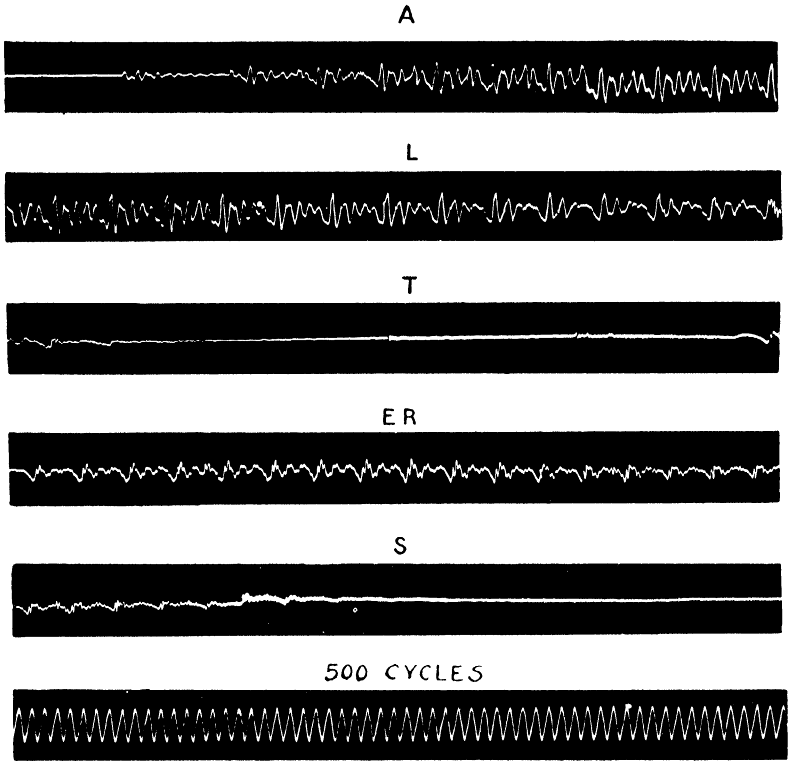


FIG. 20.—WAVE FORM OF THE WORD "ALTERS."

first ones but the phase displacements are proportional to the square root of the frequency. This kind of phase distortion is produced by a non-loaded cable telephone line. As stated, if the phase displacement is proportional to the frequency of the component, the wave picture does not change. The two wave forms are quite different although the acoustic spectra

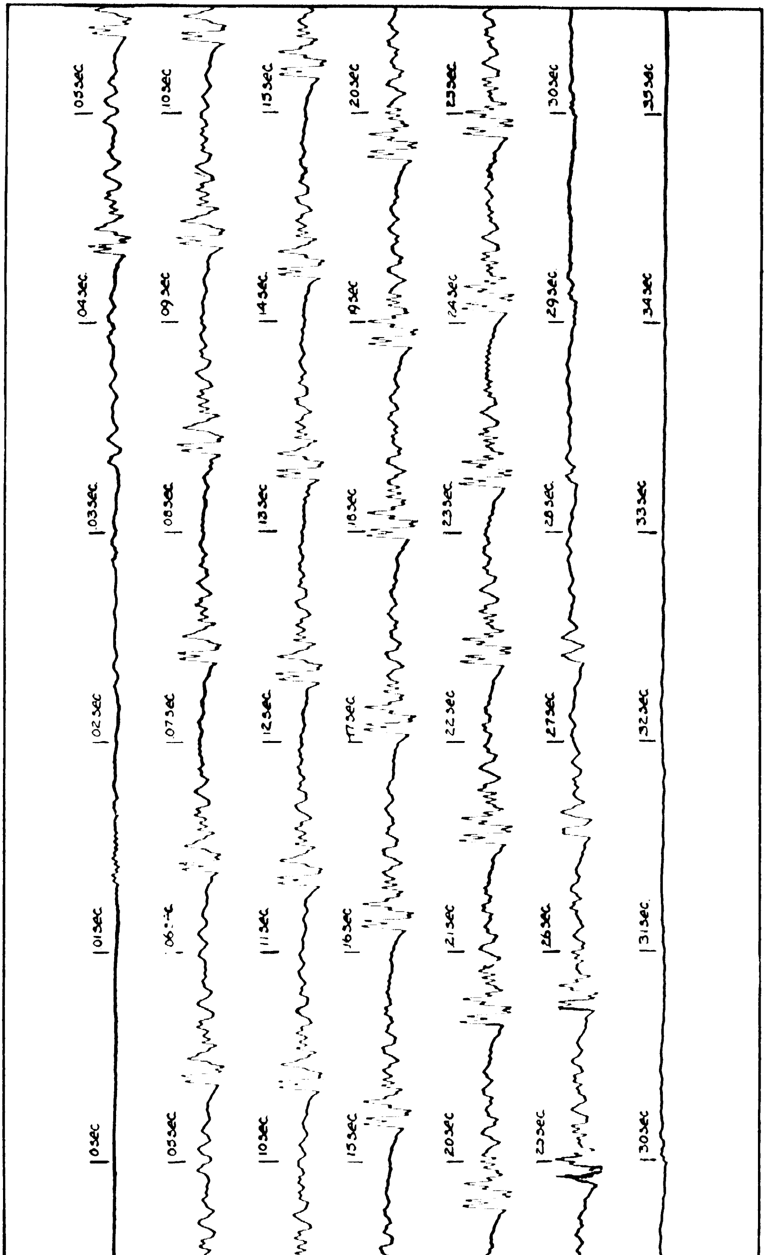


FIG. 21.—"A" AS IN "FATHER." SPOKEN BY M.A.—MALE, LOW-PITCHED.

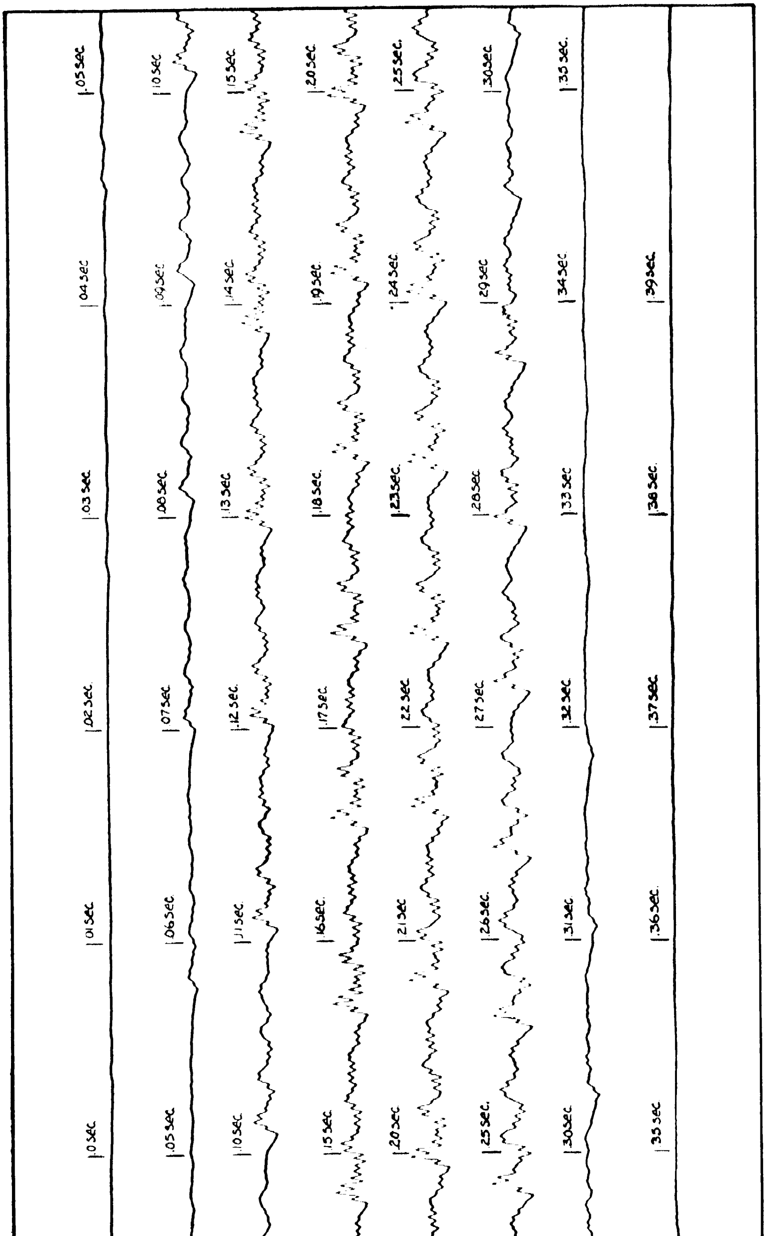


FIG. 22.—"U" AS IN "PUT." SPOKEN BY M.A.—MALE, LOW-PITCHED.

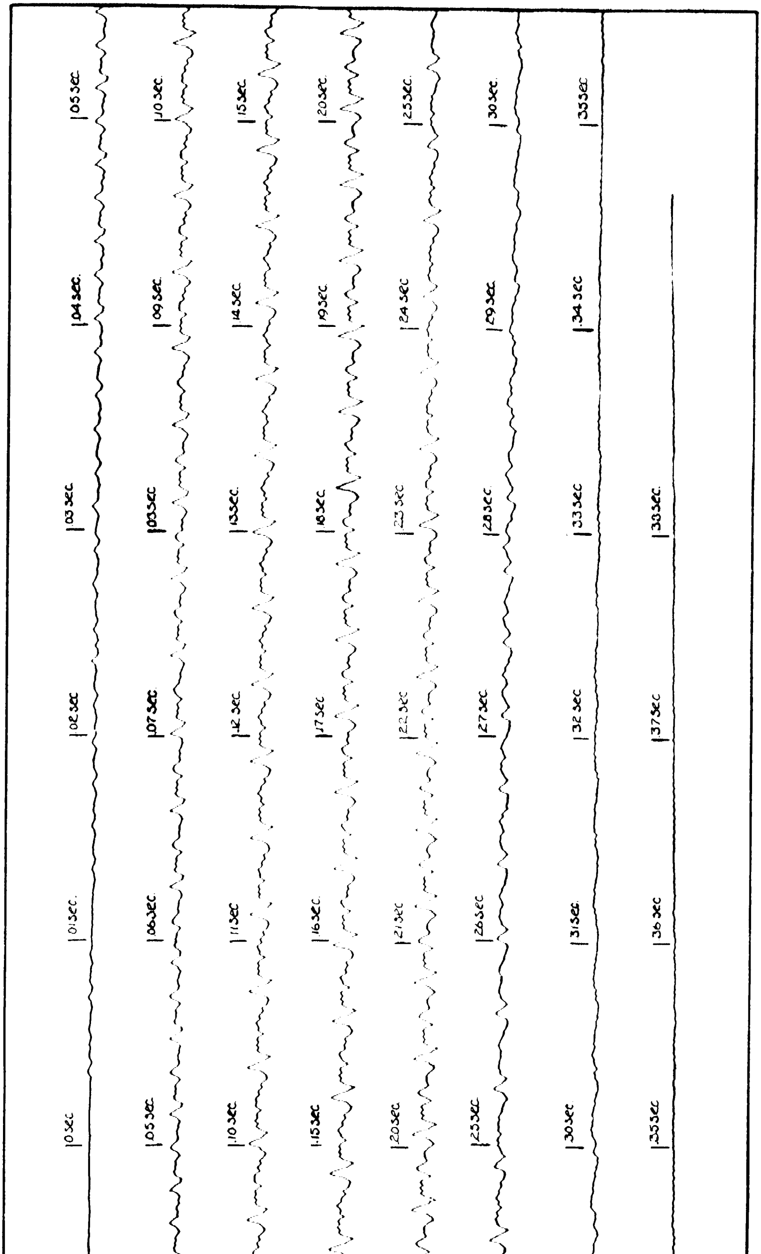


FIG. 23.—"O" AS IN "TON." SPOKEN BY F.D.—FEMALE, HIGH-PITCHED.

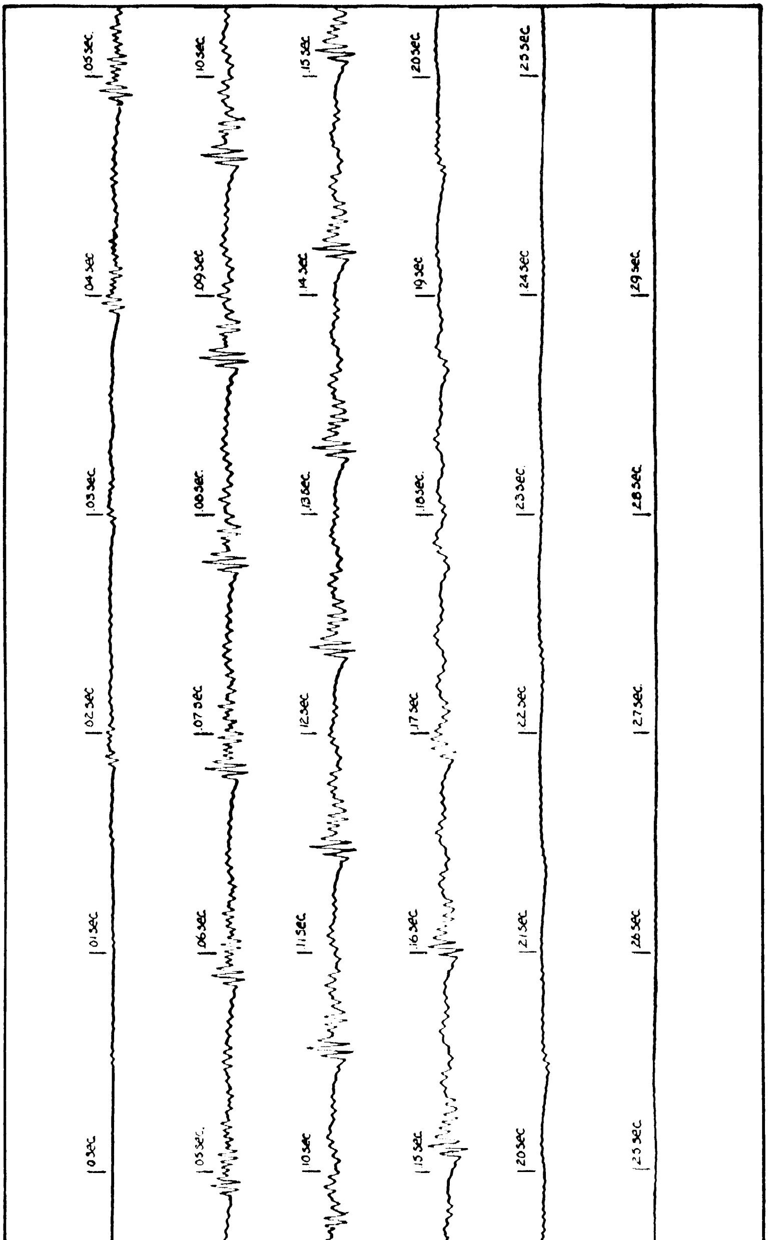


Fig. 24.—"i" as in "tip." Spoken by M.A.—Male, Low-Pitched.

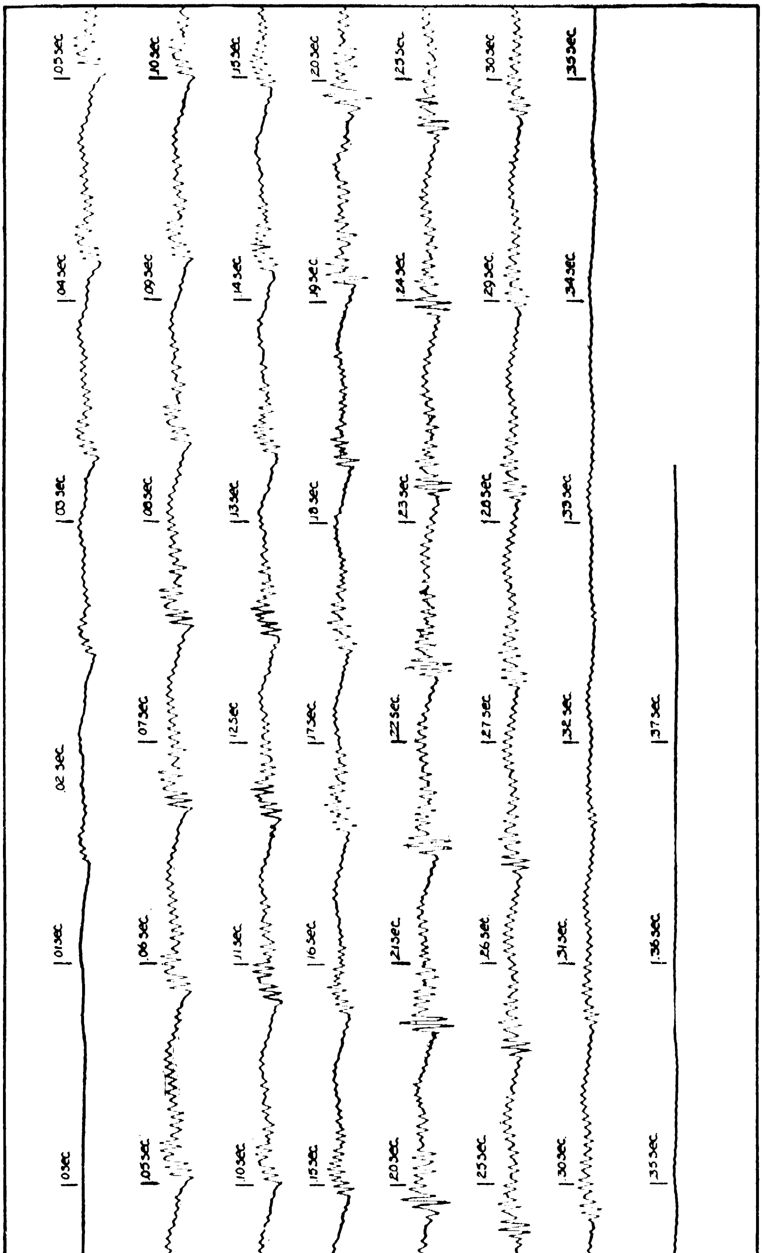


FIG. 25.—"LEE." SPOKEN BY M.B.

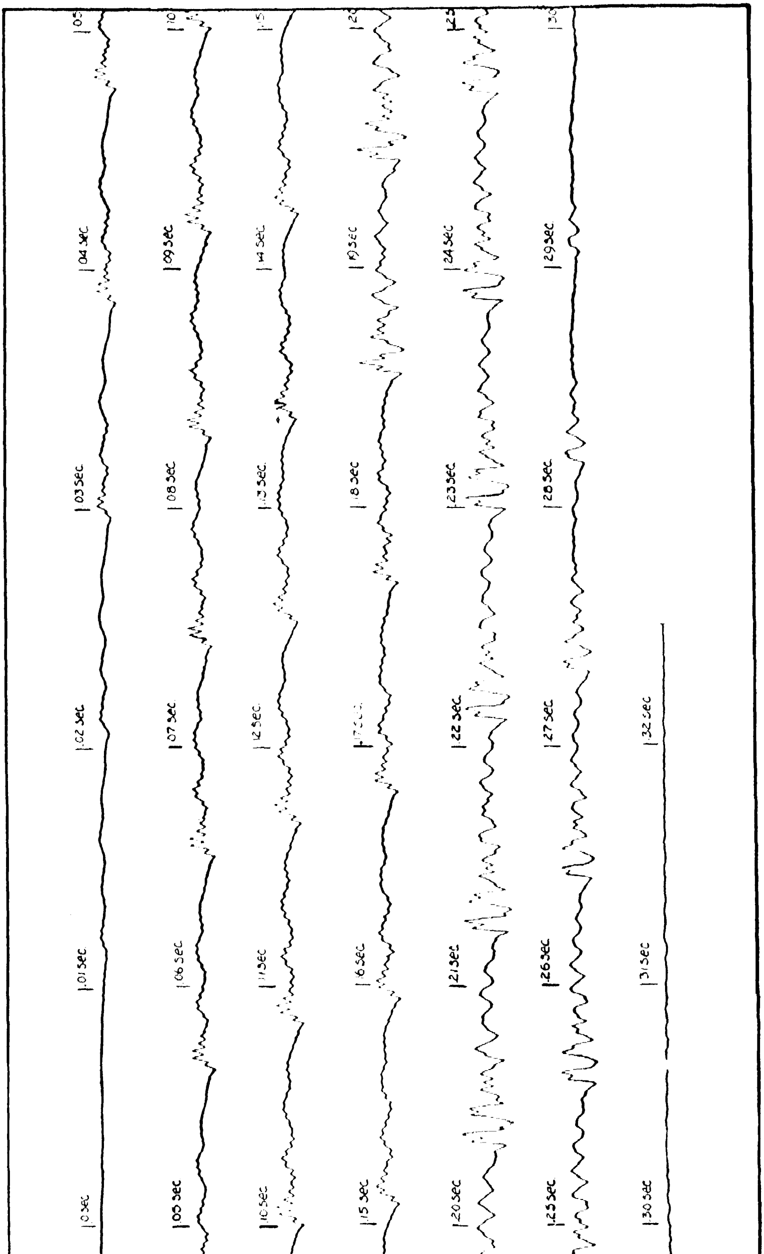


FIG. 26.—"LA." SPOKEN BY M.B.

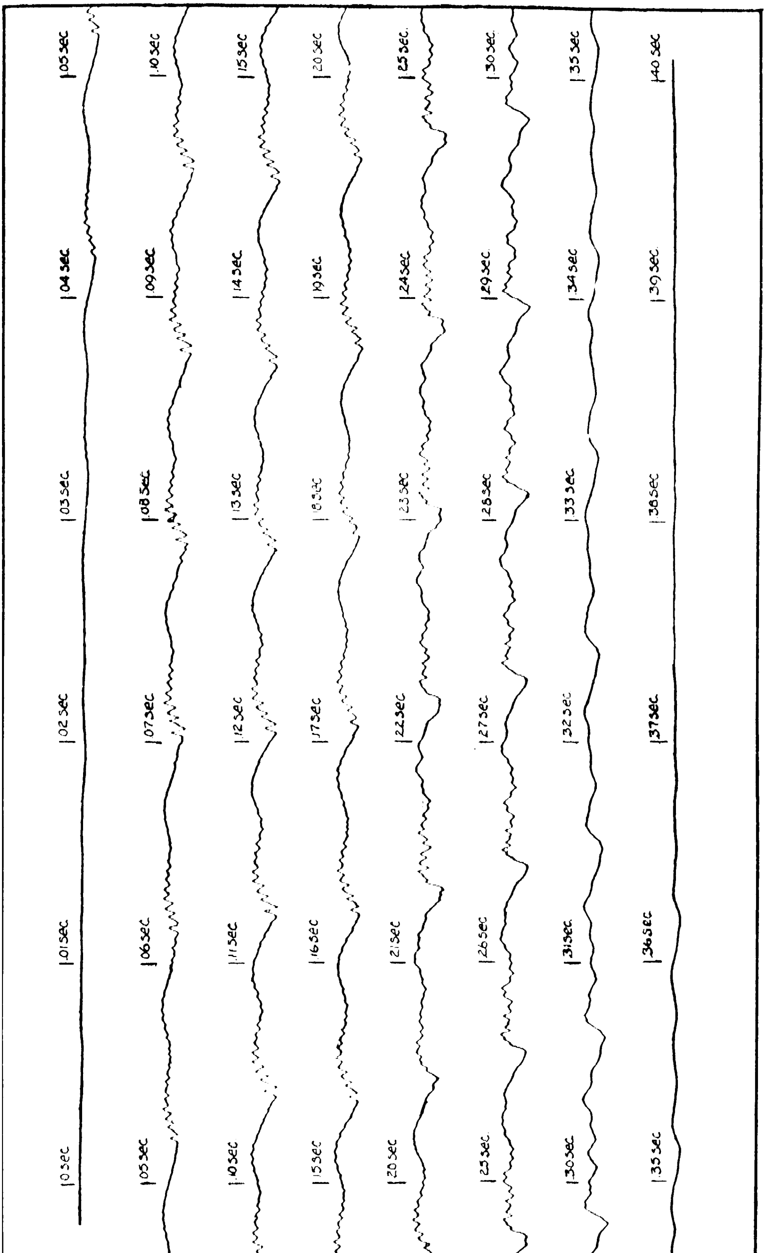


FIG. 27.—"MOO." SPOKEN BY M.B.

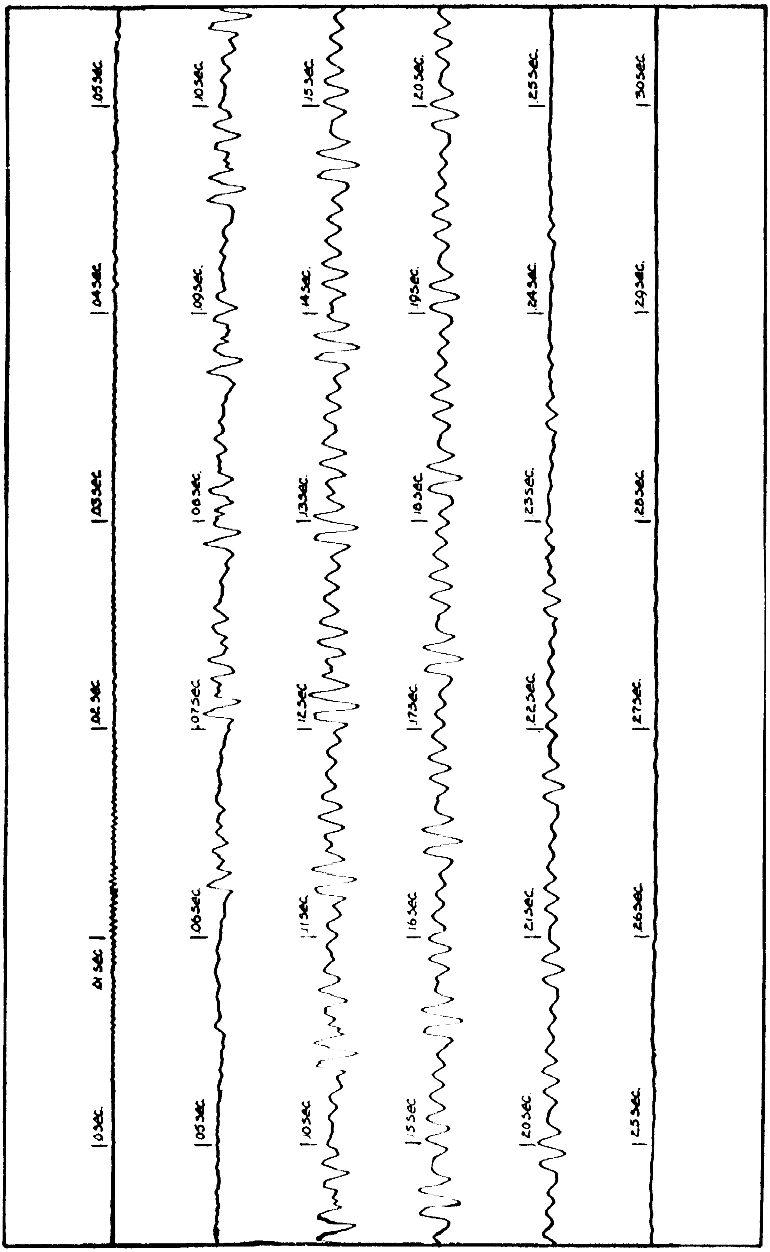


FIG. 28.—"TA." SPOKEN BY M.B.

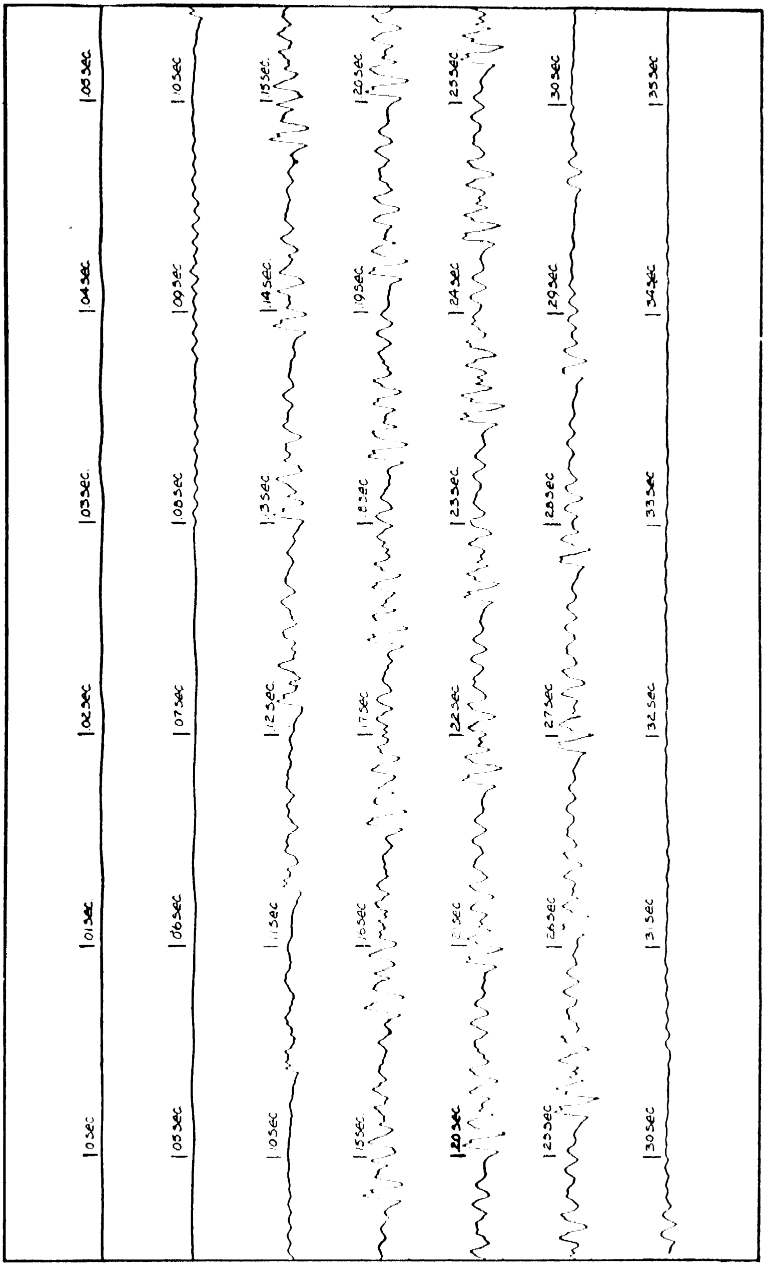


Fig. 29.—"ga." Spoken by M.B.

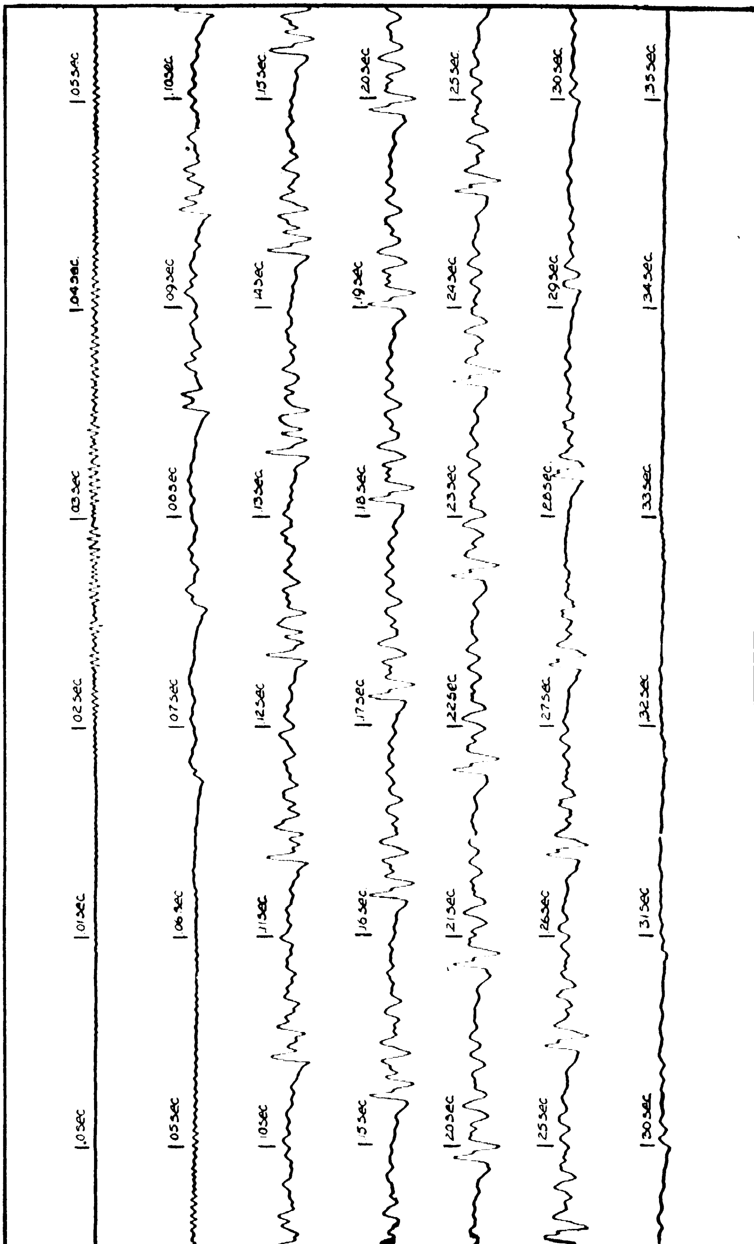


FIG. 30.—"CHA." SPOKEN BY M.A.

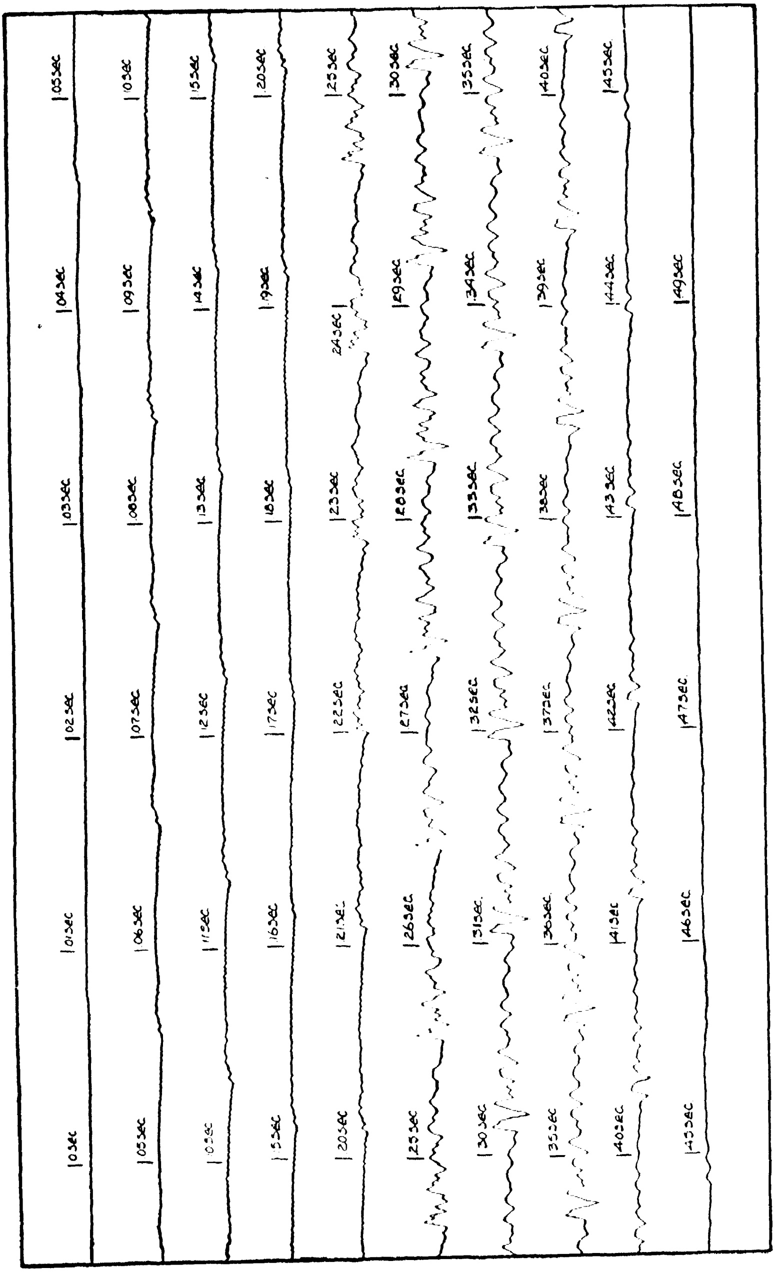


Fig. 31.—"za." Spoken by M.B.

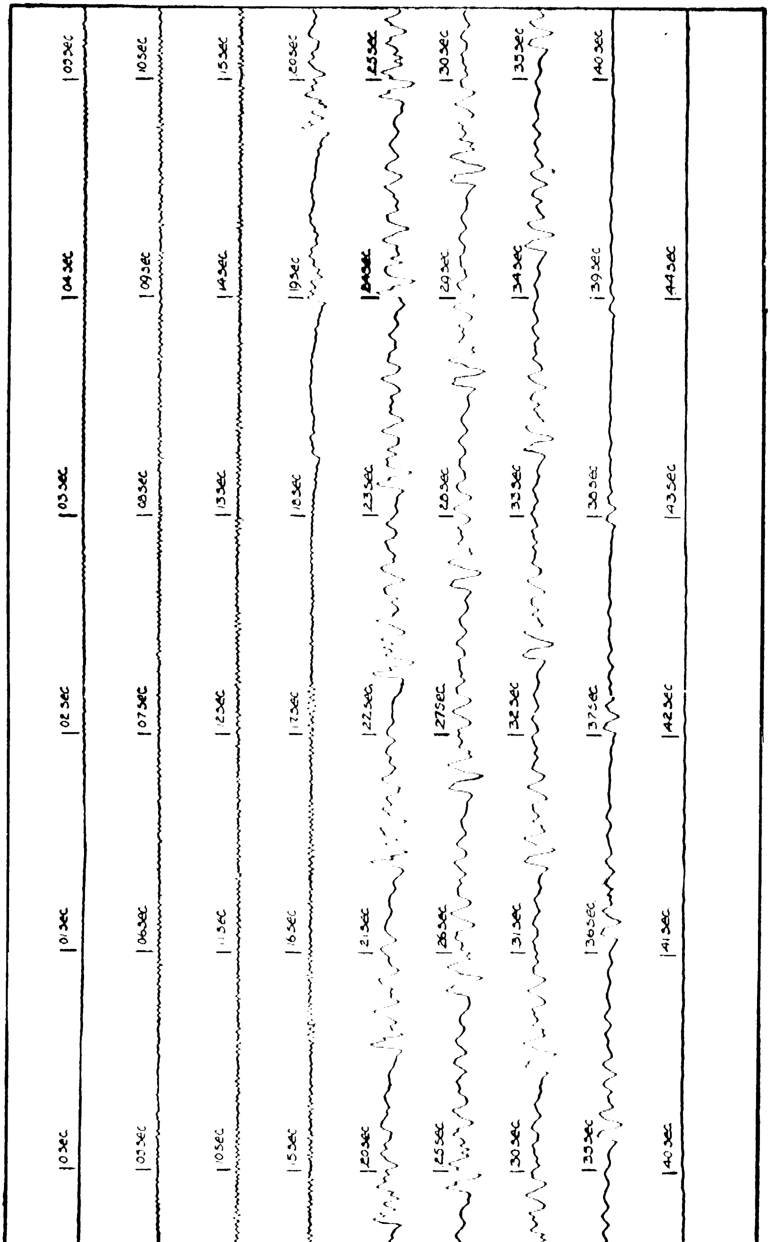


FIG. 32.—"SA." SPOKEN BY M.B.

obtained from them would be the same and the ear under most circumstances would identify them as the same sound.

In Fig. 37 the graphs of four vowel sounds are shown. The first two correspond to "a" as in "father" but pronounced at different pitches, the first by a man and the second by a woman. The wave pictures of these sounds are quite

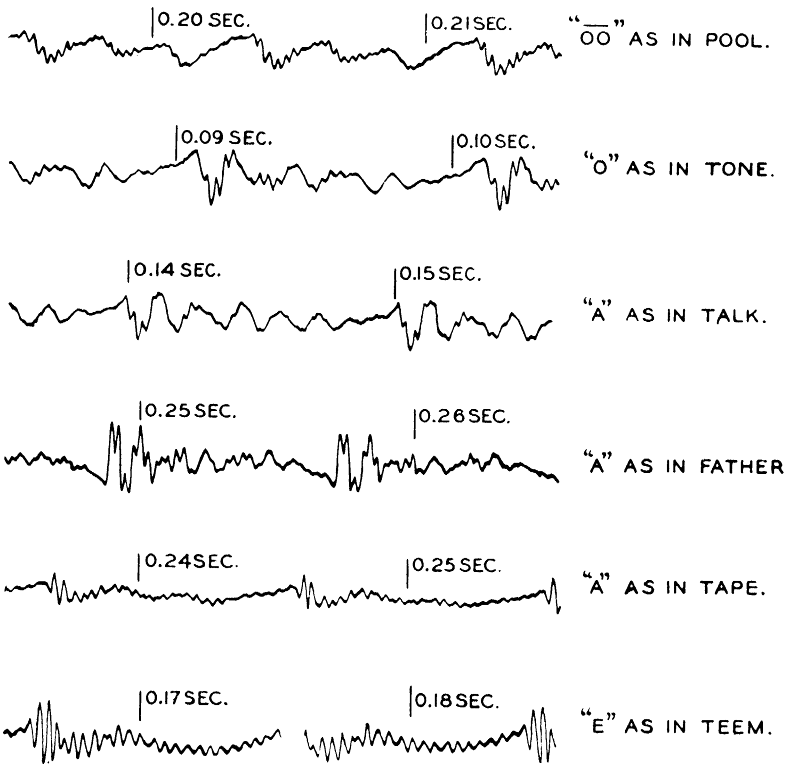


FIG. 33.—LONG VOWELS.

different yet the ear will identify both of them as the vowel "a" more than 99 per cent of the time. The first and third pictures and the second and fourth pictures look much more alike yet they are never confused by the ear. It is true that the ear recognizes some similarity between the first and third wave forms; they are both male voices and have the same pitch. These illustrations are sufficient to show that the characteristic that identifies the particular fundamental speech sound being spoken is not determined entirely by the form of the speech wave.

Theories of Vowel Production

The question therefore arises as to what characteristics of a speech sound differentiate it from another speech sound.

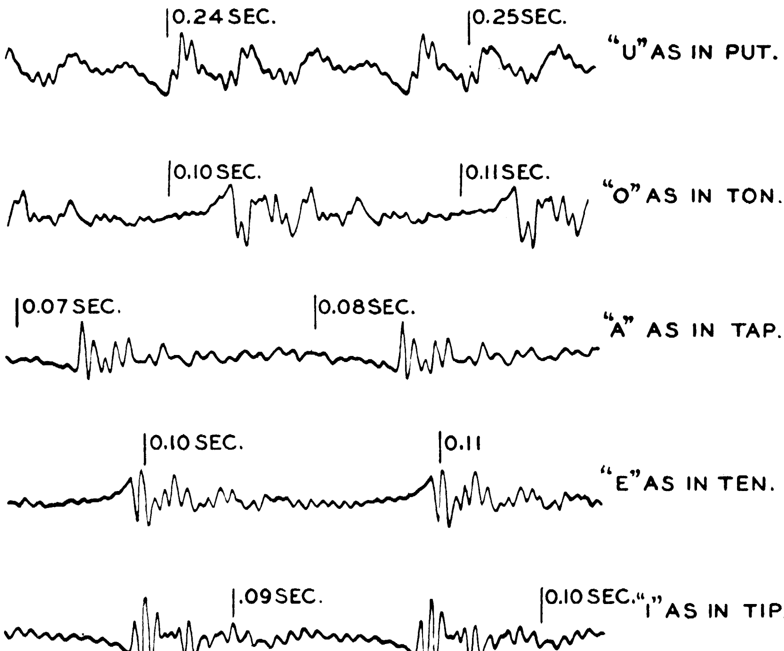


FIG. 34.—SEMI-VOWELS.

It is evident from the preceding paragraphs that the pitch or the wave form of the vowel is not its distinguishing feature. Two theories of vowel production have been advanced, namely, the harmonic or steady state theory and the inharmonic or transient theory. In spite of the fact that Helmholtz [1] showed

[1] Helmholtz, "Sensation of Tone."

that these two theories were different only in the point of view and the method of representing the same mechanism of vowel production, we still have advocates of the two theories.

The harmonic theory was first advocated by Wheatstone in 1837. According to this theory, the vocal cords generate a complex wave having a fundamental and a large number of harmonics. The component frequencies are all exact multiples of the fundamental. As described under the paragraph

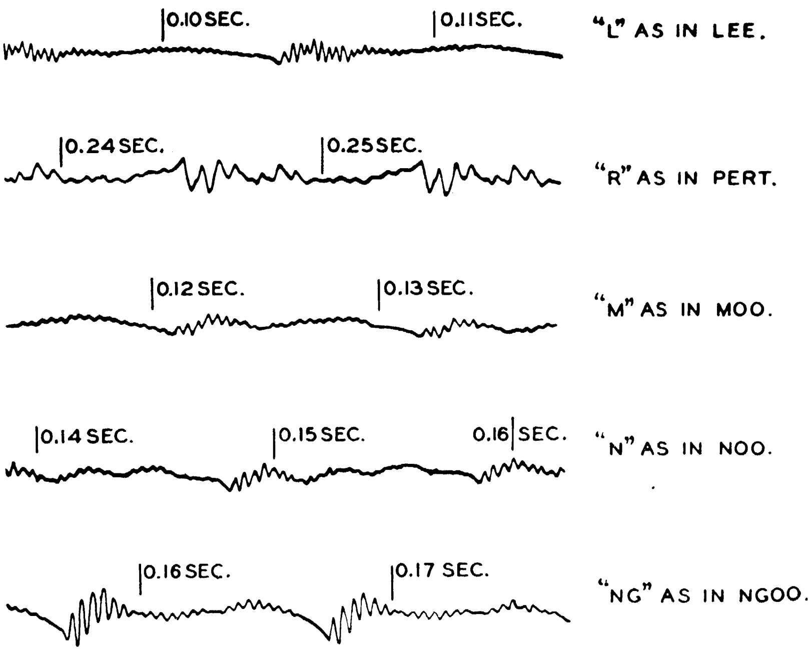


FIG. 35.—SHORT VOWELS.

on the mechanism of speaking, when these waves pass through the throat, the mouth, and the nasal cavities, those frequencies near the resonant frequencies of these cavities are radiated into the air very much magnified, the amount depending upon the damping constant of the cavity. These reinforced frequency regions determine the vowel quality.

According to the inharmonic theory of Willis (1829) and Herman and now advocated by Scripture, the vocal cords act

This is illustrated in the second and fourth pictures of Fig. 37.

It is evident then that in this theory, as well as in the previous one, the vowel quality is dependent upon the natural frequencies and damping of the vocal cavities.

The difference in the two theories is not, as some suppose, a difference in the conception of what is going on while the vowel sounds are being produced, but in the method of representing or describing the motions in definite physical terms. The second point of view enables one to visualize in a more direct way what is taking place and consequently is of greater

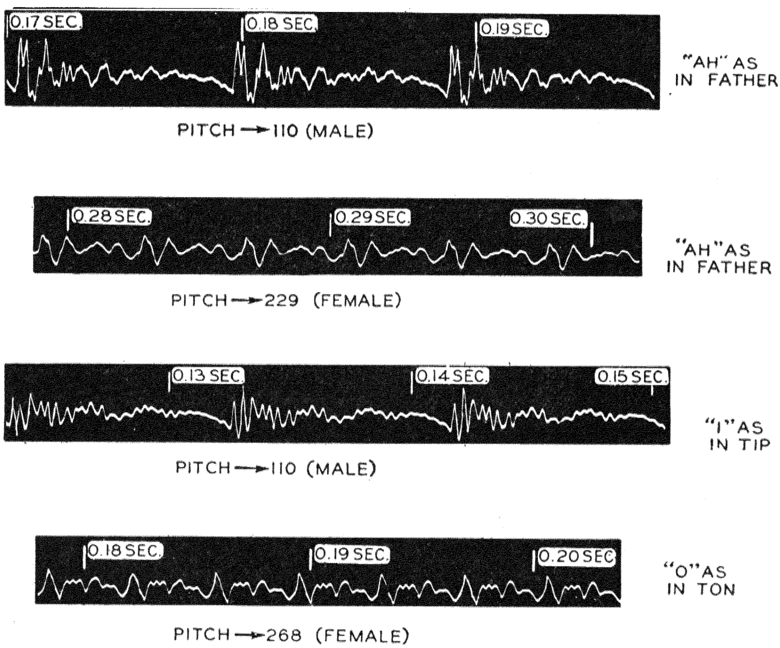


FIGURE 37.

value to the phonetician interested in the mechanism of speech production. It probably enables one better to grasp the fundamental characteristic differences between the vowels.

The first point of view is probably more useful to the engineer who is interested in designing telephone systems to properly transmit speech. The separation of the speech into its component frequencies makes it possible to see quickly only as an agent for exciting the transient frequencies which are characteristic of the vocal cavities. A puff of air from the glottis sets the air in these cavities into vibration. This vibration soon diminishes until it is started anew by a second

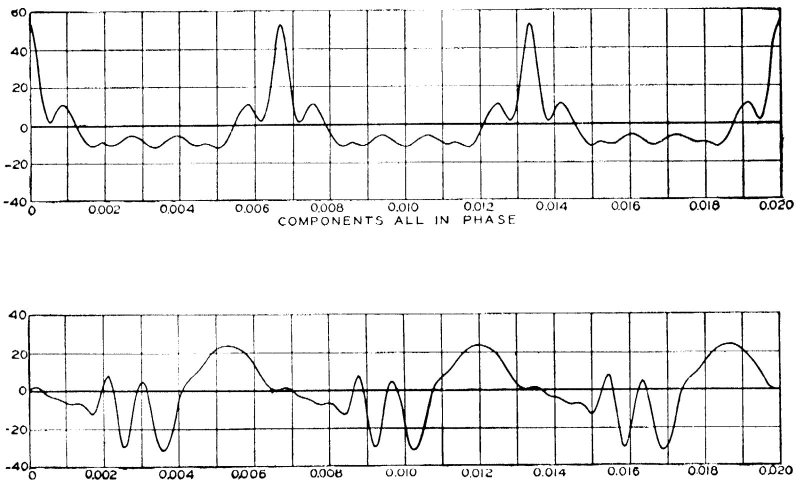


FIGURE 36.

puff. According to this theory, the puffs do not necessarily follow each other periodically and hence the name "inharmonic." However, it is hard to see how the physical mechanism in the throat can produce anything but fairly regular puffs since these are controlled by the elastic properties of the vocal cords and the two resonant columns of air on either side of them. An examination of the records of speech sounds shows that this is true. The different waves succeed each other quite regularly. On the other hand this examination also supports the view that these regular puffs do excite the transients of the mouth and throat cavities, for the amplitudes are large at the beginning of the wave and gradually die away toward the end. This is shown on the records and particularly on the three records in Figs. 21, 24, and 25. When the pitch is high, the natural vibrations do not have time to die down before another pulse sets them going again.

which frequencies must be transmitted by the system to completely carry all the characteristics of speech. A numerical example may help to make this clear. If the force which is acting on the resonant cavity of the mouth due to the vibration of the vocal cords is designated by $F(k)$ for the kth component, and f_0 is a natural frequency of the mouth chamber, f the fundamental frequency of the sound, and Δ the damping constant, then it can be shown that as a first approximation the amplitude of the kth component of the pressure wave radiated into the air from the mouth is

$$A_k = \frac{CF(k)}{\sqrt{\left(k - \frac{1}{k}\frac{f_0^2}{f^2}\right)^2 + \left(\frac{\Delta}{\pi f}\right)^2}}, \tag{1}$$

where C is a constant which determines the scale used for representing the amplitudes. The function $F(k)$ varies from sound to sound and with different individuals, but for purposes of illustration it will be assumed that

$$F(k) = \frac{1}{k^{3/2}}. \tag{2}$$

This assumption seems to give results which correspond roughly to the experimental results obtained with a typical speaker. Typical values for Δ and f_0 for the sound "a" as in "father" are 500 and 900, respectively. If the sound is spoken with the fundamental $f = 125$, then the amplitudes computed from this formula are shown in the top chart of Fig. 38. When the same sound is pronounced at a pitch corresponding to $f = 250$, the amplitudes are as shown in the bottom chart of this figure. Such representations of the relative amplitudes of the different frequency components are called acoustic spectra.

According to the inharmonic theory, it is sufficient to say that the sound "a" as in "father" is characterized by a resonant frequenty of 900 and a damping constant of 500 and that the air in the mouth cavity in this condition is set into vibration by puffs from the vocal cords.

According to the harmonic theory, although necessary to give these two numbers as characteristics of the sound, it is necessary in addition to specify the kind of exciting force that gives the values of $F(k)$.

The acoustic spectra of the most important vowel sounds are shown in Figs. 39 and 40. These spectra were obtained from typical wave pictures taken with the high quality oscillograph. Only the steady-state part of the wave was analyzed.[1] Maxima similar to those in the two calculated cases are plainly evident in these charts. Those spectra shown in

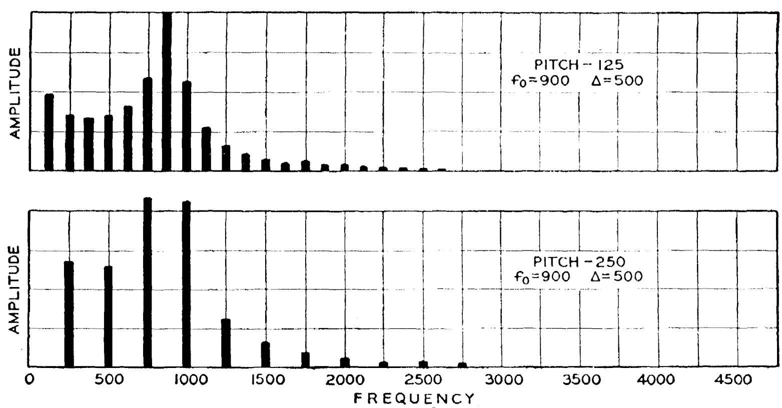


FIGURE 38.

Fig. 39 have one principal region of resonance with indications of one or more regions of less importance while those in Fig. 40 have two principal regions of resonance with other smaller ones. It is well to emphasize here the fact that these charts represent the results obtained with typical voices. When the records of several speakers are analyzed, quite different acoustic spectra are obtained, but in general the regions of maximum amplitude are approximately the same.

In order to show the effect of pitch upon the acoustic spectra of vowel sounds, an analysis was made of vowels intoned at pitches corresponding to the notes of the major

[1] These acoustic spectra were computed by W. Koenig of the American Telephone and Telegraph Company.

chord, namely, at frequencies 128, 160, 192, and 256. The resulting spectra for ē and ā are shown in Figs. 41 and 42. It will be noticed that for the sound ē the frequency regions 300 and 2300 cycles and for the sound ā the regions 500 and 1900 cycles are magnified. For obtaining an analysis of these sounds they were recorded on phonograph records by the new

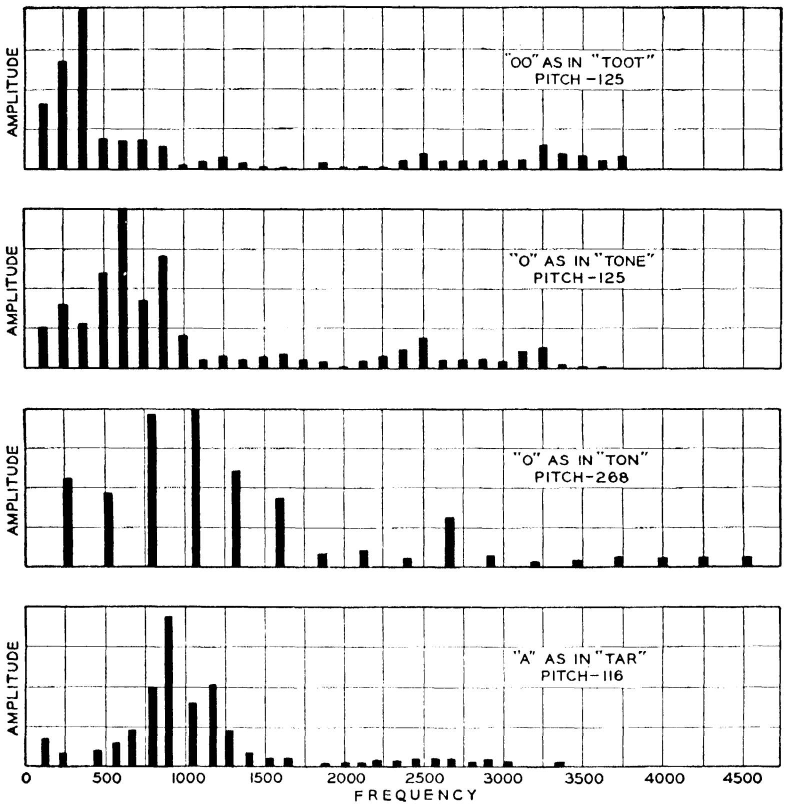


FIGURE 39.

electrical process and then analyzed by means of the electrical harmonic analyzer which is described in Part II, Chapter I. For making the analyses, two electromagnetic reproducers were arranged to play simultaneously on the same record. A key in the electrical circuit was arranged to switch from one reproducer to the other. One of the reproducers was returned to the beginning of the cut while the other was playing, thus making it possible to keep a continuous tone on the input of the analyzer. By means of this arrangement, it was possible to keep the vowel sounding for the five-minute period required to make the analysis.

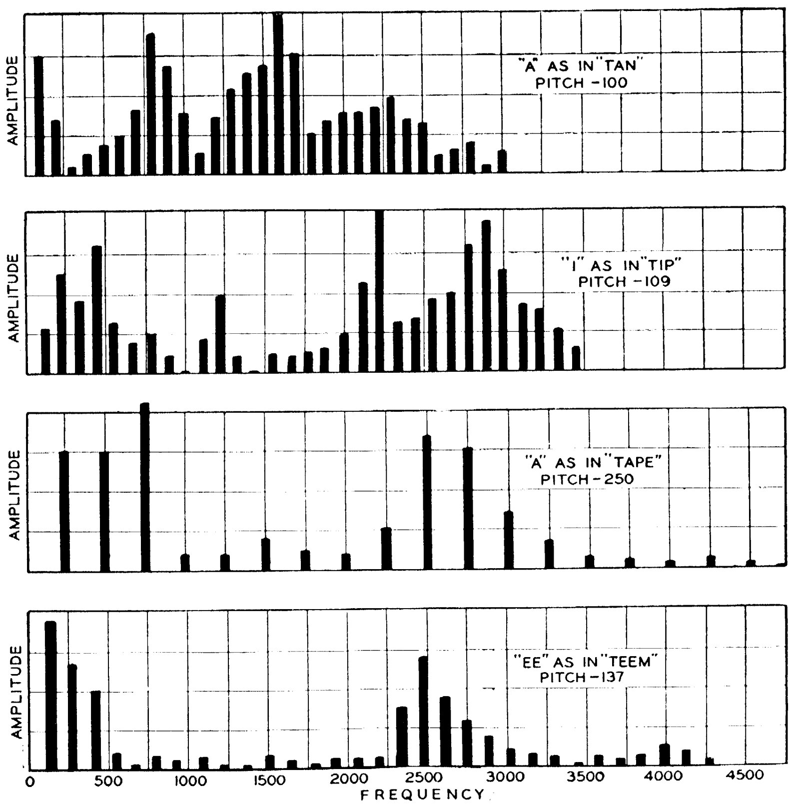


FIGURE 40.

General Characteristics of Speech

The pitch of the voice when speaking the vowels varies with different individuals, corresponding to about 90 cycles per second for a very deep-voiced man and to about 300 cycles per second for a high shrill-voiced woman. The average pitch used by a woman is near middle C or 256 cycles per second. The oscillograph records show that there is usually, although not always, a rise in pitch as the sound progresses. Speaking concerning the general characteristics as deduced from these records, Crandall [1] says:

"Consider now the general properties of the spoken vowel sound,

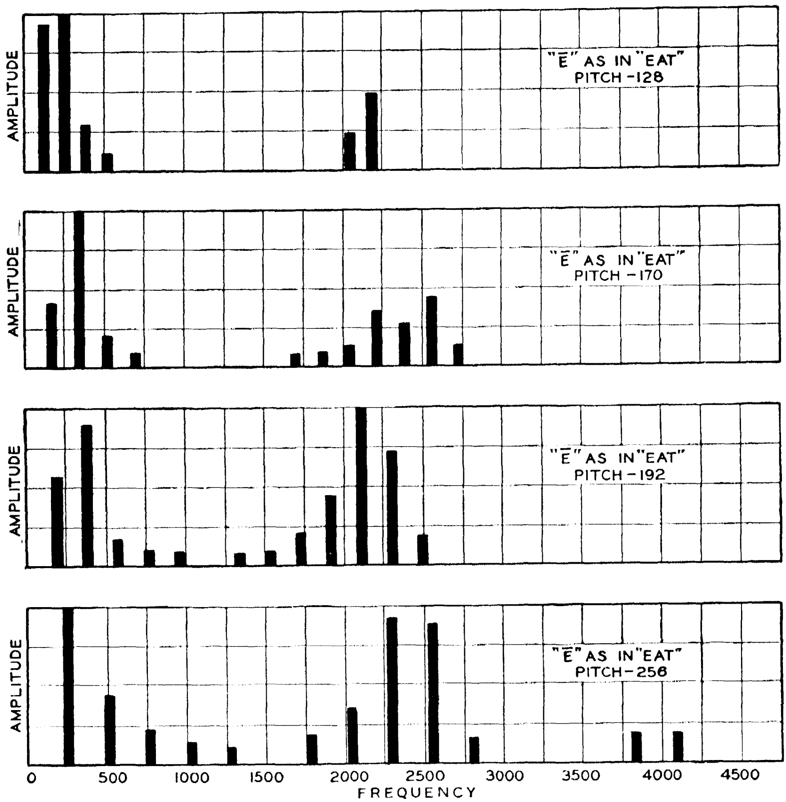


FIGURE 41.

as deduced from these records. First there is a period of rapid growth in amplitude, lasting about 0.04 second, during which all components are quickly produced, and rise nearly to maximum amplitude; second, the middle period, the characteristics of which have been noted, lasting about 0.165 second, followed by the period

[1] *Bell System Technical Journal*, October, 1925.

of gradual decay lasting about 0.09 second, bringing the total length to approximately 0.295 second. There is a tendency to short duration among the 'short' vowels (e.g., short o, e, i) and a tendency to longer records among the broader sounds, as might be expected.

"The behavior of the fundamental frequency (or 'cord tone') during the course of the record will follow normal or individual characteristics as has been described.

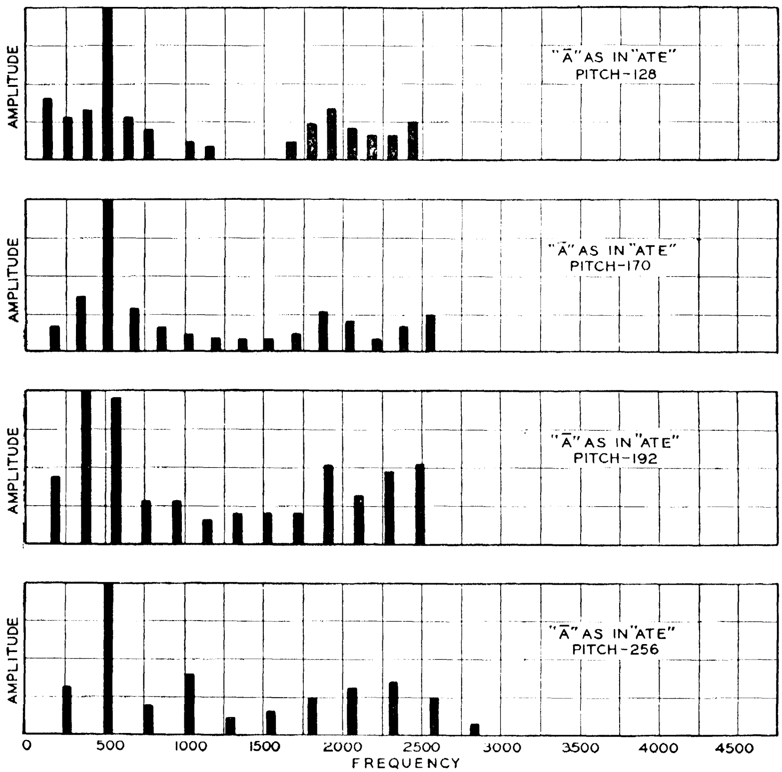


FIGURE 42.

"The low frequency characteristic appears early, usually before the fourth cycle (for men) or before the seventh (for women) and normally is in harmonic relation with the fundamental. In the eleven pure vowel sounds this point was examined at 264 locations in 88 records with the result that the harmonic relation obtained in at least 214 cases. On the other hand the normal behavior of

13238

TABLE II

STATISTICAL DATA FROM 104 RECORDS OF VOWEL SOUNDS

Sound	Duration				Mean Fundamental Frequency		Mean Low Characteristic Frequency		Scattered Low Frequency		Mean High Characteristic Frequency		Scattered High Frequency	
	Start	Middle	Decay	Total	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female
I *oo* (p*oo*l)	.061	.164	.126	.351	140	270	411	581	750 (1)	1200 (1)	......	......	3700 (4)	4412 (4)
II *u* (p*u*t)	.057	.115	.077	.249	138	250	457	691	988 (4)	1100 (3)	......	......	3637 (4)	4250 (4)
III *o* (t*o*ne)	.053	.139	.133	.325	116	237	520	729	830 (3)	1112 (4)	......	......	3475 (4)	3700 (4)
IV *a* (t*a*lk)	.034	.191	.065	.290	112	243	722	801	950 (2)	1150 (2)	......	......	3612 (4)	4075 (4)
V *o* (t*o*n)	.046	.179	.061	.280	118	253	654	854	1100 (4)	1188 (4)	......	......	3212 (4)	3353 (3)
VI *a* (f*a*ther)	.029	.199	.078	.306	113	234	955	1036	1150 (2)	1425 (2)	......	......	3683 (4)	4200 (3)
VII *ar* (p*ar*t)	......	......	......	.345	110	231	630	701	917 (3)	1012 (4)	1965	2162	3800 (2)	4150 (1)
									Note 1		Note 1			
VIII *a* (t*a*p)	.038	.180	.076	.294	123	232	796	960	......	......	1900	2165	3150 (3)	3175 (2)
IX *e* (t*e*n)	.034	.119	.066	.219	121	247	612	775	......	......	1800	2000	2925 (4)	2925 (4)
											Note 2		Note 2	
X *er* (p*er*t)	......	......	......	.331	131	239	570	712	......	......	1688	2188	3050 (2)	3500 (1)
XI *a* (t*a*pe)	.042	.172	.091	.305	125	235	494	614	......	......	3000	2800		
XII *i* (t*i*p)	.036	.126	.049	.211	137	253	450	523	......	......	2950	2962		
XIII *e* (t*e*am)	.036	.189	.116	.341	136	252	296	332	......	......	2987	3266	......	4800 (1)
Means or "Normals"	.042	.161	.085	.288 (11) .296 (13)	125	244								

NOTE 1—Both of these sets of frequencies must be characteristic of *ar*.

NOTE 2—The high frequency characteristics are less definitely located, for short *e*, than for any other doubly resonant vowel sound. The two sets of frequencies given above define a band of frequencies centered about 2400 cycles within which the characteristic high frequency must be contained.

the amplitude of the low frequency characteristic suggests the decay of a transient oscillation during each fundamental cycle—this effect being noticeable in at least 64 of the 88 pure vowel records. This transient effect was also noticeable in 13 of the 16 records of ar and er, where the harmonic effect was not so noticeable.[1] The appearance of the transient effect depends to some extent on the relative frequencies of the fundamental and the characteristic; where the fundamental period is short (as often in the case of the women's records) there is not sufficient time for decay of the characteristic tone before it receives a new impetus in the next cycle of the fundamental.

"As noted above, all the records contain high frequency vibrations which are of such amplitude that they suggest characteristic frequencies. A general mean of these frequencies would be in the neighborhood of 3200 cycles, and in the case of two records by speaker FC (Group I and Group XIII) the frequency rises to about 5000 cycles. Recalling the usual classification of the vowel sounds into two groups—(1) those of 'single' resonance, placed on the left leg of the triangle, and (2) those of 'double' resonance placed on the right leg of the triangle—there are some differences in the behavior of the high frequency components which can be related to these broad classes. In the sounds of the first class the high frequency component is usually small in amplitude, more subject to individual bias in its frequency, and may or may not build up in amplitude as early as the low frequency characteristic. In the sounds of the second class the high frequency characteristic is usually prominent from the start and builds up very rapidly; while there is less variation in its frequency with the individual speaker. In sounds of the first class there is no decided suggestion of a transient in the high frequency while in sounds of the second class the transient effect is pronounced.

"With these considerations in mind there is presented in Table II a summary of the data obtained from this preliminary examination of the vowel records. The mean duration time, and its subdivisions, are shown in the second column for each pure vowel sound, with mean duration only for the sounds ar (Group VII) and er (Group X). The fundamental and characteristic frequencies of each sound are shown in the three columns headed 'Mean Fundamental,' 'Mean Low Characteristic Frequency' and 'Mean High Characteristic Frequency' respectively. Each mean is taken from four records. The two columns headed 'Scattered Low Frequency' and 'Scattered High Fre-

[1] These are very apt to be diphthongal in character which may account for this lack of harmonic effect.

quency' contain mean values of additional components, occurring in one or more records, in certain frequency ranges, the number of records in which such components are noted being shown in parentheses following the mean. The table illustrates and emphasizes many points which have been brought out in the preceding discussion, particularly the closeness with which the high frequency characteristics are defined in the vowels of the second or 'doubly-resonant' class."

After considering the work of Stumpf, Miller, Paget, and Crandall, Table III was constructed which gives the charac-

TABLE III

Characteristic Frequency of the Vowel Sounds

Speech Sound	Low Frequency	High Frequency
ū (pool)	400	800
u (put)	475	1000
ō (tone)	500	850
a (talk)	600	950
o (ton)	700	1150
a (father)	825	1200
a (tap)	750	1800
e (ten)	550	1900
er (pert)	500	1500
ā (tape)	550	2100
i (tip)	450	2200
ē (team)	375	2400

teristic frequency regions of the vowel sounds. Considerable variations from these average values are obtained for different speakers or for the same speaker at different times. The first six of these vowel sounds are frequently referred to as singly resonant and the last six as doubly resonant. In this table two characteristic frequencies are given for all of these sounds. The intensities of the components in the characteristic high frequency range, however, are very much weaker for the sounds in the first than for those in the second group. For this reason the earlier experimenters did not detect them.

In a similar way the records for r, l, ng, m, and n were examined. All of these sounds seemed to have three characteristic frequency regions of resonance, the third being probably due to the intense vibrations of the nasal cavities for these sounds. Table IV gives approximately the resonant frequencies as deduced from Crandall's and Paget's work.

Sixteen consonant sounds were also studied by means of these records. A summary of the results as taken from Crandall's article is as follows:

"B/P.—(See Fig. 22.) Both Paget and Miller have noted the essential impulsive quality of these sounds, and have produced them by sudden closing and opening of the mouth of a resonator. Paget considers p to be the more suddenly released, *i.e.*, to have the steeper wave-front. From the records this is not evident; following the voicing period, the b would seem to be more suddenly produced, as judged by the growth in amplitude of the 'a' sound following.

TABLE IV

Sound	Throat Resonance	Nasal Resonance	Mouth Resonance
r	500–700	1000–1600	1800–2400
l	250–400	600	2000–3000
n	200–250	600	1400–2000
ng	200–250	600	2300–2600
m	250–300	600	900–1700

"D/T.—(See Figs. 22, 23, and 24.) For these (see either Table V or the records themselves) we note a high-frequency characteristic of about 4000 cycles. Paget observed 'an upper resonance 5 to 8 semi-tones higher than that of the associated vowel, and a low resonance of about 362.' We note in the records a low frequency of the order of 500 in the case of d. Paget notes a 'greater amplitude in t due to higher air pressure' and the records show a greater amplitude for the high frequency in the case of t, except right at the transition point, where d shows the high frequency of large amplitude. No conclusion can be given as to relative steepness of wave-front, d *vs.* t, because in both cases we note for speaker MB a steeper wave-front than for MA. The difference between d and t may depend entirely on the voicing and on the complicated phenomena at the transition point.

TABLE V

Group XVI—Six Stop Consonants; Transitional dth/th

Plate No.	Sound	Speaker	Consonant Characteristics					Transitional Characteristics				Vowel Fundamental	
			Duration	Near Start		Mid Portion to End							
				Voicing (Fundamental and Harmonics)	High Frequency	Voicing	High Frequency	Low Frequency	High Frequency (Note 6)	No. of Cycles	First Cycle Short	Near Start	Near End
129	ba	MA	.12	90,180	none	90,180	none	700	2700	1	......	100	115
130	ba	MB	.19	100,200	none	92,184	none	700	3100	1	yes	116	107
131	pa	MA	.02	unvoiced	none	unvoiced	2800 (Note 2)	1000	3600	1	yes	100	111
132	pa	MB	.04	(one 60-cycle vibration)	none	(one 60-cycle vibration)	3800	900	3600	1	yes	119	114
133	da	MA	.13	90,180	none	79,158	3800 (Note 3)	500	2800	3	yes	103	115
134	da	MB	.10	98,196	none	98,196	3600	600	3200	2	......	112	109
135	ta	MA	.07	unvoiced	none	(one 100-cycle vibration)	4300 (Note 3)	....	3200	4	yes	104	112
136	ta	MB	.06	unvoiced	none	unvoiced	3600	900	3000	2	yes	120	113
137	ga	MA	.12	100,200,300	none	84,252	1600, 2800 (Note 4)	550	3000	3	......	101	111
138	ga	MB	.10	100,200,300	none	95,190	1400, 4000	600	3600	2	......	112	112
139	ka	MA	.07	unvoiced	none	unvoiced	1500, 4000 (Note 5)	1200	3800	4	yes	109	118
140	ka	MB	.08	unvoiced	none	unvoiced	1600, 4200	1300	4000	4	yes	125	116
141	dtha	MA	.20	83,166	4000 (Note 1)	95,189	4200 (Note 1)	600	3000	2	......	104	116
142	dtha	MB	.18	100,200	2600	100,200	2700	600	2600	4	......	109	107
143	tha	MA	.02	unvoiced	none	unvoiced	none	600	3200	1	yes	110	110
144	tha	MB	.02	unvoiced	none	unvoiced (one 100-cycle vibration)	none	600	3200	1	yes	113	107

Note 1—A trace of these at beginning of the early fundamental cycles. Note 2—One faint transient. Note 3—Transients; longer for ta than for da. Note 4—One transient. Note 5—Irregular transients. Note 6—Possibly due in some cases to the a sound.

"G/K.—(See Fig. 29.) k shows the characteristic transients (1500, 4000; Table V, notes 4 and 5) to much more pronounced degree than g. From the records it would seem that g, in addition to the voicing, disclosed a steeper wave-front, the four transitional cycles required for k emphasizing this point. No other generalizations seem warranted, on account of the complicated series of events recorded. These sounds are treated at length by Paget who observes considerable variation in their resonant ranges, depending on the associated vowel. It will be noted, however, that in these four records, particularly consonant characteristics are persistent and of large amplitude before the vowel sound begins to appear.

"DTH/TH.—The high frequencies (2600, 3000, 3200) culminating at the transition point seem to be the key to these records. They are more persistent for dth, while th appears to show the steeper wave-front. Paget states that 'in δ [dth] the middle resonance is overblown, . . . louder than the corresponding resonance in θ [th].' He gives also an 'upper sibilant of 3444–5950' louder for dth than th, and 'difficult to identify.' It will be noted that in one record for dth there is during the voicing period a faint high frequency which has been set down in Table V as 4000 cycles. This faint 'sibilant' (which may always be audible though it fail to be recorded) establishes a certain kinship between these two sounds and those following (the fricative consonants) which are rich in sibilant sounds.

"V/F.—v shows a pronounced voicing, and previously noted, a less prominent high-frequency component than its partner f, or any of the other fricative consonants. Comparing v/f with dth/th it seems from the records that the former pair are of higher frequency (particularly f) and that for v/f as a unit the higher frequency characteristic is more pronounced; just the opposite conclusion to that reached by Paget. f may indeed differ more from v than v from dth, thus raising difficulties of classification both physically and phonetically, which cannot be resolved on the basis of the few records available. The exceedingly fine distinction between the sounds v and dth could be no more strikingly shown than it is in the records given, for both speakers.

"J/CH.—(See Fig. 30.) Some of the recorded phenomena of this pair suggest correspondences between them and the pair g/k; but the pair j/ch shows a higher frequency characteristic during the important mid-portion of its history. Of the pair, ch seems to show the steeper wave-front, that is, the more rapid transition to the vowel sound.

"ZH/SH.—With this pair we pass to the field of pure sibilants, in which there is no evidence of impulsive action or steepness of

TABLE VI
Group XVII—Fricative Consonants

Plate No.	Sound	Speaker	Consonant Characteristics					Transitional Characteristics				Vowel Fundamental	
			Duration	Near Start		Mid-portion to End							
				Voicing (Fundamental and Harmonics)	High Frequency	Voicing	High Frequency	Low Frequency	High Frequency (Note 3)	No. of Cycles	First Cycle Short	Near Start	Near End
145	va	MA	.20	97,195,390	3000	87,174	none	600	2700	3	....	101	116
146	va	MB	.25	112,224	3200 (trace)	100,200	none	600	3400	2	....	112	107
147	fa	MA	.15	unvoiced	3100	unvoiced	3500, 7000	500	2800	4	yes	112	121
148	fa	MB	.30	irregular	3200, 6400	unvoiced	3200, 6400	600	3600	3	yes	111	104
149	ja	MA	.22	81,243	3400	81,162	2600, 5200	450	2700	4	....	110	110
150	ja	MB	.14	trace	3300	90,179	2000, 4800	500	3100	4	....	115	111
151	cha	MA	.07	unvoiced	4800	unvoiced	2800, 4800	{500, 1500}	3000	2	yes	104	111
152	cha	MB	.08	unvoiced	3600	unvoiced	3600, 6400	{500, 1600}	trace	2	yes	119	115
153	zha	MA	.28	86,172,344	3000, 4000 (Note 1)	87	3000, 4000 (Note 1)	450	2900	4	....	100	111
154	zha	MB	.13	96	2600, 4200	99	3000, 4200	{500, 2000}	....	4	....	114	111
155	sha	MA	.18	unvoiced	2800, 3600 (Note 2)	unvoiced	2800, 4600 (Note 2)	450	3200	3	yes	104	104
156	sha	MB	.17	unvoiced	2200, 5000	unvoiced	2600, 500	{500, 1800}	2800	3	yes	117	112
157	za	MA	.24	96,384	2800, 5600 (Note 1)	89,178	5200, 7000 (Note 1)	400	3100	4	....	98	108
158	za	MB	.22	100,300	2200, 4400	100,200	2800, 5600	550	2800	5	....	111	107
159	sa	MA	.27	unvoiced	5600, 8000	unvoiced	6000, 7800	500	2900	2	yes	114	114
160	sa	MB	.19	unvoiced	4000, 6400	unvoiced	4200, 6600	650	2900	2	yes	117	108

Note 1—Alternating; lower frequency in first part of fundamental cycle, higher frequency in latter part of cycle.
Note 2—Alternating, irregularly.
Note 3—Possibly due to the a sound.

wave-front. The action seems to be that in the voiced sound there is, in addition to the presence of the fundamental tone, a breaking up of the characteristic high frequency wave-train into discrete units corresponding to the fundamental tone, whereas in the unvoiced sound the high frequency characteristic is continuous, though irregular. Thus noting that the characteristic frequency is of 3000 to 4600 cycles the outstanding phenomena of zh/sh are well defined. In addition to frequencies of 2048–3249 noted by Paget, he gives a 'pronounced middle resonance of 1625–2048.' This latter observation of Paget's may correspond to the 1800–2000 frequency in the records of MB in the transition region, but this component does not seem to be prominent in the records.

Z/S.—(See Figs. 31 and 32.) The general properties of these sounds can be inferred from the discussion of the preceding pair (zh/sh), adding only the fact that their principal characteristic is of much higher frequency. From Table VI we note a range of 4200–8000 cycles; Paget gives 'a characteristic upper resonance of 5790–6886.' Paget also gives 'a middle resonance of 1084–2298.' The records do not show as low a range of characteristic frequencies unless it be the frequency range 2200–2800 (see note 1, Table VI), within which fall certain vibrations occurring in the early parts of the fundamental cycles of the voiced sounds zh and z. The true s sound is, as Paget has stated, 'a relatively complex hiss' and this is true of sh as well. And to complete the record, we must observe that zh and z are even more complex, if possible, and thus not inappropriate examples of the sounds of speech with which to conclude this survey."

CHAPTER III

Speech Power

Definition of Instantaneous, Average, Mean, Syllabic, Phonetic, and Peak Speech Powers

For purposes of engineering telephone transmission systems, it is desirable to know both the acoustic and electrical power of the speech being transmitted. If this power becomes too small, it is masked by extraneous noise. If it becomes too high, parts of the transmitting apparatus become overloaded; that is, they fail to properly transmit the speech. Inasmuch as speech power is so variable, it has been found convenient to use several quantities in describing it, such as the instantaneous, the average, the mean, the syllabic, the phonetic, and the peak speech powers.

The instantaneous power is the rate that the sound energy is being radiated at any instant, and it frequently rises to values h gher than one hundred times the average power.

The average speech power is the total speech sound energy radiated while a person is speaking divided by the time interval during which he speaks.

It is of interest to know the slow variations of the speech power such as would be recorded by the usual type of voltmeter or ammeter when placed in the telephone circuit over which speech is being transmitted. For describing these variations the notion of mean power is useful. The mean power is a function of the time which shows the slow variations of the speech power without showing the periodic fluctuations of the wave. It can be determined from the wave-form pictures such as are shown in Chapter II. To do this the average power over each one-hundredth second is calculated. Then a curve is plotted using these short interval averages as ordinates and the corresponding time values as abscissas. Such a curve for a syllable, a word, a sentence, or a series of sentences gives a mean speech power curve and indicates to the eye the variations in loudness as sensed by the ear.

The syllabic speech power is useful in describing the power used in various syllables. Inasmuch as it is difficult to determine the exact beginning or ending of a syllable, the maximum mean power attained while it is spoken is taken as a measure of the syllabic power.

The phonetic speech power is the maximum value of the mean power while one of the fundamental vowel or consonant sounds is being spoken. It is useful in comparing the relative amounts of power used in producing the different phonetic sounds. The syllabic power is usually the phonetic power of the vowel in the syllable.

The peak speech power is the maximum value of instantaneous power during the interval considered. In Fig. 43 is shown a chart [1] which illustrates these kinds of speech power for the word "quite." The instantaneous power varies from zero to high values for each cycle of the wave. The mean power slowly rises to about 40 microwatts, which is the syllabic power; it then decreases to zero. The peak power rises to 1500 microwatts. The phonetic power of the sound ī is 40 microwatts or the same as the syllabic power.

Speech Power Definitions

All Measured in Microwatts

Instantaneous S.P.	—Rate sound energy is being radiated at any instant.
Average S.P.	—Total speech energy radiated over any period divided by the length of the period.

[1] Taken from article by C. F. Sacia, "Speech Power and Energy," *Bell System Technical Journal*, October, 1925.

Mean S.P.	—Average S.P. over each one-hundredth of a second period.
Syllabic S.P.	—Maximum value of mean S.P. of one syllable.
Phonetic S.P.	—Maximum value of mean S.P. of one fundamental vowel or consonant.
Peak S.P.	—Maximum value of instantaneous power over the interval considered.

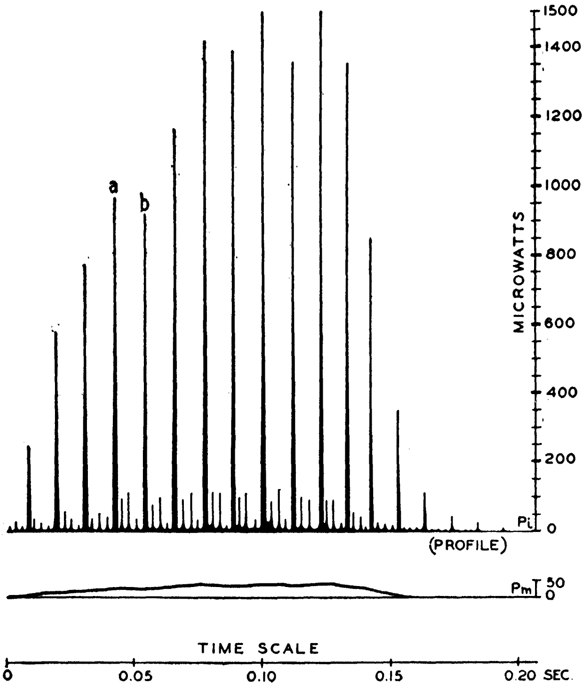


FIG. 43.—ENLARGED COPY OF ORIGINAL OSCILLOGRAM OF THE WORD "QUITE."

Average Speech Power

The speech power may be most conveniently measured by means of a calibrated condenser transmitter. A description of the method of calibrating a condenser transmitter by means of a thermophone is given in Appendix A. From the voltage developed at its terminals, the pressure on its diaphragm can be calculated. It may be shown (see Appendix B) that during the passage of a sound wave through any medium the power J in ergs per second radiated through a unit area of the wave front is given by

$$J = \frac{p^2}{r} \tag{1}$$

where p is the r.m.s value of the pressure variation expressed in bars (dynes per square centimeter) and r is the radiation resistance of the medium. The value of the radiation resistance is equal to the velocity of propagation of the sound wave multiplied by the density of the medium in which it is travelling. When r is expressed in c.g.s. units its value for air at 20° is 41.5; for water it is 143,000. It is evident then that a sound wave in water must produce fifty-eight times as much pressure variation to carry the same sound energy as a similar sound wave in air. For speech waves in air at 20° the intensity J in microwatts per square centimeter is given by

$$J = \frac{p^2}{415} \tag{2}$$

where p is expressed in bars.

The power passing through a square centimeter at a convenient distance from the mouth can thus be determined. Multiplying this by the area of a hemisphere having a radius equal to this distance from the mouth gives approximately the total power of the speech sounds.

By taking the average speech power for a number of individuals talking in their usual conversational manner, it has been found that the average speech power for American speech is approximately 10 microwatts. If the silent intervals during conversation are excluded, this average is increased approximately 50 per cent. To carry this amount of power the air particles near the mouth vibrate through a distance of the order of $\frac{1}{100}$ millimeter. When this amount of sound energy is received directly into the ear, it seems rather large due to the large excitation it produces on the auditory sense.

However, it is really very small in comparison with the other powers ordinarily encountered. For example, it takes power equivalent to that produced by more than one million voices to light an ordinary incandescent lamp. It is therefore evident that the electrical currents used to transmit speech are of a different order of magnitude than those used to transmit power for lighting and heating purposes. It is only in some of the larger broadcasting stations that electrical speech currents are comparable in size with those used in power work.

When one talks about as loudly as possible the average speech power increases approximately to 1000 microwatts. When one talks in as weak a voice as possible without whispering it drops to .1 microwatt. A very soft whisper is about at .001 microwatt.

Tests made with a large number of speakers talking into the transmitter of a commercial telephone system showed that variations in the characteristic average speech power used by different individuals are approximately as shown in Table VII. The values [1] are given in fractions or multiples of the average speech power. These figures are based upon the electrical powers flowing from the terminals of the transmitter while it is being agitated by the various speakers.

TABLE VII

RELATIVE SPEECH POWERS USED BY INDIVIDUALS IN CONVERSATION

Region of average speech power.....	below $\frac{1}{16}$	$\frac{1}{16}-\frac{1}{8}$	$\frac{1}{8}-\frac{1}{4}$	$\frac{1}{4}-\frac{1}{2}$	$\frac{1}{2}-1$	1–2	2–4	4–8	above 8
Per cent of speakers	7	9	14	18	22	17	9	4	0

Decibel—Sensation Unit—Sensation Level

When discussing problems involving mainly differences in powers, either electric or acoustic, it is convenient to use a

[1] These values were obtained by L. J. Sivian of Bell Telephone Laboratories.

logarithmic unit called the decibel and recently adopted for use in telephone engineering. If J and J_0 are the two different amounts of power being compared, then the difference in power level α expressed in bels is given by the equation

$$\alpha = \log_{10} \frac{J}{J_0}. \tag{3}$$

A unit which is one-tenth of the bel is a more convenient size for most practical work and is the one used in telephone engineering in this country. It is called the "decibel" and is usually designated by the abbreviation "db." For example, in the discussion on speech power, if the average speech power is taken as a zero level, that is, taken as the comparison level, then the level of very loud speech would be + 20 db, of weak speech − 20 db, and of a soft whisper − 40 db. The range of levels from a soft whisper to very loud speech is 60 db. From the figures shown in Table VII it is seen that using this unit, the range of average powers used in conversation by 93 per cent of speakers is about 21 db.

This unit is also used by otologists, psychologists, and physiologists in describing the magnitude of sounds being listened to by the ear. It is a well-known psychological law that equal steps on such a logarithmic scale sound *approximately* like equal loudness steps. A change of the power level of a sound by one decibel is approximately the smallest that the ear can detect. When this unit is used in this connection the term "sensation unit" has come into use. However, since the telephone companies of both Europe and America have adopted the names bel and decibel, it seems desirable that these names be used universally if possible. Unless otherwise stated, the decibel will be used in this book for representing levels of intensity.

The sensation level of any sound reaching the ear is the number of sensation units it is above the threshold level for audition. If the lips of an average speaker are held within half an inch of the ear of a person having normal hearing, the sensation level of the speech received is about 100 db.

The intensity of the speech or the speech power per unit area which actuates the ear under such circumstances is called the initial speech intensity. Although, as mentioned before, the average speech power is only one-millionth that used for an electric light, it can still be attenuated 10^{10} times or to one ten-billionth before it ceases to affect the auditory sense. This matter will receive further consideration in a later chapter.

Power in the Fundamental Speech Sounds

In the course of conversation the fundamental vowel and consonant sounds are produced with varying degrees of power depending upon their position in the sentence and the emphasis desired. In spite of this variation some of the speech sounds are always much more powerful than others and it is interesting to know typical values used in conversation.

The phonetic and peak powers of the individual speech sounds can be obtained by means of a calibrated condenser transmitter. Sacia and Beck [1] have obtained in this way from measurements of oscillograms some values for most of the speech sounds. Although sixteen people were used in obtaining these data and the sounds made in various combinations, they are still insufficient to give average values which can be said to be typical. However, the values obtained do give a good notion of the range of powers involved. These data are given in Table VIII under the columns headed "Phonetic Power" and "Peak Power." The figures are in microwatts power radiating from the mouth of the speaker.

As a check against the results obtained by this method there is given in the last column of Table VIII a set of figures which are the average values of two other methods used for determining phonetic power. These other methods are described in Part IV, Chapter IV, and will be only briefly outlined here. The first method uses a form of articulation test, taking advantage of the fact that as the speech level is decreased the weaker sounds will be the first to be misunder-

[1] *Bell System Technical Journal*, July, 1926.

stood since the ear fails to hear their essential characteristics first. By recording the amount of attenuation necessary to

TABLE VIII

POWER IN MICROWATTS IN THE FUNDAMENTAL SPEECH SOUNDS

Phonetic Sound	Key Word	Phonetic Power		Peak Power		Calculations from Threshold and Articulation Measurements
		Average	Maximum	Average	Maximum	
ū	tool	23	60	235	700	38
u	took	26	100	470	890	50
ō	tone	25	80	435	1300	74
o′	talk	45	120	615	1500	87
o	ton	24	110	450	1700	83
a	top	41	120	700	1600	68
á	tap	25	90	650	1800	57
e	ten	22	90	500	1700	34
ā·	tape	23	60	525	1700	35
i	tip	20	50	350	1300	22
ē	team	20	80	310	1500	16
m	me	1.8	17	110	200	2.9
n	no	2.1	18	47	70	4.1
ng	ring	·3	3·6	97	170	12
l	let	·3	9·6	130	230	18
r	err	16	30	200	600	33
v	vat	.03	2.4	25	30	1.0
f	for	.08	3·6	3	4	1.0
z	zip	·7	7·2	30	40	1.2
s	sit	·9	8.7	30	55	·9
th	thin	....	....	1	1	·3
t͟h	that	....	....	9	10	2.3
zh	azure	....	....	40	55	
sh	shot	1.8	6.0	110	130	11
b	bat	....	....	7	7	1.1
p	pat	....	....	6	7	1.0
d	dot	.08	2.9	4	7	1·7
t	tap	.1	6.0	16	19	2.7
j	jot	·5	3·6	24	36	4·1
ch	chat	1.4	19	52	60	6.1
g	get	....	....	8	9	3·3
k	kit	·3	4·8	6	9	3·0

reduce each speech sound to the point where it is misunderstood some arbitrary per cent of the number of times uttered, it is possible to obtain approximately the relative power of each sound.

The second method reduces each sound to the level at which it may no longer be heard. This is done by using a telephone circuit giving a very faithful reproduction of the speech sounds and into which is introduced suitable attenuators. The number of db that each sound must be attenuated to make it inaudible, is thus a measure of its phonetic power.

Table IX shows the individual results of these last two methods and also their averages. The figures in the last column of Table VIII were obtained from the averages of these two methods by use of the "threshold" or minimum audible sound pressures which are determined as described in Part III, Chapter II. Here it is shown that the minimum pressure at from one to three thousand cycles is approximately .0006 bar. Substituting this value in equation (2) gives

$$J = \frac{(.0006)^2}{415} = 8.7 \times 10^{-10}. \tag{4}$$

The area of the hemisphere through which the sound is passing is approximately 10 sq. cm. when the lips of a speaker are $\frac{1}{2}$ inch from the ear. Consequently, the power of this minimum sound is approximately 87×10^{-10} microwatts. Having, from the "average" column of Table IX, the amounts the various sounds must be attenuated to reach this minimum level, a simple calculation gives the phonetic power of each at normal levels and these are the values listed in the last column of Table VIII.

After considering the different sets of data, Table X was constructed which gives the relative amounts of power in the different sounds using the power in the faintest sound as a basis of comparison.

It is seen that the most powerful sound is ó (awl), and the faintest sound th (thin), the ratio of powers between these two being 680. The difference in level expressed in db corresponding to this figure is 28. From the data available the indications are that in an average room in the city the noise is such as to raise the threshold approximately 30 db. Also

TABLE IX

SENSATION LEVELS PRODUCED BY AN AVERAGE SPEAKER FOR THE FUNDAMENTAL SPEECH SOUNDS

Speech Sound	Threshold	Articulation	Average
ó (talk)	100.0	100.0	100.0
o (ton)	99.6	100.0	99.8
ō (tone)	99.6	98.9	99.3
ī (bite)	99.5	100.0	99.8
ou (bout)	99.2	100.0	99.6
á (tap)	99.2	97.2	98.2
e (ten)	98.4	93.5	95.9
a (top)	97.4	100.3	98.9
u (took)	97.1	98.1	97.6
ū (tool)	95.9	94.3	95.1
ā (tape)	93.3	98.2	95.8
i (tip)	92.6	95.5	94.0
ē (team)	89.4	96.3	92.9
r (err)	96.0	95.5	95.8
l (let)	93.5	92.6	93.1
ng (ring)	88.9	93.8	91.4
sh (shot)	88.9	93.2	91.1
ch (chat)	87.2	89.7	88.5
n (no)	86.8	86.7	86.75
m (me)	85.4	85.1	85.3
t̲h̲ (that)	84.2	.....	84.2
t (tap)	84.1	86.4	85.3
h (hat)	83.9	81.7	82.8
k (kit)	83.8	85.3	84.6
j (jot)	83.7	89.7	86.7
f (for)	83.6	77.7	80.7
g (get)	82.9	86.9	84.9
s (sit)	82.4	78.1	80.3
z (zip)	81.6	81.6	81.6
v (vat)	81.4	80.1	80.8
p (pat)	80.6	81.4	81.0
d (dot)	78.9	87.8	83.4
b (bat)	78.8	83.7	81.3
th (thin)	78.7	71.2	75.0

the sound is attenuated more than 40 db if the speaker is about 10 feet away from the listener. Consequently, under such circumstances the sound th is barely audible.

The pure vowels are the most powerful sounds and have a range of intensity of 3 to 1. As would be expected, the open vowels ó, a, o, and á have the largest phonetic powers. The diphthongs are not given but they have about the same power as the vowels which compose them.

The semi-vowels are next to the pure vowels in phonetic power. Of these, n is the weakest and r the strongest. It is interesting to note that the unvoiced fricatives, sh and ch, have powers comparable to the semi-vowels. Next follow

TABLE X

Relative Phonetic Powers of the Fundamental Speech Sounds as Produced by an Average Speaker

ó	680	ū	310	ch	42	k	13
a	600	i	260	n	36	v	12
o	510	ē	220	j	23	th	11
á	490	r	210	zh	20	b	7
ō	470	l	100	z	16	d	7
u	460	sh	80	s	16	p	6
ā	370	ng	73	t	15	f	5
e	350	m	52	g	15	th	1

the stop and fricative consonants; z, s, t, g, v, and th having about the same power which is about one-fifth that of the semi-vowels and then b, d, p, and f having a slightly lower power.

The syllabic power varies more with the emphasis given than with the vowel sound used. A vowel in an accented syllable has usually three or four times as much phonetic power as one in an unaccented syllable. This difference is dependent upon the speaking habits of the individual.

The peak power varies considerably with the type of voice, the values given in Table VIII being typical. For engineering purposes, it may be considered to be about five times the syllabic power. In this connection Sacia [1] says

"I have become able to associate peak factors with vocal qualities in the following way: the voices with the higher peak factors are those which in the ordinary terminology are said to be 'resonant' or 'vibrant'; they have the greater carrying power, especially over the telephone; they are rich in the musical sense and are, therefore, well suited to singing, although many such voices, unfortunately, are never applied to the art."

It is seen then from Table VIII that for an accented syllable the peak power frequently rises to 700 microwatts. For the 4 per cent of speakers who are in the class producing an average power of from four to eight times the average for the entire group, this peak value might reach as high as 5000 microwatts.

From the data given it is estimated that the average phonetic power in the faintest speech sound, namely, th (as in thin), is approximately .05 microwatt. This value, of course, is for one speaking with a typical average voice. For those speaking with softer voices a much smaller power for this sound would be produced. The reduction, however, would not be as great as for the vowel sounds. A round figure of about .01 microwatt probably represents the faintest sound and of about 5000 microwatts the peak value of the loudest sound that will be encountered in conversation. This represents a range in intensity of 500,000 to 1 or 56 db. When dealing with only one speaker, the range of intensity of the speech sounds is usually between 35 and 40 db.

All of the figures given are based upon average American speakers. Although no measurements have been made upon persons who are specially trained to speak distinctly, such as actors and public speakers, it is very probable that such tests would show that the weaker sound would be given considerably more power by such trained speakers than is ordinarily used by the average speaker. Due to this cause, the range

[1] Sacia, C. F., "Speech Power and Energy," *Bell System Technical Journal,* October, 1925.

of intensities used by such speakers would be narrower than indicated by these figures. However, the greater range of emphasis used by them would tend to make the necessary intensity range wider.

Relative Distribution of Speech Power into Frequency Bands

The frequency range necessary for the faithful transmission of speech is of considerable importance. In Chapter II were

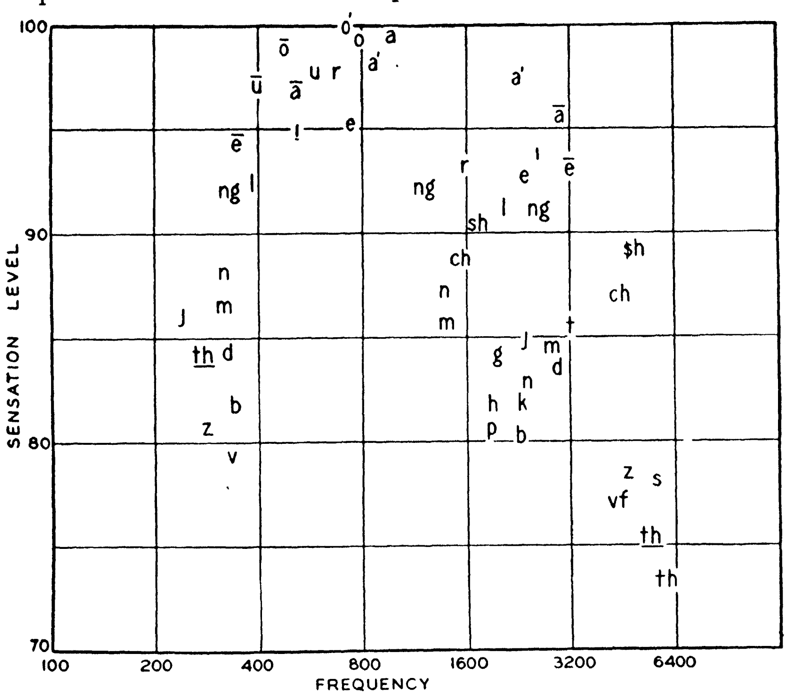


FIGURE 44.

given the characteristic frequency regions, and in the present chapter the characteristic intensity regions for the various speech sounds are given. Figure 44 is a plot showing these combined characteristics of most of the fundamental sounds of speech. The ordinates give the sensation level of the principal components and the abscissas the characteristic frequency regions of each speech sound. When a sound has several principal components the position of each is indicated. Although it cannot be claimed that this chart gives more than a very rough picture of the true facts, it may serve to give a general picture of the intensities and frequencies involved in the transmission of speech.

More accurate data for the frequency-energy distribution for speech as a whole have been obtained by Crandall and MacKenzie.[1] The method consisted of analyzing the speech waves which were impressed upon a condenser transmitter by using a resonant circuit to transmit narrow frequency bands of energy and pronouncing the separate syllables of the connected speech so slowly that the kick of a direct current galvanometer could be separately read for each syllable. The

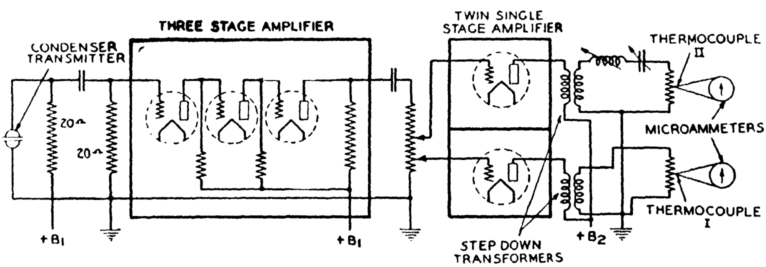


FIG. 45.—CIRCUIT FOR DETERMINING THE FREQUENCY-ENERGY DISTRIBUTION IN SPEECH.

circuit used for this purpose is shown in Fig. 45. The sound waves which fall upon the condenser transmitter are transmitted through the three-stage amplifier and then into the twin single-stage amplifiers. Connected to the output of one of these amplifiers is the resonant circuit which limits the band of frequencies being transmitted. Connected to the other amplifier is a circuit which transmits all the frequencies. By changing the tuning of the first circuit the different bands of speech are transmitted. When a syllable is spoken, simultaneous readings are taken of both meters, one reading corresponding to the total energy of the syllable uttered, the other to the energy of the syllable lying within the limits of trans-

[1] *Physical Review*, March, 1922, pp. 221–232.

mission of the tuned circuit. Twenty-three bands of frequencies were used in the experiment so that each syllable was repeated this number of times. Any variations in loudness of the syllable produced by the speaker could be detected

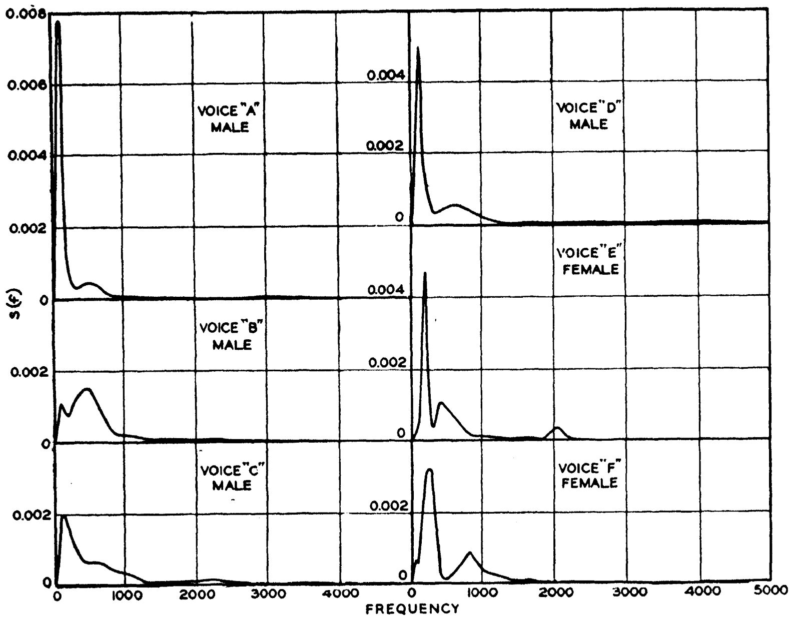


FIG. 46.—ANALYSES OF INDIVIDUAL VOICES.

by the first meter and the necessary corrections made. The properties of the circuit were such that for the different frequency settings the widths of the band were not the same. A correction was therefore made to make the readings for these different widths comparable. After these corrections were made and the data reduced to a comparable basis, the relative amounts of energy in the various regions were computed. Figure 46 shows the results [1] of such an analysis for six voices. The curves are arranged so that energy in any particular frequency band is proportional to the area included between the two ordinates erected at the limits of this fre-

[1] *Physical Review*, March, 1922, p. 227.

quency band, the curve, and the "X" axis. These curves were obtained by each of the six speakers pronouncing the test sentence of fifty syllables for each of the twenty-three frequency settings, making 6900 observations. As would be expected, the energy-frequency distribution for the different voices shows characteristic differences. The curve giving the average of these six voices is shown in Fig. 47.

The region containing the maximum amount of energy is at a frequency near that corresponding to the fundamental pitch ordinarily used in producing the vowel sounds. Although the experiments were carried to frequencies no higher than 5000 cycles, it is well known from other sources that there is

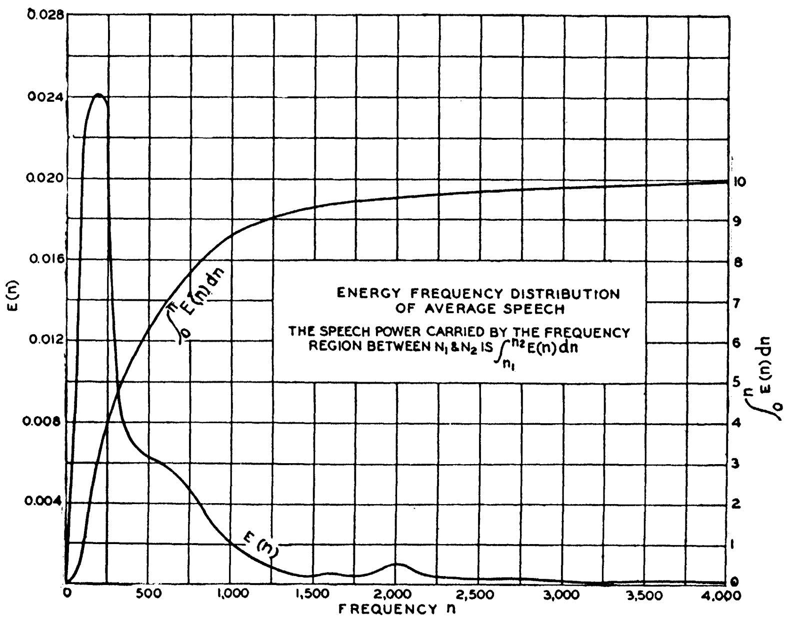


FIGURE 47.

still energy in some of the speech sounds up to as high as 10,000 cycles. It is important to notice that this distribution curve was obtained by the particular method just described of obtaining an average from a group of speakers producing a group of speech sounds. The occurrence of considerable energy in any particular frequency region does not necessarily imply that high sound intensities have been produced in this region. Such high energy values may be obtained either by the occurrence of a large number of speech sounds having components of comparatively low intensities in this region or by the occurrence of a few sounds having components of comparatively high intensities in this region. For some engineering work it is important to distinguish between these two possibilities.

CHAPTER IV

FREQUENCY OF OCCURRENCE OF THE DIFFERENT SPEECH SOUNDS

Words

AFTER learning the physical characteristics of the speech sounds it is natural to inquire how frequently they are used in conversational speech. It is evident that such knowledge will be useful in telephone engineering as well as in other fields. Godfrey Dewey [1] has made an extensive study of the frequency of occurrence of words, syllables, and fundamental vowel and consonant sounds in written material. This material was taken from representative sources such as modern newspapers, fiction, American speeches, personal correspondence, business correspondence, modern advertising, religious English, scientific English, and American magazines. As a result of this study, Dewey found the 100 most frequently occurring words to be those shown in Table XI. The numeral gives in per cent the frequency of occurrence of the word.

The definite article, "the" accounts for more than 7 per cent of all the words occurring on an average written page. Some words, such as winter, to-morrow, succeed, and railroads, which seem very familiar, occur only once in 10,000 words. The first 10 words given in this list account for more than 25 per cent and the 100 words account for more than 50 per cent of all the words occurring.

Syllable Combinations

Dewey's studies indicated also that the 100 most frequently occurring syllables are as shown in Table XII. The figure at

[1] Dewey, Godfrey, "Relative Frequency of English Speech Sounds," 1923, Harvard University Press, Cambridge, Mass.

TABLE XI

RELATIVE FREQUENCY OF OCCURRENCE OF WORDS

7.31	the	.58	not	.31	their	.20	time	.15	these
3.99	of	.58	at	.30	there	.20	up	.14	two
3.28	and	.57	this	.30	were	.20	do	.14	very
2.92	to	.54	are	.30	so	.20	out	.13	before
2.12	a	.52	we	.29	my	.19	can	.13	great
2.11	in	.51	his	.26	if	.19	than	.13	could
1.34	that	.50	but	.25	me	.18	only	.13	such
1.21	it	.47	they	.25	what	.18	she	.13	first
1.21	is	.46	all	.25	would	.17	made	.12	upon
1.15	I	.45	or	.24	who	.16	other	.12	every
1.03	for	.45	which	.23	when	.16	into	.12	how
.84	be	.44	will	.23	him	.16	men	.12	come
.83	was	.43	from	.22	them	.16	must	.12	us
.78	as	.41	had	.22	her	.16	people	.12	shall
.77	you	.39	has	.21	war	.16	said	.11	should
.72	with	.36	one	.21	your	.16	may	.11	then
.68	he	.33	our	.21	any	.15	man	.11	like
.64	on	.33	an	.21	more	.15	about	.11	well
.61	have	.32	been	.21	now	.15	over	.11	little
.60	by	.32	no	.20	its	.15	some	.11	say

the left of the phonetic syllable gives in per cent the frequency of occurrence.

Fundamental Sounds

The analysis of the phonetic pronunciation of the words enabled Dewey to find the frequency of occurrence of each of the fundamental speech sounds given in Table I. His values are given in Table XIII.

It is seen from this table that the sound i (as in tip) is the most frequently occurring phonetic sound. The sounds n, t, r, and o (as in ton), are the next four sounds in the order of their frequency of occurrence and they account for more than 36 per cent of all the sounds found on a written page. It is seen from the tables given above that a comparatively small part of the more common words comprise the large part of our ordinary speech. That this is true is emphasized by the following summary taken from Dewey's book.

General Summary

9 words are found to form over	25 per cent	of the total words
12 syllables are found to form over	25 per cent	of the syllables
4 sounds are found to form over	25 per cent	of the sounds
69 words are found to form over	50 per cent	of the words
70 syllables are found to form over	50 per cent	of the syllables
9 sounds are found to form over	50 per cent	of the sounds
732 words are found to form over	75 per cent	of the words
339 syllables are found to form over	75 per cent	of the syllables
19 sounds are found to form over	75 per cent	of the sounds
1027 words occurring over ten times form	78.6 per cent	of the words
1370 syllables occurring over ten times form	93.4 per cent	of the syllables
41 + 1 sounds form	100 per cent	of the sounds

TABLE XII

Relative Frequency of Occurrence of Syllables

7.3	the	.85	waz	.57	ol	.39	on	.28	os
4.0	av	.84	with	.56	kan	.38	men	.28	wud
3.3	in	.84	di	.54	wē	.38	érz	.28	som
3.3	ánd	.83	ti	.53	ez	.38	our	.28	what
3.2	i	.82	an	.52	bot	.34	en	.28	if
3.2	o	.78	áz	.52	hiz	.34	mī	.28	ōn
3.2	tū	.71	ō	.48	thā	.34	thár	.27	kom
2.4	ing	.70	hē	.48	nō	.33	op	.27	yu
2.1	or	.69	á	.47	wil	.33	out	.27	dā
1.6	ri	.68	ól	.47	on	.33	bin	.26	nes
1.4	it	.68	en	.46	ōā	.33	wor	.26	el
1.3	thát	.66	háv	.46	án	.32	thār	.26	si
1.3	iz	.64	bī	.45	which	.31	ev	.26	them
1.3	ī	.64	ar	.45	sō	.31	mē	.26	dis
1.2	li	.62	bi	.43	fram	.31	tu	.26	oth
1.1	fór	.61	át	.41	hád	.30	ex	.25	hū
.97	bē	.60	ór	.41	won	.29	its	.25	vér
.92	shon	.60	nat	.40	ment	.29	kán	.25	when
.91	ed	.58	tor	.39	ház	.29	dér	.24	dū
.86	yū	.57	this	.39	bul	.29	him	.24	pē

TABLE XIII

Relative Frequency of Occurrence of Speech Sounds

Speech Sound	Key	Relative Frequency	Speech Sound	Key	Relative Frequency
ū	tool	1.60	m	....	2.78
ō	tone	1.63	n	....	7.24
ó	talk	1.26	ng	hang	0.96
a	top	3.33	v	....	2.28
ā	tape	2.35	z	....	2.97
ē	eat	3.89	th	then	3.43
u	took	0.69	zh	azure	0.05
o	ton	5.02	f	....	1.84
á	tap	4.17	s	....	4.55
e	ten	3.44	th	thin	0.37
i	tip	7.94	sh	shell	0.82
ī	dike	1.59	b	....	1.81
ou	our	0.59	d	....	4.31
oi	oil	0.09	j	....	0.44
ew	few	0.31	g	....	0.74
w	....	2.08	p	....	2.04
y	....	0.60	t	....	7.13
h	....	1.81	ch	chalk	0.52
l	....	3.74	k	....	2.71
r	....	6.88			

PART TWO

Music and Noise

CHAPTER I

Physical Properties of Musical Sounds

Characteristics of Typical Musical Sound Waves

Musical sounds, like speech sounds, are also transmitted through the air by very complicated wave forms. They are characterized by being sustained at definite pitches for comparatively long times. When a change in pitch is made it takes place in definite steps called musical intervals, such as thirds, fifths, and octaves. For producing certain musical effects, occasionally the pitch is changed continuously from one position on the musical staff to another, but this is exceptional and not the rule.

There are two outstanding physical mechanisms for producing musical tones, namely, vibrating strings and vibrating air columns. The piano and the violin are examples of the first type of mechanism and the pipe organ, the flute and horn are examples of the second. Although the human voice is of sufficient importance to be considered by itself, it is really a mechanism of the second type.

It is well known that a single note sounded by one of these musical instruments contains more than one frequency. The lowest component frequency, called the fundamental, usually determines the pitch but there is, in addition, a large number of component frequencies called harmonics, each one being a simple multiple of the fundamental frequency. It is this abundance of harmonics that produces the richness of musical tones.

The Electrical Harmonic Analyzer

Unfortunately no good wave pictures of sounds from musical instruments seem to be available. Tones from some of the common musical instruments have been analyzed into their component frequencies by means of an electrical harmonic analyzer.[1] In using this instrument to analyze sound waves, a condenser transmitter is used to transfer the acoustic wave into an electrical wave which is a faithful copy of the original. This electrical wave is then sent into a selective network, the essential feature of which is a sharply tuned circuit whose frequency of tuning is controlled by varying its

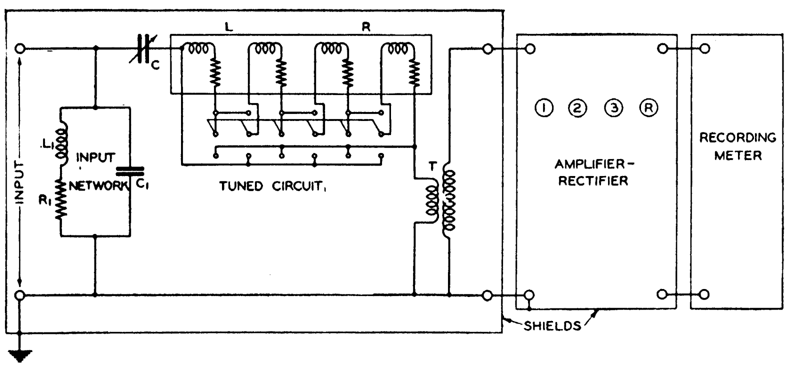


FIG. 48.—SCHEMATIC ANALYZER CIRCUIT.

capacity in small steps by means of a pneumatic apparatus similar to that used in a player piano. Maximum responses of the circuit occur at frequencies of tuning which coincide with the frequencies of the components of the complex wave. A schematic of the analyzer circuit is shown in Fig. 48.

An automatic photographic recorder registers as a permanent record the amount of current getting through the tuned circuit at each frequency. From this record the relative amplitudes of the components of the complex wave are readily determined. For convenience of operation an automatic con-

[1] Wegel, R. L., and Moore, C. R., "An Electrical Frequency Analyzer," published in *The Bell System Technical Journal*, April, 1924.

trol apparatus is provided so that it is only necessary to connect the complex source or sources to be analyzed and press the starting button. Then the completed record of the analyses is delivered after the machine has passed through the entire

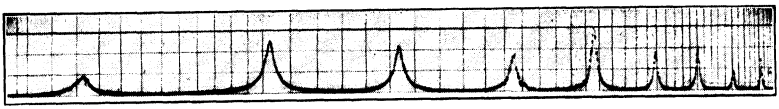


FIG. 49.—HARMONIC ANALYSIS OF TONE FROM TROMBONE ORGAN PIPE.

range of frequencies. In Fig. 49 is shown the record obtained by means of this machine of the tone from an organ pipe.

In Fig. 50 are shown the essential mechanical features of the analyzer. The pneumatic arrangement is a modification of a player-piano mechanism in which a paper roll of standard

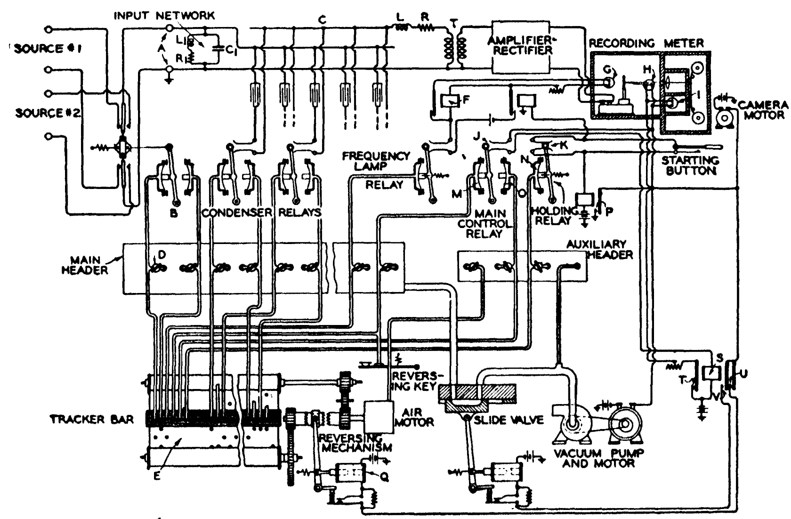


FIG. 50.—ARRANGEMENT OF PNEUMATIC AND ELECTRICAL APPARATUS.

dimensions is used. By proper perforation of the roll special pneumatic relays are operated in proper sequence to switch the condensers of the tuned circuit, flash frequency lines on the record, stop the mechanism after a record has been completed, rewind the piano roll, and perform other functions necessary to leave the analyzer in the starting position.

The photographic recording apparatus consists of the camera motor for moving the sensitized record paper at a constant rate, a proper arrangement of lenses and lamps for illuminating the mirror galvanometer and tracing the scale and frequency lines, and suitable baths for developing and fixing the record. The record is drawn through the mechanism by means of the two motor-driven rubber rollers which serve also to remove excess solution. A more detailed description of this apparatus will be found in the paper by Wegel and Moore, the inventors of this analyzer.

Acoustic Spectra of Typical Musical Instruments

The next four figures, 51, 52, 53, and 54, show the results obtained by means of this apparatus for the analysis of some musical tones. The acoustic spectrum shown in each case was obtained directly from an experimental chart similar to that shown in Fig. 49. The wave picture was then constructed from the component frequencies, assuming that all of the components had the same phase. Due to the time required to make a complete analysis of sung vowels and piano tones, the sound is necessarily interrupted but this does not affect the final result. It is seen that the low-pitched piano tone has a large number of harmonics. For all of the strings in the lower pitch register on the piano, the tone is largely carried by the harmonics rather than the fundamentals. It is interesting to note that the component frequencies between 2500 and 3000 in the case of the clarinet are very much magnified. It is seen that the tenth harmonic has about one-half the amplitude of the fundamental. Also, for the 'cello organ pipe the third harmonic has about five times the amplitude of the fundamental. The trombone organ pipe is very rich in harmonics. Experiments made with Bourdon organ pipes show that the amplitudes of the harmonic frequencies are quite small compared to the fundamental.

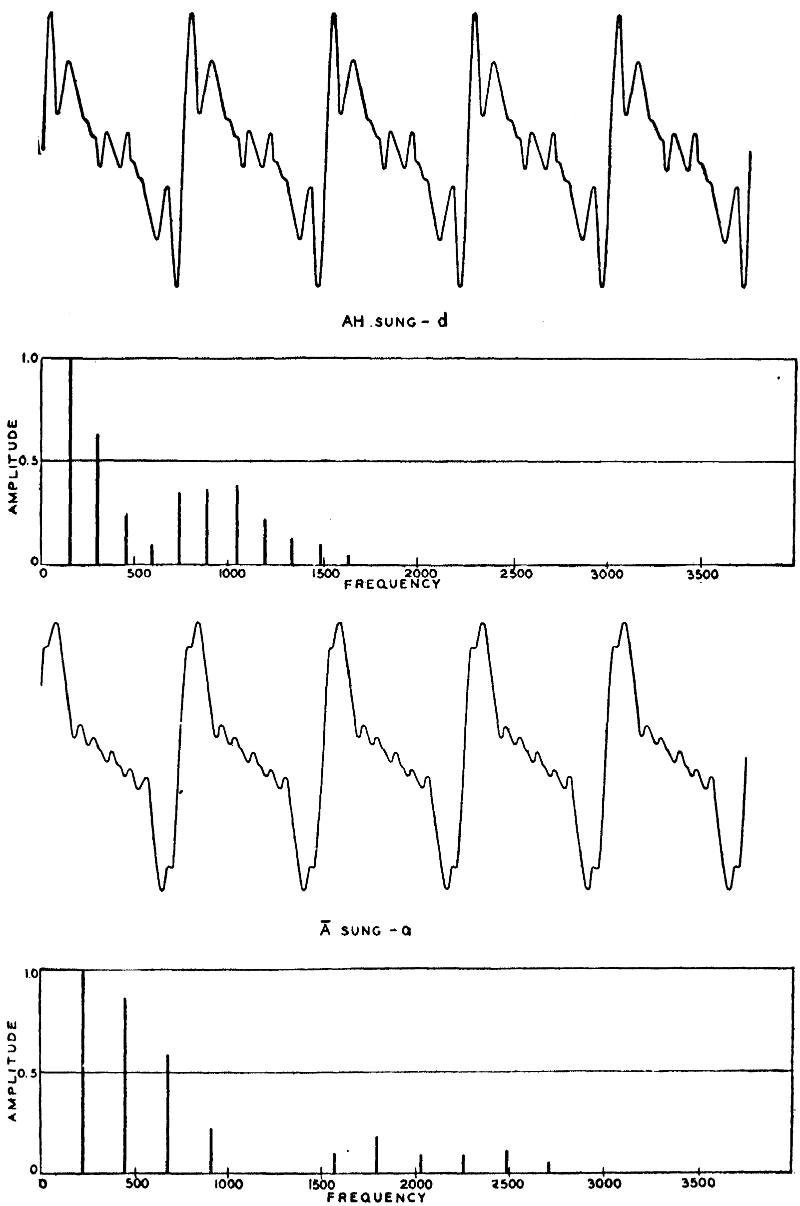

AH SUNG - d
1.0
0.5
0
AMPLITUDE
0
500
1000
1500
2000
2500
3000
3500
FREQUENCY
Ā SUNG - a
1.0
0.5
0
AMPLITUDE
0
500
1000
1500
2000
2500
3000
3500
FREQUENCY

FIGURE 51.

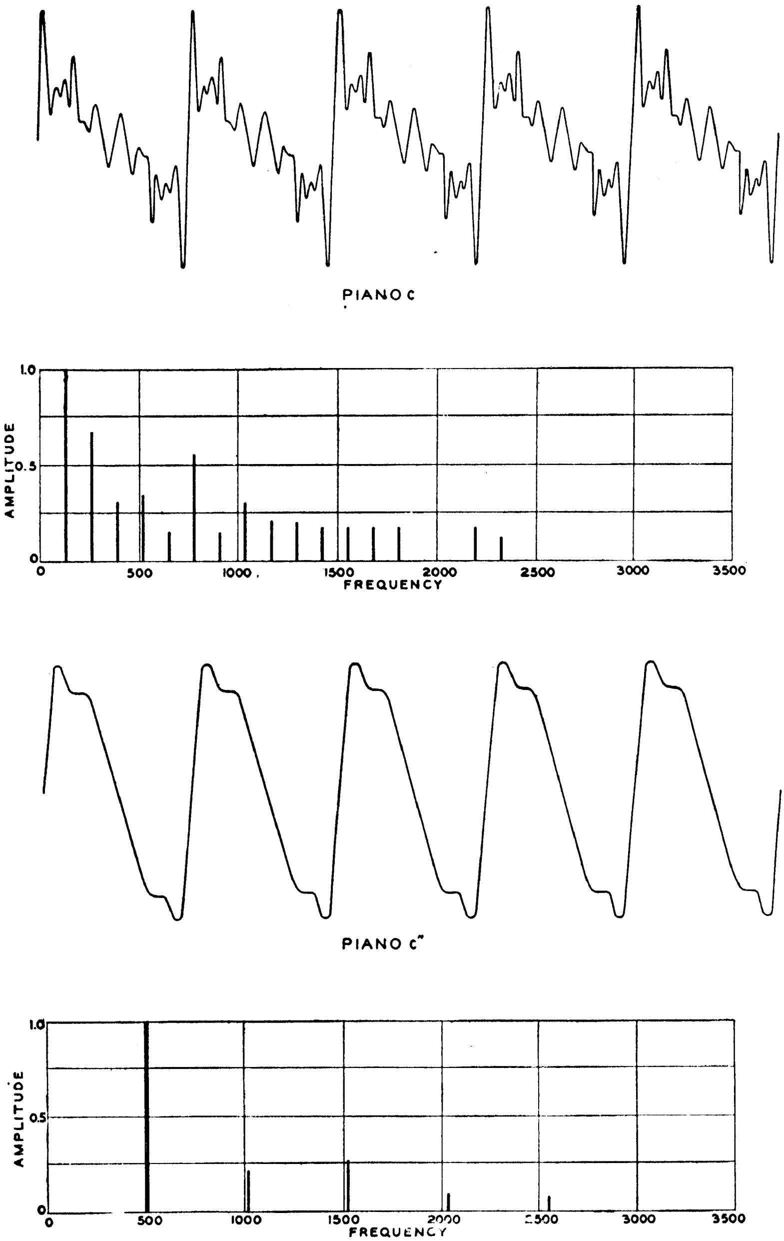

PIANO C
1.0
0.5
0
AMPLITUDE
0
500
1000
1500
2000
2500
3000
3500
FREQUENCY
PIANO C"
1.0
0.5
0
AMPLITUDE
0
500
1000
1500
2000
2500
3000
3500
FREQUENCY

FIGURE 52.

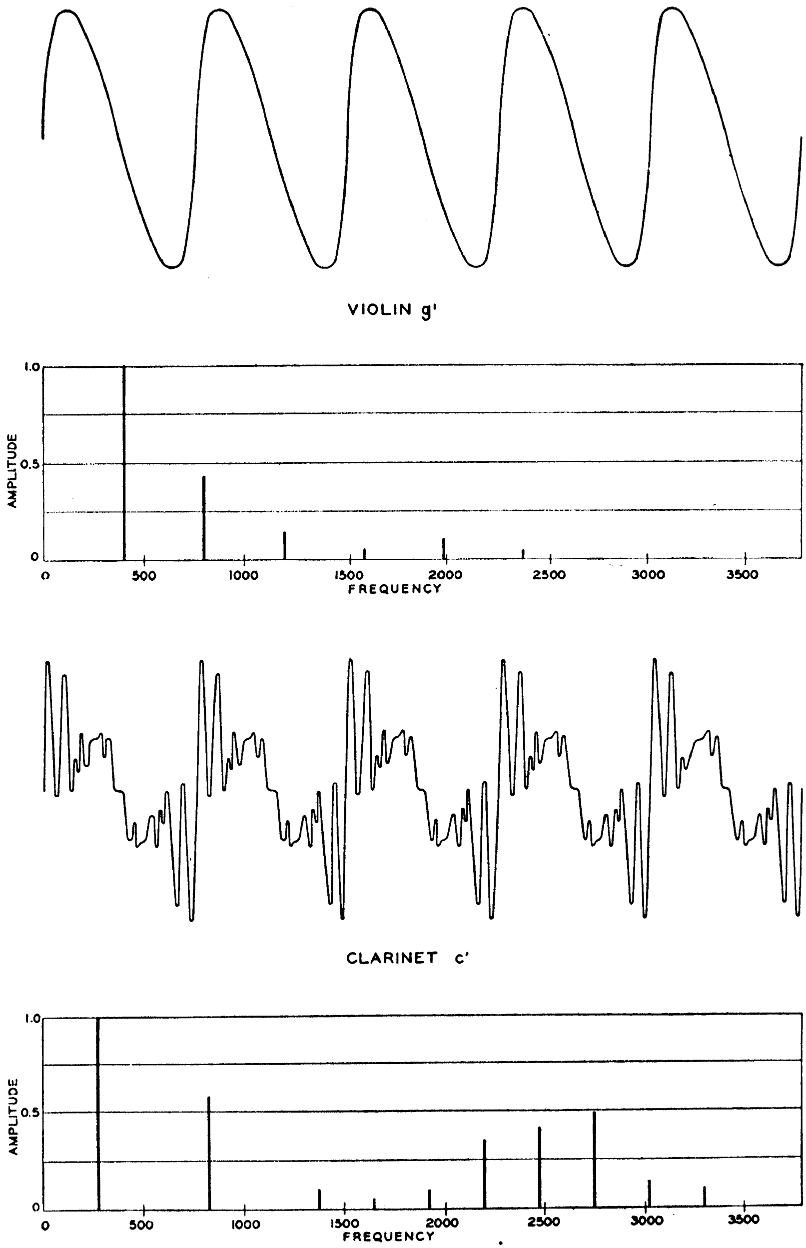

VIOLIN g'
1.0
0.5
0
AMPLITUDE
0
500
1000
1500
2000
2500
3000
3500
FREQUENCY
CLARINET c'
1.0
0.5
0
AMPLITUDE
0
500
1000
1500
2000
2500
3000
3500
FREQUENCY

FIGURE 53.

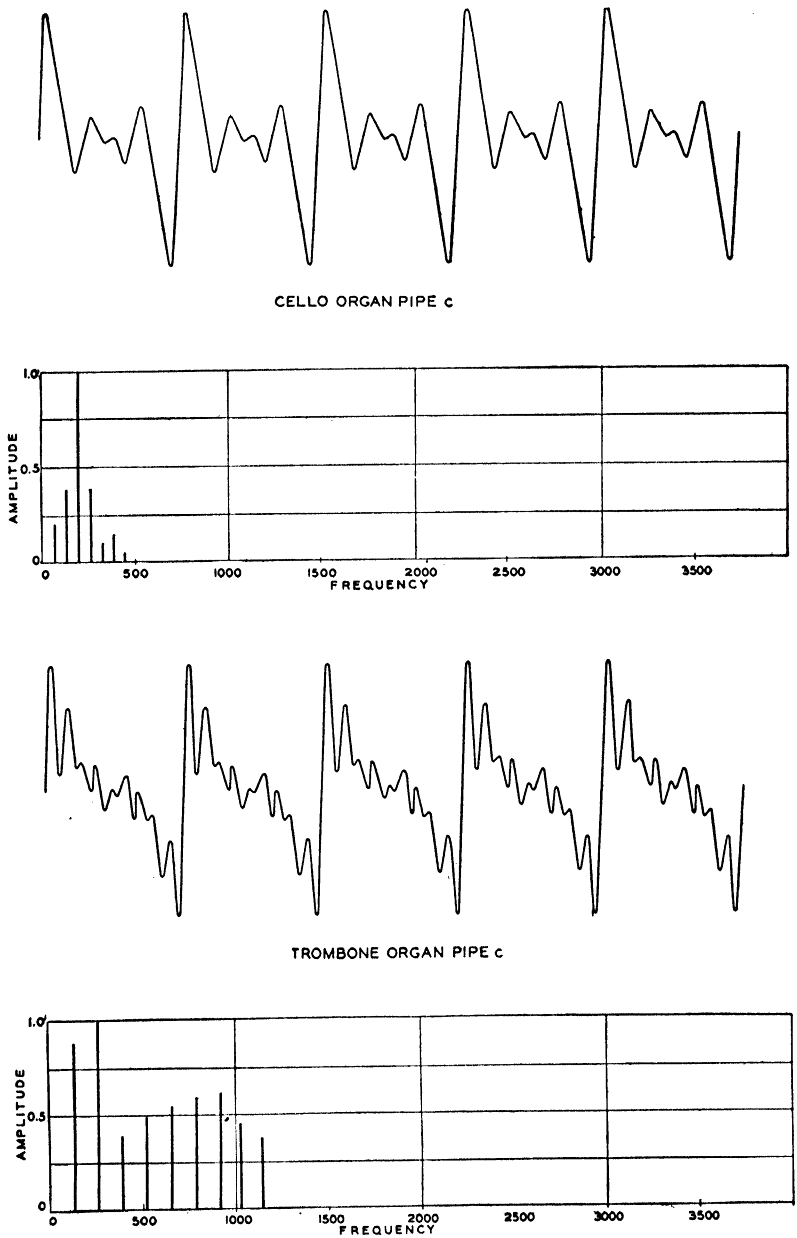

CELLO ORGAN PIPE c
1.0
0.5
0
AMPLITUDE
0
500
1000
1500
2000
2500
3000
3500
FREQUENCY
TROMBONE ORGAN PIPE c
1.0
0.5
0
AMPLITUDE
0
500
1000
1500
2000
2500
3000
3500
FREQUENCY

FIGURE 54.

Ranges of Frequency and Intensity for Music

A similar analysis was made for typical pipes taken from the pipe organ. The average results obtained from this analysis are shown in Fig. 55. The ordinates give the average pressure variation in bars which is produced in an ordinary room when the pipes having a fundamental frequency equal to the abscissas are blown with a pressure equivalent to that

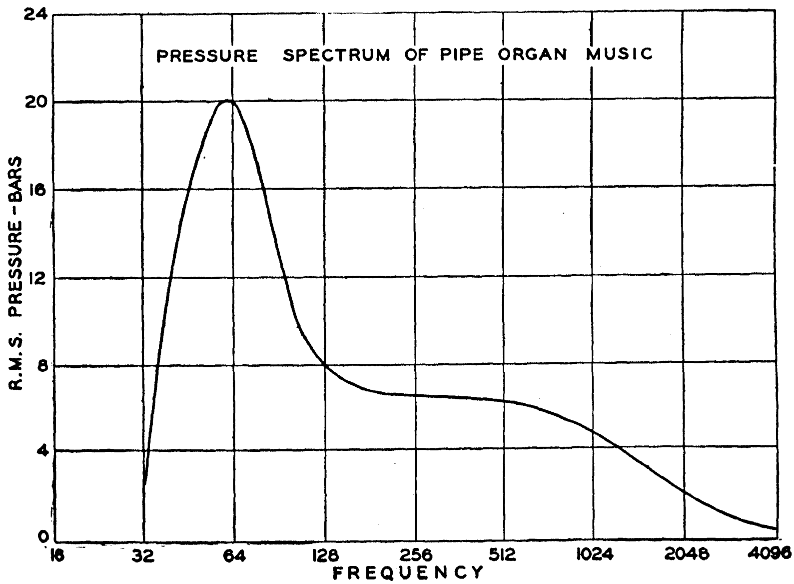


FIGURE 55.

ordinarily used. These absolute values will depend very largely upon the type of the room and also upon its size. It is seen that the maximum pressure variations in the air are produced by those organ pipes having a pitch of 64 cycles. On account of the large content of energy carried by the lower frequencies in organ music, it is difficult to build a transmission system which will faithfully reproduce this kind of music.

Experiments with the pipe organ in the Warner Brothers' Theatre of New York City indicated that the sensation levels produced in the main body of the hall were from 40 to 50 db, which are somewhat lower than those shown in Fig. 55.

To obtain a rough estimate of the range of frequencies and intensities produced by musical instruments the following experiments were performed by C. E. Lane of Bell Telephone Laboratories. To determine the sound pressures created, an apparatus was used which consisted of a calibrated condenser transmitter, a vacuum tube amplifier and rectifier, and an ammeter. From the reading of the meter an average r.m.s. pressure on the transmitter diaphragm could be determined. Measurements were made with the lips of the person singing or with the instrument which he was playing about 18 inches away from the transmitter. In each case three intensity levels corresponding to the musical notations, *pp*, *mf*, and *ff*, were determined.

For singing, a part of Wagner's Pilgrims' Chorus was used and the intensities of certain notes were measued. In this way three bass voices, two tenor voices, three soprano voices, and three alto voices were measured. These persons did not have unusually strong voices but represented about the average found in glee clubs and choirs. In Table XIV the average results are given. The figures give the pressure variation in bars created at a distance of 18 inches from the singer. The powers created by the voice corresponding to these figures range from 1000 to 30,000 microwatts, which are considerably higher than those used in conversational speech.

The singing intensity was found to remain approximately constant from the middle range of pitches to the higher range. For the low range the intensity fell off rapidly so that the lowest note that could be sung well produced a pressure variation of only about one-thirtieth that of the higher ranges. In a similar way results were obtained for the various musical instruments.

Table XV gives the values for the wind instruments for the case when these instruments are pointed at an angle of 60° from the pick-up transmitter. It is seen that the stringed instruments produce, in general, less intensity than the wind instruments, the violin producing the weakest sounds. The bass drum produces the greatest intensity. The sound from this instrument is transmitted through the air by means of very low frequencies. As a rule the bass instruments produce the greatest intensities, the tenor and alto the next greatest, and the soprano the least intensities that are used in music.

TABLE XIV

	Pressure (Bars)			Pitch Range (Octaves above 1 Kilocycle)
	pp	*mf*	*ff*	
Bass	13	18	31	− 3.7 to − 2.6
Tenor	14	21	34	− 2.8 to − 0.9
Soprano	14	20	24	− 2.1 to + 0.2
Alto	6	13	18	− 2.5 to − 0.2

TABLE XV

	Pressure (Bars)			Pitch Range (Octaves above 1 Kilocycle)
	pp	*mf*	*ff*	
Saxophone (C Melody)	20	26	36	
Trombone	12	21	35	− 3.5 to − 1.0
Cornet	17	25	34	− 2.6 to − 0.1
Clarinet	15	27	33	− 2.6 to + .6
Fife	12	24	35	− 1.5 to + 1.8
Baritone Horn	17	26	36	− 3.6 to − 1.4
Bass Tuba	21	34	41	− 4.4 to − 1.4
Organ Pipe	......	30		
Violin	8	18	25	− 2.5 to + 2.0
Banjo	24	32	36	
Mandolin	13	18	24	− 2.5 to + 0.4
Bass Viol	13	21	31	− 4.6 to − 3.5
Harp (single notes)	11	20	27	− 5.0 to + 1.6
Harp (chord in G, including thirteen notes)	25	29	32	
Piano (single notes)	..	23	31	− 5.1 to + 1.6
Bass Drum	29	40	48	
Snare Drum	10	22	33	

The small amount of data available indicates that musical instruments are built so that the tones of different pitch will have approximately the same loudness as interpreted by the ear. Since the ear is relatively insensitive to the low pitches, organ pipe tones for this low-pitch range are considerably more intense than those produced in the high-pitch range. The same is true of the tones from a piano. The mean power of the sound during the rendition of an orchestral selection varies over wide ranges, sometimes as much as 100,000 to 1. This fact makes it very difficult to handle the proper transmission of such music.

The important frequency range for music is from 50 to 5000 cycles, and for very faithful reproduction frequencies an octave below and an octave higher than these limits must be transmitted.

CHAPTER II

NOISE

Physical Properties of Noise

THOSE sounds to which no definite pitch can be assigned are usually classified as "noise." The clapping of hands, the rattling of paper, the hammering of typewriters, and the roar from the traffic in the street, are typical types of noise. Practically all types of sound which cannot be classified as speech or musical tones come under this classification. Although the noise waves are carried through the air by vibrations similar to those transmitting both speech and music, their form is very much more complex. The range both in intensity and frequency is very much greater than for the first two classes of sounds discussed. For this reason it is very much more difficult to transmit them faithfullv by means of any transmission system.

In Fig. 56 is shown a typical wave form of street noise. As will be seen, its principal characteristic is the great irregularity in the vibration. The wave form at the bottom was produced by a pure tone having a frequency of 500 cycles per second and is given for comparison.

When transmitting speech or music either directly to an audience in a large hall or over an electrical system, such as a radio or a telephone system, there is always an interference to the proper reception of such speech and music, due to other sounds being present. These extraneous sounds which serve only to interfere with the proper reception aré designated by engineers as "noise." With such a designation, the sound may be either periodic or non-periodic as long as it is something that would be better eliminated. In telephony noises result from a number of different sources. Some of these noises arise from inductive effects between telephone lines and other types of electrical transmission lines; other noises are caused by electrical disturbances originating within the telephone system itself. In addition to these, there is always, of course,

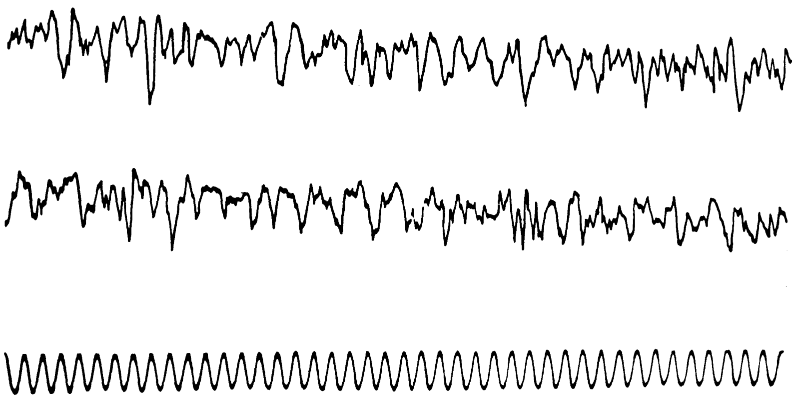


FIG. 56.—TYPICAL WAVE FORM OF A STREET NOISE AND A PURE TONE OF 500 CYCLES.

a certain amount of noise, generally classed as "room noise," in places where telephones are used.

A sample spectrum of line noise current is shown in Fig. 57. This was obtained by analyzing the current flowing in an open wire toll line which was terminated by a resistance of 700 ohms. It is seen that the components, with one exception, are all harmonics of a 60-cycle fundamental. Also it will be noticed that the odd harmonics are much stronger than the even ones, which is a notable characteristic of a power generator. It is evident then that the principal part of this line noise current is due to the inductive effect of a power line carrying a 60-cycle current upon this particular telephone line. When a subscriber's circuit is connected to such a toll line this line noise current produces a sound in the telephone receiver called "line noise." It sounds like a hum having a definite pitch and for that reason it might be classed as a musical tone. However, due to the interference which it causes to the proper recognition of the transmitted speech sounds, it is classed as noise.

Other types of electrical disturbances which may be picked up or which may originate in the telephone system itself produce sounds at the receiver which are more nearly true noise sounds as they have no periodic characteristics.

The room or booth on the receiving end of the line always has some noise present. It varies in character from the ticking of a clock in a quiet country home to the intense noise in a booth on the platform of a subway railway station. The type of room noise which is the most characteristic is that which is usually characterized as "roar" from the street. For some purposes it is desirable to know the average frequency content of the room noise. Its character is so varied, however,

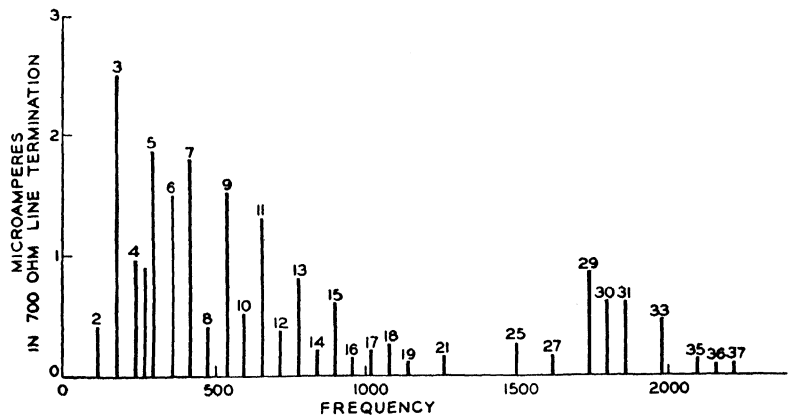


FIG. 57.—LINE NOISE SPECTRUM.

that for most purposes it is usually assumed that the energy of the room noise is scattered uniformly throughout the important range of speech frequencies.

The interference to the person carrying on a telephone conversation is due mainly to the room noise getting to the telephone ear, that is, to the ear on which the telephone receiver is placed. As will be shown later, only a small amount of interference to the proper reception of speech is caused by any noise in the non-telephone ear. Even holding the telephone serviceable which measure the deafening effect of the noise upon the ear. For this reason any instrument which has been designed for measuring the acuity of hearing can be used with slight modifications, for measuring room noise.

One such instrument which has proved very serviceable as a rough indication in this connection is known as a buzzer type audiometer. A picture of this instrument is shown in Fig. 58. In this instrument a buzzer element generates an electrical

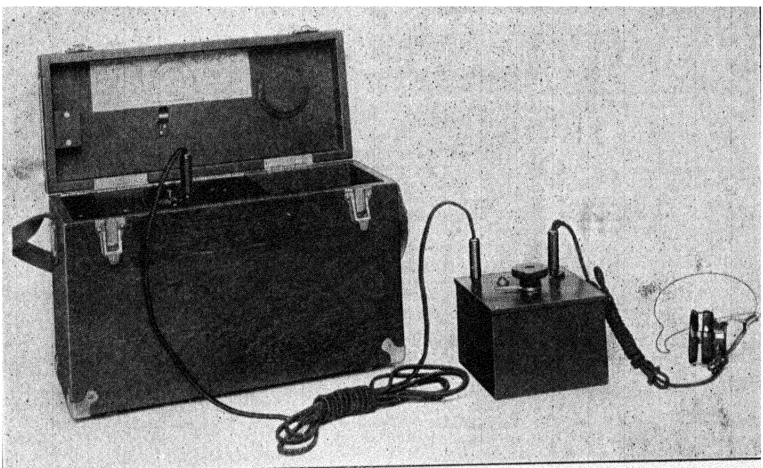


FIGURE 58.

current having component frequencies scattered throughout the entire speech range. This generator is connected through a system of networks, called an attenuator, to a telephone receiver which has a special cap designed to hold the receiver at a fixed distance from the ear. To make a measurement of the room noise, the intensity of the sound issuing from the receiver is reduced by turning the dial of the attenuator until the receiver sound is masked by the noise present in the room. In other words, the threshold of audibility of the tone from the audiometer determines the amount of noise in the room. The difference between the threshold setting obtained in a receiver as tightly as possible to the ear does not entirely eliminate the noise.

In radio transmission "room noise" affects the received sounds in the same way that it does in telephone transmission, but the line noise consists of static, squeals, and howls from spark and regenerative sets, which are improperly operated, as well as from sources in the radio system itself. In this case also, it is usually assumed that such noises are scattered uniformly throughout the audible frequency range, except for such frequency selectivity as is imposed by the characteristics of the transmission system.

When there is no obvious way of reducing the noise level, it is necessary to raise the level of the speech or music being transmitted. For example, the best remedy which has yet been proposed for static in the receiving set is the construction of powerful broadcasting stations. When a receiving set is within a short distance from such stations little trouble is experienced from static noises. Similarly, in the telephone plant it has been necessary to supply to the average subscriber more than one thousand times as much power as would be necessary for good hearing if the line and room noises were eliminated.

Method of Measuring Noise

A method of gauging line noise is to measure the electrical currents induced in the telephone transmission system when it is in operating condition but not transmitting speech. If a meter were designed to give the proper importance to the various frequency components in the noise currents, its reading would be an indication of the detrimental effect of the noise currents on the line.

For measuring room noise a high quality telephone system similar to that for recording speech sounds described in Part One, Chapter II, could be used. Such systems have been tried but they have been found to be impractical for most purposes because the apparatus is too bulky and requires expert attention to keep it in adjustment. Instruments have proved to be noisy place and that obtained in a quiet place by an individual with normal hearing gives the deafening effect of the room noise for the buzzer tone.

A similar principle can be used to determine the deafening effect of noise for each frequency. Such an instrument, which was designed primarily for measuring the degree of deafness, is shown in Fig. 59 and is known as the Western Electric 2-A Audiometer. It is designed to produce any one of eight pure tones ranging in frequency from 64 to 8192 cycles per second

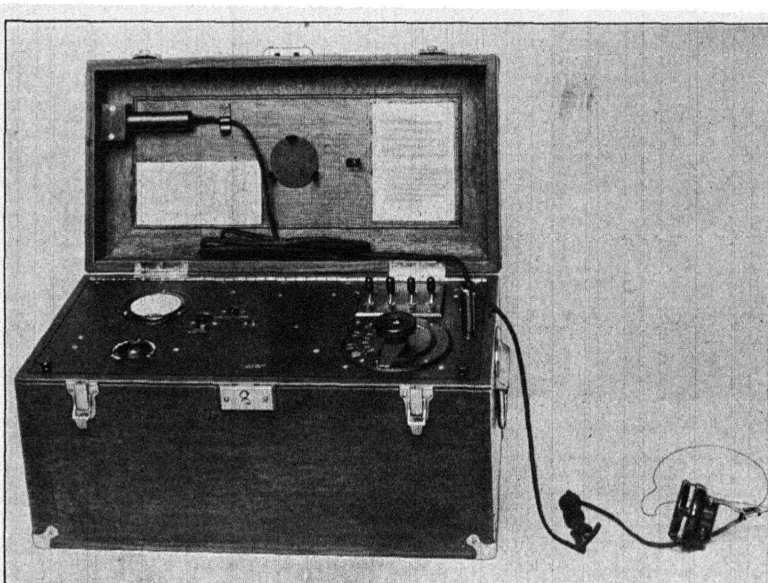


FIGURE 59.

at definite sensation levels which are obtained by properly setting the dial. A curve showing the loss of hearing at each frequency is called an audiogram. For this reason it is convenient to refer to a curve showing the deafening effect at each frequency due to noise as a noise audiogram. Such a curve if obtained for all frequencies would be more directly correlated to the effect upon the recognition of sounds than is an acoustic spectrum of the noise.

When single frequency components are present in the noise being measured and single frequency tones are used for measuring the noise audiograms, beats are produced which make the proper location of the threshold rather difficult. For this reason, a type of noise meter has been developed which produces bands of frequencies instead of single frequencies. By its use, more accurate noise audiograms can be obtained. These bands of frequencies are produced by means of another type of portable audiometer known as the phonograph audiometer. This instrument consists of a phonograph turntable, special records, and an electrical reproducer connected to a telephone receiver. An oscillating circuit in which is placed a variable condenser was designed so that bands of frequencies of any width and with the components spaced at any desired frequency interval could be produced. These bands of frequencies were then recorded on phonograph records by means of the new electrical process. From these they are reproduced by means of the electromagnetic reproducer and sent to the receiver which has the special receiver cap described above. The method of determining the masking effect of the noise is the same as that given for the other two audiometers. In practice it is difficult to measure the masking effects of varying noises.

Instead of using the deafening effect as a measure of noise, another method is to produce an artificial noise which is judged by the observer to have the same interfering effect as that existing. The measurement of noise by this means is called the "balance method." It may be used either for measuring line noise issuing from the telephone receiver at the end of the line or for measuring room noise. The difficulty with such a method is due to the inability of a person to judge accurately when two sounds which differ greatly in character are equally loud. The phonograph audiometer is very suitable for making such measurements if records are available which have a character of noise similar to the type which it is desired to measure.

Results of Noise Surveys

In all of these instruments the unit showing the degree of the deafening effect is the db or the sensation unit. For example, Dr. E. E. Free [1] found that the noisiest place in New York City was at Thirty-fourth Street and Sixth Avenue, where the buzzer audiometer registered 50 db. This means that sounds originally near the threshold of hearing which one desires to hear at this place must be magnified 100,000 times

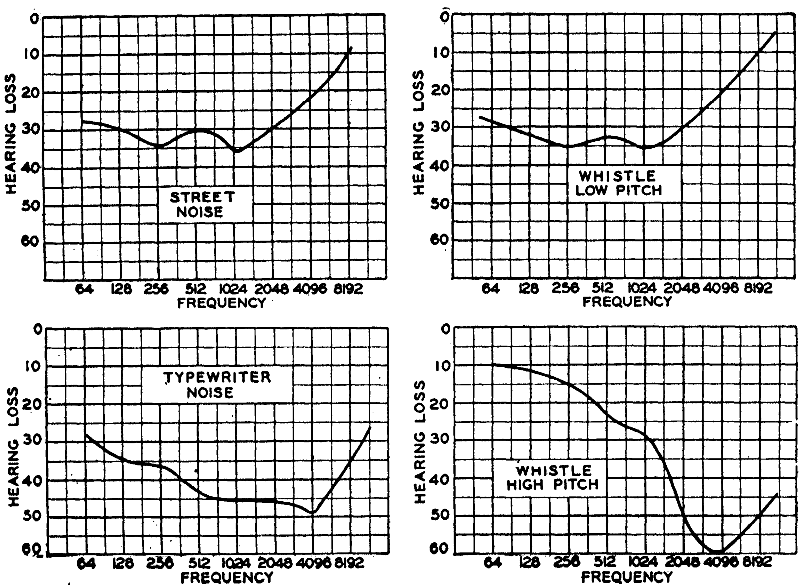


FIGURE 60.

before they can be heard. In many places the noise becomes much greater than this.

Measurements in a number of offices and places of business in New York City have indicated that the deafening effect usually encountered is about 30 db. For this reason, a man who is permanently deafened by this amount will scarcely notice his defect except when he goes into a quiet place such

[1] Free, E. E., "Noises You Never Hear," *Pop. Sci. Monthly*, v, 109, pp. 16–17, 110, August, 1926.

as a church or a theater. Such a deafened person would also have little difficulty in using the telephone for the system is designed to reproduce speech sufficiently loud to override the deafening effect of the noise.

To illustrate the type of noise audiograms which one would obtain for certain kinds of noise, four such audiograms are shown in Fig. 60. As indicated, the first is for typical street noise, the second for noise coming from the typewriter which is being rapidly operated, the third for a low-pitched whistle, and the fourth for a high-pitched whistle.

Part Three

Hearing

CHAPTER I

Mechanism of Hearing

How we hear has been a subject for discussion by men in the various branches of science for a long time. Although there is good agreement concerning the principal structures of the ear, there is still considerable controversy regarding the function of the various parts. Of the five senses it is hearing that makes us aware of the presence of physical disturbances called sound waves. For audition purposes sound may be classified into two groups, namely, pure tones and complex sounds. A pure tone is specified by two properties, namely, the pitch and the loudness. These sensory properties are directly related to the physical properties, frequency and intensity of vibration of the air particles near the ear. Some psychologists state that there is a third sensory property of a pure tone, namely, volume or extension of the tone. It is related to both intensity and pitch although the qualitative relationship has not been definitely established. Complex sounds may be considered as combinations and variations of pure tones. Vibrations in the sound wave communicate mechanical vibrations to the ear drum which, in turn, communicates the vibration to the inner ear where the nerve endings are excited.

Description of the Organs of Hearing

The ear mechanism may be divided into three general parts: the outer ear, the middle ear, and the inner ear. The outer ear consists of the external part or pinna, and the ear canal or auditory meatus. The middle ear contains three small bones or ossicles called, respectively, the hammer, the anvil, and the stirrup. The inner ear contains the cochlea, vestibule, the semi-circular canals, and the endolymphatic duct and sac. In the cochlea are located nerves which give us the sense of

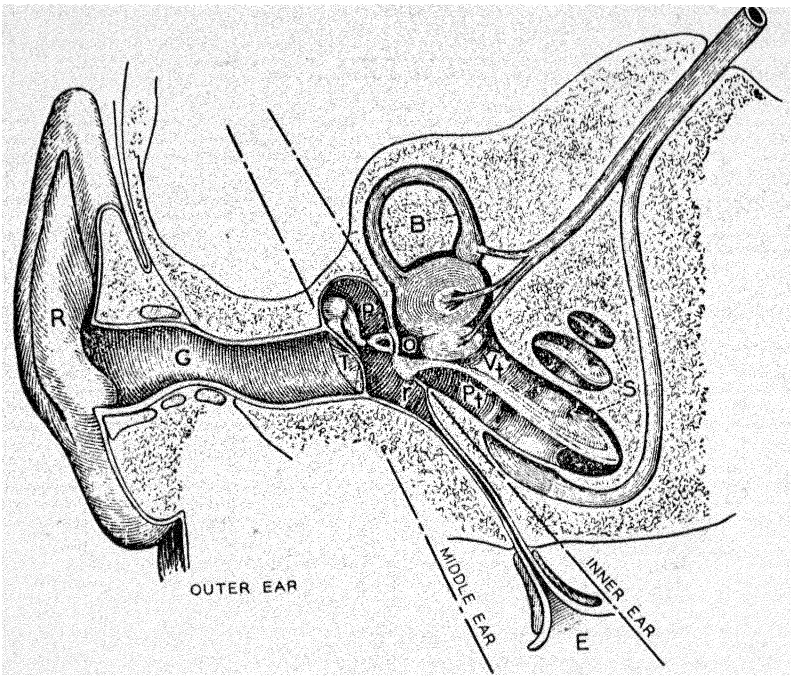


FIG. 61.—SEMI-DIAGRAMMATIC SECTION THROUGH THE RIGHT EAR (CZERMAK): *G*, EXTERNAL AUDITORY MEATUS; *T*, MEMBRANA TYMPANI; *P*, TYMPANIC CAVITY; *o*, FENESTRA OVALIS; *r*, FENESTRA ROTUNDA; *B*, SEMI-CIRCULAR CANAL; *S*, COCHLEA; *Vt*, SCALA VESTIBULI; *Pt*, SCALA TYMPANI; *E*, EUSTACHIAN TUBE; *R*, PINNA.

hearing, and in the semi-circular canals are located nerves which cause reactions concerned with the maintenance of equilibrium.

Figure 61 shows a schematic diagram of the parts of the ear with the inner ear much enlarged. The pinna is used by a number of animals to aid in collecting the sound. The human pinna has almost lost this function but a cupped hand held to the ear sometimes supplants it.

The ear canal, or auditory meatus, *G*, is about three centimeters long. It is closed at the inner end by the ear drum or tympanic membrane. Attached to the drum from its center and upwards by a long part called the handle is the first of the ossicles, called the hammer. The top of the hammer is connected with the anvil by a joint and the anvil in turn is connected to the stirrup, the small bone that conveys the motion through the oval window to the labyrinth in the inner ear. The part of the stirrup lying in the oval window is flat and is called the foot plate. It is held in place by an annular ligament of the membrane which prevents the fluid of the inner ear from coming into the middle ear. The mastoid cells are connected to the middle ear but are not concerned with hearing.

The inner ear has a dense bony wall forming an irregular cavity referred to as the bony labyrinth and is filled with fluid. It contains a smaller structure of the same general shape called the membranous labyrinth which contains a fluid that is separate and distinct from the rest of the fluid in the bony structure. Its walls are formed by a very soft membrane so that sound waves pass through them with little obstruction. The cavity of the inner ear is encased in solid bone and has only two small openings into the middle ear, one, at the oval window into which fits the stirrup and one at the round window indicated at *r*. An elastic membrane is stretched across the round window and is sometimes referred to as the secondary ear drum. The middle ear is connected to the outside air by means of a small tube called the Eustachian tube, which opens into the upper part of the throat behind the nasal cavity. Infectious germs sometimes travel up this tube from the nasal cavity and cause a "gathering" in the middle ear.

The inner ear consists of three principal parts, namely: (1) the semi-circular canals which take no part in the mechanism of hearing, but serve as an organ of balance, (2) the vestibule, the space just behind the oval window, and (3) the cochlea which is really the end organ of hearing. Cross-sections of the cochlea as it twists into a relatively long spiral of two and three-quarter turns like a snail shell are indicated at S in Fig. 61. The center of the spiral is a bone called the modiolus, and is perforated to allow space for the auditory nerve. The nerve enters the base of the cochlea and outside it unites with the nerves from the semi-circular canals into two parts forming the eighth cranial nerve. The cochlea

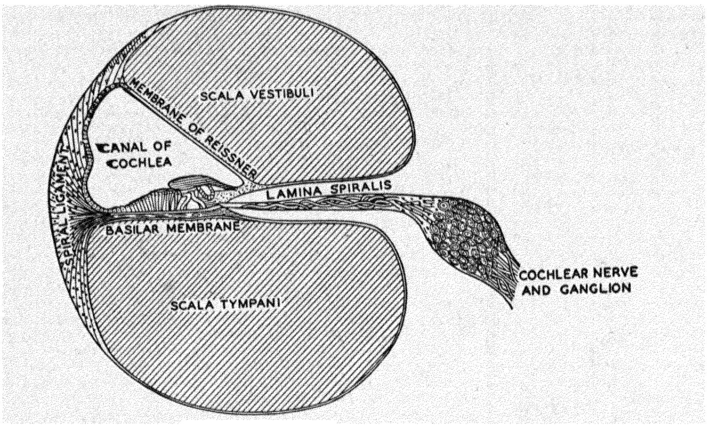


FIG. 62.—COCHLEA IN TRANSVERSE SECTION. OBSERVE ESPECIALLY THE CANAL OF THE COCHLEA WHICH IS A PART OF THE MEMBRANOUS LABYRINTH. (Testut.)

is divided along its length into three parts by the basilar membrane and Reissner's membrane. These form three parallel canals which are wound into the spiral. A cross-section showing the shape of these canals is given in Fig. 62. The oval window is at one end of the scala vestibuli and the round window at the end of the scala tympani.

As indicated in this figure, the canals are called the scala media or canal of cochlea, scala tympani, and scala vestibuli. As stated before, the membrane of Reissner is a very thin flexible membrane which will very readily pass any sound waves, so that from a dynamical consideration, the canal of cochlea and scala vestibuli may be considered as a single chamber filled with fluid. The partition between the scala tympani and the other two chambers is composed of a bony projection called the lamina spiralis for about half the distance, the remainder being a flexible membrane called the basilar membrane. It is seen from this figure that if any vibratory energy is communicated from one side of this partition to the other, it must vibrate the basilar membrane. On one side of the basilar membrane is the organ of Corti, which contains the nerve terminals in the form of small hairs extending into the canal of cochlea. Attached to the lamina spiralis and lying over the hair cells is another soft loose membrane called the tectorial membrane. The details of this part of the inner ear are made clearer by Fig. 63, which is a greatly magnified cross-section of these two membranes. It is seen from this figure that there are five rows of hair cells at the terminals of the so-called rods. There are about 5000 rods in each of the four outer rows and about 3500 in the inner row, making a total of about 23,500 rods. At the end of each rod there is a hair cell from which project twelve to fifteen hair cilia

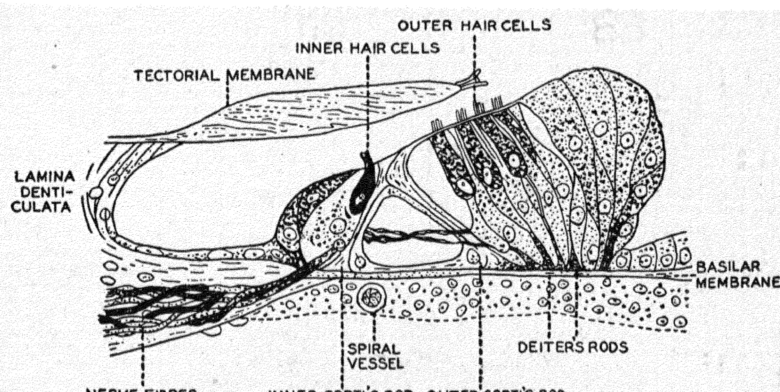


Fig. 63.—Corti's Organ (after Retzius). The tectorial membrane is shown contracted in the process of hardening the tissue, and torn away from the plateau of Corti.

into the liquid of the cochlea. When a sound excites the sense of hearing there is a relative motion between the basilar membrane and the tectorial membrane which causes the hair cells to stimulate the nerve endings at their base. The base of the inner rod of Corti is supported on the edge of the bony projection, called the lamina spiralis. For this reason, according to some authors[1] the motion which stimulates the hair cells is a lateral one between the rods and the tectorial membrane due to the rocking motion of the former. According to Dr. Shambaugh, the stimulation is principally due to the vibration of the tectorial membrane. Helmholtz, as well as many other writers on the subject, assumed that the basilar membrane was the principal vibrator carrying the rods of Corti with it. Thus the hair cells are excited by their relative motion to the tectorial membrane. Any of these points of view are still possible even if we assume that the ends of the hair cells are imbedded in the tectorial membrane. In any case the nerve endings are stimulated when a sound vibration is conducted from the canal of cochlea to the scala tympani.

A nervous impulse is then conducted by means of nerve fibres through the base of the rods to the cochlear nerve and then to the brain, causing the sensation of hearing. The chambers separated by the basilar membrane are connected by a small opening at the apex of the cochlea called the helicotrema.

The drum of the ear and the ossicles of the middle ear act as a sort of transformer to communicate the vibratory energy from the air, a light medium, into the liquid, a dense medium. Due to the fact that the area of the stirrup which plunges into the fluid of the inner ear is about one-twentieth of that of the ear drum and also due to the lever action of the three bones, the pressure exerted by the oval window of the middle ear upon the fluid of the inner ear is from thirty to sixty times that exerted by the air upon the ear drum. This transformer action permits the sound wave to pass more readily from the air into the liquid.[2]

[1] Emile ter Kuile was one of the first to maintain this view (Pflüger's "Archives," 1900, Vol. LXXIX, p. 146).

[2] The impedance of the air is 43 and that of the water 144,000. The transformer ratio which will transform a maximum amount of power from one medium to the other is the square root of the ratio of impedances, or 58. However, the body of liquid in the ear is so small that it probably moves bodily back and forth and is not compressed. Consequently, the impedance at the oval window may be very different from that offered by the same area in a large body of water.

The relative sizes of the parts of the inner ear may be judged from Fig. 64,[1] which shows the cochlea uncoiled. It is seen that the length of the uncoiled cochlea is about 31 millimeters. The cross-sections of the cochlear passages on each side of the basilar membrane vary as one goes from the oval window to the helicotrema as indicated in the figure. The area of the stapes where it fits into the oval window is seen to be about 3 square millimeters. The opening between the two chambers at the helicotrema is about one-quarter millimeter. The greatest cross-section of either of these canals is

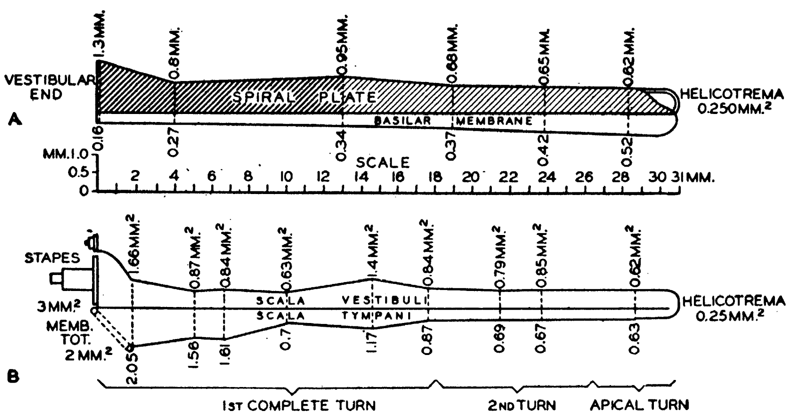


FIG. 64.—*A.* THE DIMENSIONS AND SHAPE OF THE HUMAN BASILAR MEMBRANE AND BONY SPIRAL LAMINA. *B.* DIAGRAM OF THE SECTIONAL AREAS OF THE COCHLEAR PASSAGES (DRAWN TO SCALE). IN EACH CASE THE ACTUAL MEASUREMENTS AND SCALE OF MEASUREMENTS ARE GIVEN.

less than 2 square millimeters. These figures emphasize the fact that this important mechanism of hearing is really very small. As mentioned above, the nerve terminals are scattered along the basilar membrane, and all the differentiations of complex sounds which are heard are made possible by the corresponding stimulation patterns produced in this membrane only one-quarter millimeter wide and about 31 millimeters long.

[1] The dimensions for this figure were obtained from the book by Wrightson and Keith entitled "An Inquiry into the Analytical Mechanism of the Internal Ear."

Functions of the Various Parts of the Ear While Sensing a Sound

There have been many theories proposed which describe the various functions performed by the different parts of the ear when sensing a sound. Most of these theories originated before there was much quantitative data concerning the facts of audition. During the last few years, due to the accumulation of such data, the evidence has been overwhelmingly in favor of a theory which is an extension of that originally proposed by Helmholtz. Only this theory will be described here. Those interested in some of the other theories will find good discussions of them in various recent publications.[1]

The Helmholtz theory is frequently called the resonance theory or the Harp theory. Either of these names gives rise to a wrong conception as to what Helmholtz really intended. In order to obtain a true picture of the Helmholtz theory I am giving below some abstracts taken from his book entitled "Sensations of Tone." These are sufficient to give the essential elements of his theory.

"When the drumskin is driven inwards by increased pressure of air in the auditory passage, it also forces the auditory ossicles inwards, as already explained, and as a consequence the foot of the stirrup penetrates deeper into the oval window. The fluid of the labyrinth, being surrounded in all other places by firm bony walls, has only one means of escape,—the round window with its yielding membrane. To reach it, the fluid of the labyrinth must either pass through the helicotrema, the narrow opening at the vertex of the cochlea, flowing over from the vestibule gallery into the drum gallery, or, as it would probably not have sufficient time to do this in the case of sonorous vibrations, press the membranous partition of the cochlea against

[1] Boring, E. G., "Auditory Theory with Special Reference to Intensity, Volume, and Localization," *American Journal of Psychology*, April, 1926, Vol. XXXVII.

Fletcher, H., "Physical Measurements of Audition and Their Bearing on the Theory of Hearing," *Journal of the Franklin Institute*, Vol. 196, No. 3, September, 1923.

Knudsen, V. O., and Jones, I. H., "Facts and Theories of Audition, Annals of Otology, Rhinology and Laryngology," December, 1925, and March, 1926.

Wilkinson, George, and Gray, Albert A., "The Mechanism of the Cochlea."

Wilkinson, George, "Is the Question of Analysis of Sound by Resonance in the Cochlea by Central Analysis Still an Open One?," *American Journal of Psychology*, April, 1927, Vol. XXXVIII.

the drum gallery. The converse action must take place when the air in the auditory passage is rarefied.

"Hence the sonorous vibrations of the air in the outer auditory passage are finally transferred to the membranes of the labyrinth, more especially those of the cochlea, and to the expansion of the nerves upon them."

. . .

"Hence when we hereafter speak of individual parts of the ear vibrating sympathetically with a determinate tone, we mean that they are set into strongest motion by that tone, but are also set

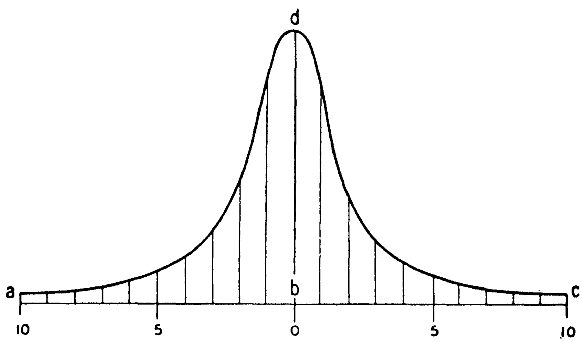


FIGURE 65.

into vibration less strongly by tones of nearly the same pitch, and that this sympathetic vibration is still sensible for the interval of a Semitone. Figure 65 may serve to give a general conception of the law by which the intensity of the sympathetic vibration decreases, as the difference of pitch increases. The horizontal line *a b c* represents a portion of the musical scale, each of the lengths *a b* and *b c* standing for a whole (equally tempered) Tone. Suppose that the body which vibrates sympathetically has been tuned to the tone *b* and that the vertical line *b d* represents the maximum of intensity of tone which it can attain when excited by a tone in perfect unison with it. On the base line, intervals of 1/10 of a whole Tone are set off, and the vertical lines drawn through them show the corresponding intensity of the tone in the body which vibrates sympathetically, when the exciting tone differs from a unison by the corresponding interval."

. . .

"Under these circumstances the parts of the membrane in unison with higher tones must be looked for near the round window, and those with the deeper, near the vertex of the cochlea, as Hensen also concluded from his measurements. That such short strings should be capable of corresponding with such deep tones, must be explained by their being loaded in the basilar membrane with all kinds of solid formations; the fluid of both galleries in the cochlea must also be considered as weighting the membrane, because it cannot move without a kind of wave motion in that fluid."

. . .

"The 4200 Corti's arches appear then, in this respect, to be enough to apprehend distinctions of this amount of delicacy. But even if it should be found that many more than 4200 degrees of pitch could be distinguished in the Octave, it would not prejudice our assumption. For if a simple tone is struck having a pitch between those of two adjacent Corti's arches, it would set them both in sympathetic vibration, and that arch would vibrate the more strongly which was nearest in pitch to the proper tone. The smallness of the interval between the pitches of two fibres still distinguishable, will therefore finally depend upon the delicacy with which the different forces of the vibrations excited can be compared. And we have thus also an explanation of the fact that as the pitch of an external tone rises continuously, our sensations also alter continuously and not by jumps, as must be the case if only one of Corti's arches were set in sympathetic motion at once."

. . .

"The sensation of different pitch would consequently be a sensation in different nerve fibres. The sensation of a quality of tone would depend upon the power of a given compound tone to set in vibration not only those of Corti's arches which correspond to its prime tone, but also a series of other arches, and hence to excite sensation in several different groups of nerve fibres.

"Physiologically it should be observed that the present assumption reduces sensations which differ qualitatively according to pitch and quality of tone, to a difference in the nerve fibres which are excited."

. . .

It is seen that according to this view when a sound wave impinges upon the ear drum its vibrational motion is communicated through the middle ear, and, as stated above, its amplitude is decreased to about one-sixtieth and the force or pressure variation increased correspondingly as it enters the inner ear or cochlea. Here the vibration is communicated to the fluid of the scala vestibuli. If the pitch of the tone is low, say, below 20 cycles per second, this fluid in the scala vestibuli and the scala tympani is moved bodily back and forth through the helicotrema. The motion between the round window and the oval window is just opposite in phase, the former moving inward while the latter moves outward. Thus, at the very low frequencies the mass reaction of the fluid is not sufficient to cause any appreciable transverse motion of the basilar membrane and consequently limits the lower pitch range of audibility.

At the very high frequencies the mass of the ossicles is so great that very little energy can be transmitted to the cochlea. When the elastic forces are negligible, it requires a force ten thousand times larger to produce a given amplitude at 10,000 cycles than at 100 cycles. For this reason it is probable that the factor which controls the upper limit of pitch audibility is the mass reactions involved in the ear rather than any lack of nerve sensitivity. For intermediate frequencies the mass reactions, the elastic restoring forces, and the frictional resistances which are brought into play are such that the vibratory energy is transmitted through the basilar membrane [1] at certain points causing the nerves to be excited.

From experiments on the differential sensitivity of the ear for pitch, which will be discussed in Chapter IV, it has been calculated that tones of various pitch are sensed by nerves at various positions along the basilar membrane as indicated in Fig. 66. If a pure tone of vibrational frequency of 1000 is communicated to the ear, the energy passes through the ossicles of the middle ear and sets the fluid in the scala vestibuli into vibration. The vibration is communicated through the canal until it gets half way up the cochlea, where, due to resonance, the vibrational energy is conducted through the

[1] As described above, the tectorial membrane lies over the basilar membrane so that if either one or the other or both vibrate when a wave is transmitted from the scala vestibuli to the scala tympani, the hair cells will be stimulated. In the above discussion, these two membranes are considered as one and referred to as the basilar membrane.

basilar membrane and then down through the scala tympani to the round window. The nerves which are stimulated most are those which are near the mid-point of the basilar membrane. The pitch or position of a tone on the musical scale is then dependent upon the position of the maximum stimulation along the basilar membrane. The character of the sound depends upon the positions and the relative intensities of agitation of the various parts of the basilar membrane.

Although the mechanics of the cochlea as sketched above is undoubtedly correct in its essential feature, it does not follow that the brain depends entirely upon the space pattern of stimulated nerves for determining the pitch and quality of

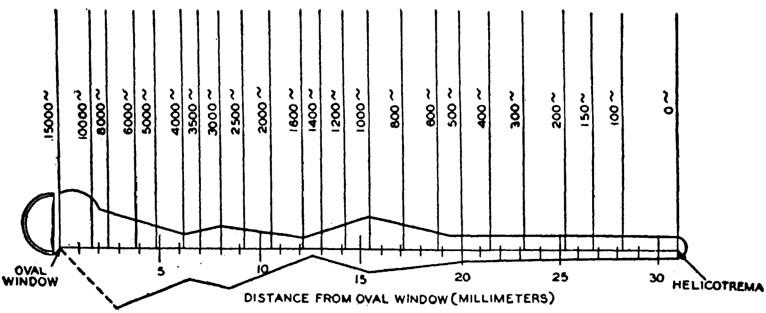


FIG. 66.—CHARACTERISTIC FREQUENCY REGIONS ON THE BASILAR MEMBRANE.

tones. Some of the features of the original sound wave may be preserved in the combination of the nerve impulses being sent to the brain by the individual nerve fibres. The time pattern of stimulation which is impressed upon the brain may therefore aid in making the proper interpretations. This is certainly true for long-time intervals. Some authors state that the first of these, namely, the space pattern, is sufficient to account for the recognition of pitch and quality while others insist that the second of these, namely, the time pattern, is the thing which is used to recognize pitch and quality. No doubt they both aid in varying amounts depending upon the character of the sound and the condition under which one listens to it.

There are some direct experimental evidences which indicate in a general way that the positions for sensing the tones of various pitches given in Fig. 66 are correct. The best known are those of Yoshii, Wittmaach, and Marx [1] on guinea pigs. These guinea pigs were kept in a continuous sound of fixed pitch for several hours a day for a long time. They were then killed and a histological examination made. All these observers noted that the low tones caused a degeneration of the end organ near the apical end, while the high tones caused a degeneration near the basal end, that is, near the oval window. They also observed that when tones in the middle register were used, a degeneration occurred near the middle of the basilar membrane.

There are three anatomical facts which show how the cochlea can act in a mechanical way to separate the tones of different pitch in the manner described: first, the basilar membrane increases in width more or less regularly and continuously from the base to the apex of the cochlea; second, the transverse fibres constituting the basilar membrane decrease in tension from the basal to the apical end of the cochlea; and third, the vibrating mass of fluid becomes greater as the stimulated spot goes from the basal to the apical end.

As a first approximation, we can treat the mechanical system as one having a single degree of freedom, that is, having a single moving mass constrained by elasticity and resistance. The vibrating mass is principally the fluid in the scala vestibuli from the oval window to the stimulated spot and back through the scale tympani to the round window. From Fig. 64 it was estimated that this length varied from 2 millimeters to 60 millimeters or a ratio of 30 from one end of the basilar membrane to the other. The elasticity is furnished mainly by the transverse fibres of the membrane which have a variation in length of 1 to 4. The variation in tension is unknown although the structure of the membrane at various positions along its length indicates clearly that the tension decreases rapidly

[1] Yoshii, *Zeitsch. f. Ohrenheilk*, Bd. 59, 1909, pp. 201–501.
Wittmaach, *Zeitsch. f. Ohrenheilk*, Bd. 54, 1907, and Bd. 59, 1909.
Marx, *Zeitsch. f. Ohrenheilk*, Bd. 59, 1909.

from the base to the apical end. Since for such a system the frequency is given by

$$f = \frac{1}{2\pi}\sqrt{\frac{E}{m}} \tag{1}$$

where E is the elasticity and m the mass, the variation in the vibrating mass and in the length of the transverse fibres over the length of the basilar membrane would cause a frequency variation of 11-fold. Inasmuch as one can hear through a frequency range 1000-fold, the tension must vary through a range of 8000 to 1. This seems like a rather large variation although it is not entirely impossible as is pointed out by Wilkinson and Gray.[1] However, the treatment of the cochlea as such a simple vibratory system may lead to results which are very far from the truth. In the above computation no account was taken of the frictional forces which may be the controlling forces for such small openings. For example, the canal on either side of the membrane is less than a square millimeter in cross-section. When all the factors are taken into account, it may not require such a large variation in the tension of the fibres to satisfactorily cover the frequency range that is used in hearing. Some further investigation along this line is badly needed. The problem is difficult because exact information of the mechanical constants involved is not available.

Another fact of importance in the hearing mechanism is that the chain of bones in the middle ear has a non-linear transmission characteristic. In the process of transmitting complex tones this part of the ear acts like a detector tube in a radio circuit. When a single tone is transmitted through to the cochlea, not only the impressed frequency but, in general, all its harmonics are sent into the cochlea, and the magnitude of the harmonics becomes greater compared to the fundamental as the intensity of the tone is increased. Similarly, when two tones are transmitted, the harmonics of each tone and also of the summation and difference tones are

[1] Wilkinson and Gray, "Mechanism of the Cochlea," p. 67.

transmitted to the cochlea. It is thus seen that when loud complex sounds are sent into the ear a very complex pattern is set up on the basilar membrane. For this reason one might be justified in saying that the character of the sound is interpreted according to the pattern existing on the basilar membrane. A recognition of this fact has led Scripture and others to reject the Helmholtz resonance theory and replace it by what they call a pressure pattern theory.

However, in speaking of the elements of the ear mechanism as having resonant elements, it must not be understood that this resonance is of the same type that would exist in a tuning fork or stretched string where the tone exists a considerable time after the driving force is removed. Due to the very small dimensions in the ear the frictional forces are very large and consequently the damping is great. It probably is as great, if not greater, than that existing in a telephone receiver. Consequently, the hangover, that is, the vibration existing after the stimulating tone has ceased, is so small that it is almost imperceptible. In listening to a continually varying source of sound the form of the vibration of the basilar membrane is set up and dies down so quickly that it follows changes with no perceptible delay. For this reason when the ear mechanism is compared to a harp or a piano a wrong impression is usually created.

Mechanism of Nerve Conduction

A complete theory of hearing must include an explanation of the nerve action which takes place after stimulation. In it probably lies the answer to the question: "How does the ear sense loudness?" The auditory nerve is very similar to a cable trunk. It contains about 3000 medullated nerve fibres, each consisting of an "axis cylinder" surrounded by a fatty substance called the myelin. The axis of this cylinder has a diameter of about .001 centimeter and forms only about 9 per cent of the fibre. It is thus seen that nerve fibres are constructed very much like insulated telephone wires and bound together in a strikingly similar manner to telephone cables. This analogous structure led some physiologists to the conclusion that all nervous impulses were electrical in origin and that their transmission was very similar to the electrical transmission on telephone lines. This theory, however, was found to be untenable.

Most of the earlier experiments on nerve conduction were made with motor nerves so that the mechanism of nerve conduction described here is based upon such experiments by several investigators. However, the recent work of Adrian [1] shows that the action in the sensory nerves is essentially the same. A nervous impulse may be excited by heat; chemical, electrical, or mechanical stimuli; or by reflex stimuli. Touching the nerve with a red-hot iron, with an acid, or pinching it, sets up a nervous impulse. The most common method in the physiological laboratories of exciting such an impulse is to use the shock obtained from a "make-and-break" induction coil. In order to set up such an impulse, the strength of the stimulus and its rate of change must be greater than a certain minimum. It has been found that the nervous impulse which travels along the nerves is not at all analogous to an electrical current travelling along a wire. An elemental nerve fibre has no impulse at all or else it fires with its full force. In other words, in normal nerve fibres the impulse is either of normal strength or zero strength throughout its entire course and seems to be the same regardless of how it is stimulated. The minimum value for starting the full nervous impulse is different for the different nerve fibres constituting the nerve.

In describing nerve conduction, physiologists frequently say it is similar to what goes on when a gunpowder fuse is lighted. The rate of the fire travelling down the fuse and the intensity of the heat which it creates are in no way dependent upon the way the fuse was lighted at the end. From this point of view it is seen to be a necessary condition that the loudness produced by a tone exciting the ear must be directly related to the number of fibres being excited and the rate at which the

[1] Book entitled "The Basis of Sensation," published in 1928.

excitations occur, since each fibre always carries its maximum impulse. It would seem necessary also that the minimum stimulus to excite each fibre must differ greatly.

That this is true is beautifully illustrated by some experimental work [1] of Porter and Hart. The nerves controlling muscle contraction were stimulated by electric shocks. The currents producing these shocks were gradually increased. The successive contractions of the muscles did not increase gradually but in definite steps as shown in Fig. 67. This figure is a record taken from their experimental work. The height of each line is a measure of each successive contraction.

If the auditory nerves act in a similar way, then, as a tone is gradually increased in intensity from below the threshold to loud values, the excitations reaching the brain must increase in definite steps, the threshold corresponding to the first nerve fibre being excited. When all of the nerve fibres are excited and firing at their maximum rate no further increase in loudness is possible.

After a nervous impulse has passed down the nerve, there is a "refractory" phase during which time the nerve is unable to respond or conduct. Then follows a "relative refractory" period during which the excitability, the conductivity, and the speed of propagation gradually return from zero to normal. As the condition of the nerve returns to normal, it overshoots the mark and becomes supernormal; that is, it is more sensitive, more highly conductive, and the speed of propagation is greater. This supernormal condition gradually dies away, until the nerve is once again in its normal stage. It is only during the relative refractory phases that the nerve conducts impulses reduced in magnitude. The length of this refractory period has been measured by several observers, and, although there is a wide disagreement, the best estimate at the present time seems to place it at about .001 second and the relative refractory period at about .003 second. According to these figures, the maximum number of nervous impulses which a

[1] Porter, E. L., and Hart, V. W., "Reflex Contractions of an All or None Character in the Spinal Cat," *American Journal of Physiology*, October, 1923.

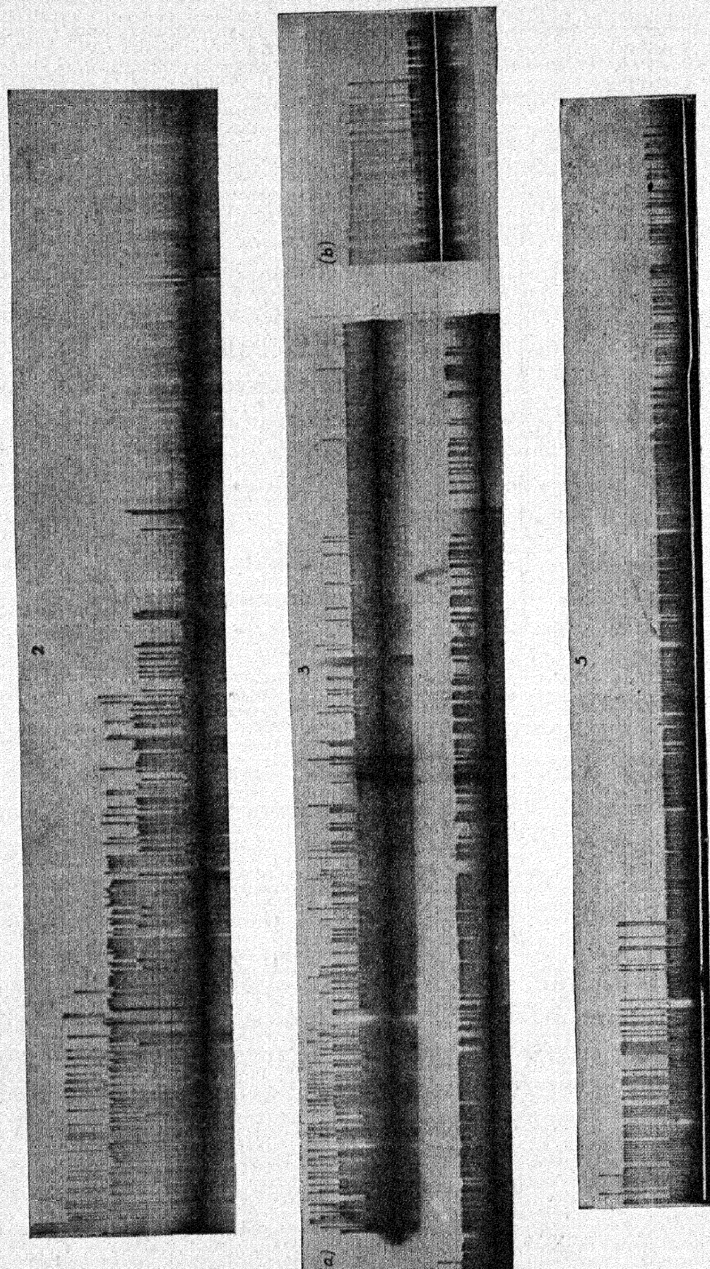


FIGURE 67.

single nerve fibre can send to the brain is 1000 per second. Those periodic excitations greater than 300 per second will not be transmitted as normal impulses, since each succeeding excitation will lie in the relative refractory period. According to Adrian's work the nerve ending has a greater relative refractory period than the nerve fibre itself. In his work on sensory nerve endings no frequencies of discharge greater than 150 cycles per second were observed. Consequently when a pure tone having a frequency of 2000 or 3000 cycles excites the ear it is probable that the number of nervous impulses being sent to the brain per second by each nerve fibre is considerably less than the exciting frequency. This is a very important fact, but up to the present no direct experimental evidence is available to determine the rapidity of the impulses being sent to the brain when the auditory nerve is excited by tones of various pitch. Experiments with transmission circuits indicate that no serious interference to the recognition of speech or music occurs when differences between arrival times of various component frequencies are less than .001 second.

The experimental work of Adrian shows conclusively that as the intensity of stimulation increases, the rate of nervous discharge also increases. This rate also depends upon the rate of change in the intensity of stimulation. To explain some of the facts of binaural audition, it seems reasonable to suppose that when the nervous discharge takes place it will always occur at the same phase of vibration, that is, when the basilar membrane is at a maximum amplitude or a maximum velocity or at some definite time interval between these two extremes.[1] The discharge will not take place at every vibration, but the stimulation may be stored up until enough has accumulated to produce the discharge. This may occur only once in 10 or once in 100 vibrations for a given fibre, or it may occur when highly stimulated at every vibration, but it cannot ever

[1] According to Sir Thomas Wrightson (see his book, "An Inquiry into the Analytical Mechanism of the Internal Ear"), a nerve stimulation takes place four times during each cycle at the crest, at the trough, and at those times when the membrane goes through the equilibrium position.

occur faster than the refractory period. According to this view different nerve fibres will discharge at different times but always at the same phase of vibration of the basilar membrane. Consequently, a nerve composed of a bundle of nerve fibres will carry a nerve current to the brain which produces a time stimulation pattern there that has the same periodicity as the sound waves producing it.

With this picture of the nervous mechanism in mind it is not unreasonable to assume that the loudness (magnitude of sensation) of a sound is directly related to the total number of nervous discharges coming to the brain. This number is dependent both on the number of nerve fibres stimulated and on the intensity of stimulation of each one. According to this view, consider then the loudness effects produced when a 250-cycle pure tone increases in intensity from the threshold of hearing to very high values. At first a single fibre is sufficiently stimulated to cause nervous discharges at a slow rate. As the intensity increases, the rate of these discharges in the first nerve fibre increases while at the same time other fibres start discharging. This continues until a small patch of nerve endings at a position 25 millimeters from the oval window is stimulated. In this patch some nerve fibres are firing these nervous discharges much faster than others, the rate depending upon their distance from the position of maximum stimulation and also upon their initial sensitivity. The pitch is determined by the position of maximum stimulation and the loudness by the total number of discharges from all the nerve fibres in the stimulated patch. As the intensity is still further increased, other patches at positions corresponding to the subjective tones are added to those already stimulated, first at 21 millimeters, then at 18½ millimeters, then at 16½ millimeters from the oval window, and so on. These subjective tones are produced during the transmission of the sound through the mechanism of the middle ear. At very high intensities these patches overlap so that some fibres at all positions along the membrane will be discharging. When two tones are impressed upon the ear a similar thing happens except that besides the patches due to the subjective harmonics, others will appear due to the subjective, summation and difference tones. For any complex tone a corresponding pattern of stimulation will be produced, which will depend upon the intensity of the sound received at the ear as well as upon its physical characteristics. This pattern determines the character of the sound which is perceived.

This general picture of how the ear works will aid in interpreting the various experimental facts of audition which will be discussed in the succeeding chapter.

CHAPTER II

LIMITS OF AUDITION

WHEN the intensity of a sound is continuously decreased it reaches a value where it produces no stimulation of the auditory sense. The intensity which is just sufficient to be heard is called the "threshold of audibility." It is the lower intensity limit of audition. If the intensity is continuously increased it reaches an intensity which stimulates the sensation of feeling. This intensity is called the "threshold of feeling." Since intensities higher than this cause pain and injure the hearing mechanism, this threshold of feeling serves as a practical upper intensity limit to sounds which can be sensed by the human ear. If a tone is kept at a given intensity and at the same time gradually raised or lowered in pitch it ceases to be sensed by the ear at both an upper and a lower pitch limit.

It is the purpose of this chapter to give data concerning the limits of audition; and also to describe experiments which enabled such data to be taken.

Threshold Intensity vs. Frequency

During the past century a number of observers have made measurements of the intensity at the threshold of audibility. The results of these measurements have been interesting not only to physicists, but also to a number of the other scientific groups. A description of the methods which have been used by some of these investigators may be of interest.

IN 1870 Toepler and Boltzmann [1] made a determination of ear sensitivity. The amplitude of vibration of the air particles

[1] Ann. der Phys., Vol. 141, p. 321, 1870.

in an organ pipe was determined by light interference methods. From the distance to the source at which sound was just audible it was possible to determine the amplitude of vibration of the tone at the threshold of audibility.

In 1877 Lord Rayleigh [1] used a whistle as a source of sound and calculated the energy emitted by it from the pressure used in blowing it. He also used a tuning fork mounted on a resonator. From the difference in the decay constants of the fork suspended freely in the air and mounted on the resonant box it was possible to calculate approximately the energy emitted by the box. He made a third measurement using a telephone receiver as a source. The deflection of the diaphragm for direct current was considered the same as for an alternating current when the period was far below the natural period. The former was measured microscopically and consequently when the volume of air enclosed in the ear is known, it is possible to calculate approximately the change in pressure on the ear drum from the current flowing in the receiver.

In 1883 Wead [2] used a vibrating tuning fork in an open field as a source of sound. The amplitudes of vibration were made large enough to be directly measured. From the decay constants and the time elapsed before the tone disappeared, the absolute value of the threshold intensity was obtained.

In 1903 Wien [3] used a telephone receiver as a source of sound, making direct measurements of the amplitude of vibration for loud sounds. By assuming that the amplitude increases proportionally with the current, it is possible to calculate the amplitude of vibration of the diaphragm at the threshold of audibility. He observed results through a range of frequencies from 50 to 16,000 cycles. These results were generally considered the most reliable until the recent work using vacuum tubes and thermophones.

In 1904 Webster [4] used for his source a so-called "Phone,"

[1] *Proceedings of Royal Society*, Vol. 26, p. 248, 1877.
[2] *American Journal of Science*, 151, Vol. 26, p. 177, 1883.
[3] Wien, *Archiv für die gesamte Physiologie*, 97, pp. 1–57 (1903).
[4] Boltzmann, F. L., *Festschr.*, Leipzig, 1904, p. 1866.

an instrument so constructed that the amount of sound energy emitted by it can be calculated.

In 1905 Abraham [1] used as a source of sound a telephone receiver attached to a brass cylinder, the diaphragm forming its base and an ear piece its top. The change in pressure in the cylinder for a direct current in the receiver was determined by a sensitive manometer. He obtained approximately the same sensitivity for the two frequencies, 250 and 500 cycles. These frequencies were well below the natural period, so that the same proportionality factor was used for obtaining the pressure change as was obtained by the direct current measurement.

Due to the importance of knowing the absolute sensitivity very accurately, Bell Telephone Laboratories took up the problem in 1920 and published the results of their work in 1922.[2] During this same period the problem was also being worked upon by Kranz whose results were published in 1923.[3]

The methods used in this recent research work are based upon the thermophone formula developed by Arnold and Crandall [4] and later modified by Wente.[5] (See Appendix A.) When an alternating electrical current is superimposed upon a direct current and sent through a very thin metal strip, it generates a sound wave. By means of the formula mentioned one can calculate the amplitude of the pressure variation produced when the metal strip is enclosed in a small gas chamber.

The new tools besides the thermophone which made it possible to obtain more accurate data concerning threshold intensities were vacuum tubes (in the form of oscillators, amplifiers, and rectifiers), condenser transmitters, and attenuators accurately calibrated throughout a wide range of intensity. In the investigations described in the previous chapters these same

[1] *Comptes Rendus*, Vol. 144, p. 1099, 1907.

[2] Fletcher, H., and Wegel, R. L., "The Frequency-Sensitivity of Normal Ears," *Physical Review*, June, 1922.

[3] Kranz, F. W., "Minimum Intensity for Audition," *Physical Review*, Vol. 21, No. 5, May, 1923.

[4] Arnold, H. D., and Crandall, I. B., "The Thermophone as a Precision of Source of Sound," *Physical Review*, Vol. X, No. 1, July, 1917, pp. 22–38.

[5] Wente, E. C., *Physical Review*, Vol. 19, April, 1922, pp. 333–345.

tools were almost indispensable and are now being used in a great many other lines of investigation in acoustics.

In the work at Bell Telephone Laboratories two methods of determining the absolute values of the threshold of audibility were used. In the first method advantage was taken of the availability of a calibrated high quality telephone system. A schematic of such a system is shown in Fig. 68. By adjusting the potentiometer it was possible to make the tone coming from the system receiver, *B* (Case 1), sound as loud as that produced when the ear was held in the same position as the condenser transmitter, (Case 2). This reading of the attenuator was found for all frequencies from 100 to 2000 cycles. Since

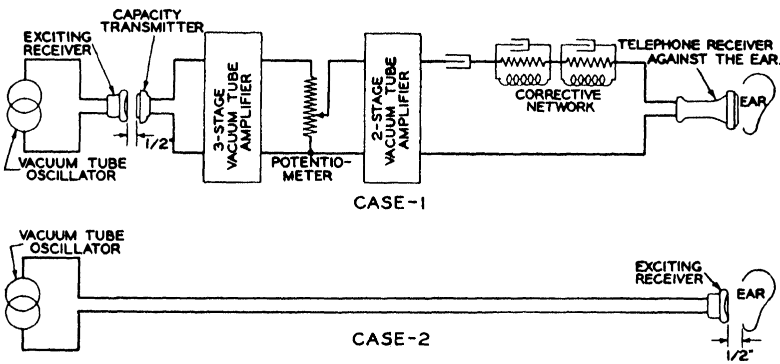


FIGURE 68.

the sound energy striking the condenser transmitter was known in terms of the voltages generated by it, being previously calibrated by means of the thermophone, the energy going from the receiver of the system into the ear of the observer could be calculated from the potential difference at the terminals of the condenser transmitter and the reading of the attenuator.

To make a measurement a potential difference of known magnitude and frequency was applied at the terminals of the condenser transmitter and sufficient attenuation was introduced into the system to make the sound from the receiver inaudible. The attenuation was then gradually removed until the sound just became audible. From the amount of attenuation for this condition and the voltage impressed upon the terminals of the condenser transmitter, the pressure variation in the ear was calculated. Four or five readings were taken in this way, which gave an average value having a probable error of 20 or 30 per cent in the determination of the pressure variation. This method was used for making tests with eleven different observers—seven men and four women—through a range of frequencies from 130 to 2000 cycles per second. In this work and also in that using the second method specially constructed sound-proof booths were used.

In the second method a thermal receiver unit small enough to be inserted in the external auditory meatus of the ear and to completely close it was used. It consisted of a series of short Wollaston wires enclosed in a small brass capsule with small holes communicating with the outside air. To calibrate it a small chamber was made in front of the diaphragm of a calibrated condenser transmitter by means of a coupler designed to fit over the face of the transmitter. The volume of air thus entrapped was made equal to that in the ear canal. The thermal receiver was inserted in a small hole through the coupler and then all the joints were sealed so as to make the chamber air-tight.

The determination of the relation between the voltage impressed upon the thermal receiver and the pressure exerted by it on the condenser transmitter diaphragm was then a simple matter. The voltage at the terminals of the thermophone was measured directly and the pressure exerted by it was calculated from the voltage produced at the terminals of the condenser transmitter.

Using such a calibrated thermal receiver inserted in the ear canal, the minimum audible pressure was determined by noting the minimum current for audibility. Measurements of the threshold of audibility were made with five people using first an air-damped telephone receiver and then two of these thermal receivers. From a comparison of the results the air-damped telephone receiver was calibrated. The probable observational error in this determination was found to be 8 per cent in the range of frequencies from 500 to 3000 cycles. The air-damped telephone receiver was then used to measure the threshold intensity for 102 ears of men and women of various ages. After the measurements were made, an otological examination revealed that some of the persons tested had defects in their hearing. After all doubtful cases were eliminated, data on 72 ears remained which were used in the final averages given below. The tests were made in sound-proof booths.

It is seen from the manner in which the measurements were made that the first method of calibration gives the pressure variation at the opening of the external ear provided that the ear reflects the sound waves in the same way as the condenser transmitter when placed in the same position. Also in the second method the pressure variation which is computed is that which would be exerted upon the ear drum provided that it had the same stiffness as the condenser transmitter diaphragm. Since the ear drum moves and its mechanical impedance at some frequencies may be comparable with that of the air chamber, the actual pressure variation against the ear drum may be somewhat less than that given by these observations.

Since these two methods gave results which were approximately the same it seems reasonable to assume that the pressures which are calculated by either method are not very greatly different from those which exist near the ear drum. It would be desirable to check this result by direct measurements of the pressure within the ear canal when apparatus and technic have been developed to a stage where this is possible.

In the Kranz[1] method a small thermophone was also inserted in the ear. It was constructed, however, so that the pressure variation in the ear canal could be directly computed by means of the Wente formula. Such a calculation depends upon knowing the volume of air in the canal and upon the

[1] *Loc. cit.*

assumption that all the walls including the ear drum are rigid. A schematic of the arrangements which Kranz used is shown in Fig. 69.

Kranz also repeated the method first used by Wien. He used a microscope to measure the amplitude of the motion of the diaphragm of the telephone receiver when it was driven at large amplitudes. For this purpose a small fibre of wood was attached to the center of the diaphragm. In order that this should not affect the results, when making threshold measure-

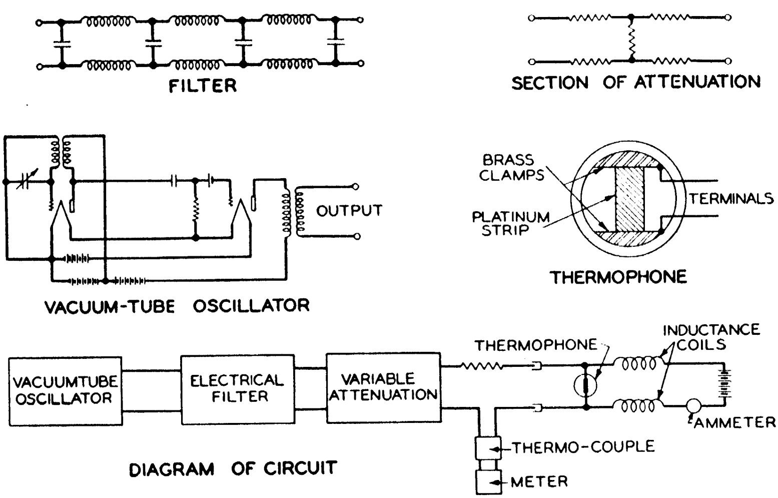


FIGURE 69.

ments, the wooden fibre was replaced by a small piece of brass which had the same weight.

Another means of determining the amplitude of the receiver diaphragm consisted in using a mirror so mounted as to rotate by movements of the diaphragm. A phosphor bronze wire of some stiffness was mounted to project from the center of the diaphragm and this rested against one sharp edge of a small rectangle of steel, another sharp edge being held against a copper block by the force exerted by a slight bending of the phosphor bronze wire. It is thus seen that movements of the diaphragm caused the wire to rock the steel piece about the edge which rested on the copper as an axis. A small mirror waxed on to the steel piece gave a broadening of a reflected line of light when the receiver diaphragm vibrated.

A fair agreement was found between the results obtained with the thermophone and with the Wien method, except at the low frequencies. In general, the thermophone method gave results which showed the ear to be slightly more sensitive. The work of Kranz was also done in sound-proof rooms.

In 1922 Lane published some results [1] for the threshold values for frequencies from 2000 to 18,000. In this investigation the tone generator developed by Hewlett[2] was used as a source of sound. The radiating surface of the tone generator consisted of a thin aluminum diaphragm 10 centimeters in diameter which was actuated by a flat coil of wire carrying an alternating current superimposed upon a direct current. The reaction between the electrical currents induced in the aluminum diaphragm and those in the coil set the diaphragm into vibration, thus causing the sound to be radiated. By making certain reasonable assumptions, it was possible to calculate the intensity of the sound at any given distance from this radiator when the alternating current actuating it was known. Lane's work on the threshold of audibility was done out of doors with observers on a small platform 5 meters above the ground and with the source of sound $1\frac{1}{2}$ meters from the ear. This was done at night so as to minimize the interference effect of noise.

In obtaining a final curve to represent the average pressure variation necessary to excite the auditory sense of persons having normal hearing, the results of only four of the observers mentioned above were considered of sufficient accuracy to be included. These observers are Wien, Kranz, Fletcher and Wegel, and Lane. For values between 64 and 4096 the final results were obtained by assigning weights of 3, 14, and 72 to the results obtained by the first three observers

[1] "Minimum Sound Energy for Audition for Tones of High Frequency," *Physical Review*, May, 1922.

[2] "A New Tone Generator," *Physical Review* (2), xix, January, 1922, p. 52.

mentioned. These weights are proportional to the number of ears tested in each case. The values for the higher frequencies were obtained from Lane's data. A correction of 6 db was applied to his data to bring them into line with the other data at frequencies between 2000 and 4000 cycles. This is justified, since the work was done out of doors where insect noise would produce a slight shift in the threshold.

The final values are given in Table XVI. The results of Wien in the region of 2000 cycles would indicate that the ear was more sensitive by about 35 db. This difference is much

TABLE XVI

Frequency (dv).......	64	128	256	512	1024
Pressure (bars).......	.12	.021	.0039	.0010	.00052
Power (watts $\times$ 10^{-12})	35	1.06	.036	.0024	.00065
Frequency (dv)......	2048	4096	8192	16,384	18,500
Pressure (bars).......	.00041	.00042	.0025	.13	4.1
Power (watts $\times$ 10^{-12})	.00040	.00042	.015	41	400,000

larger than any observational error would warrant. However, the work was carefully done and the method should yield correct results. The results of Kranz would indicate that in the middle pitch range the ear sensitivity was about 4 SU more than indicated by these figures and in the lower pitch range was about equal. Inasmuch as observations were not taken by the observers at all the octave frequencies given in the table, values for these frequencies were obtained from a curve connecting all the observed points. In this way the average values given above were obtained. In the third row of this table the values are expressed as the power in micro-microwatts passing through a square centimeter when a sound wave is travelling through air and producing the pressure changes indicated. These results are also shown graphically in the lower curve of Fig. 70.

The curve connecting the points is a smooth curve because each point represents an average of a large number of persons.

The curve representing the threshold of audibility for a single person is never such a smooth curve. Two such individual

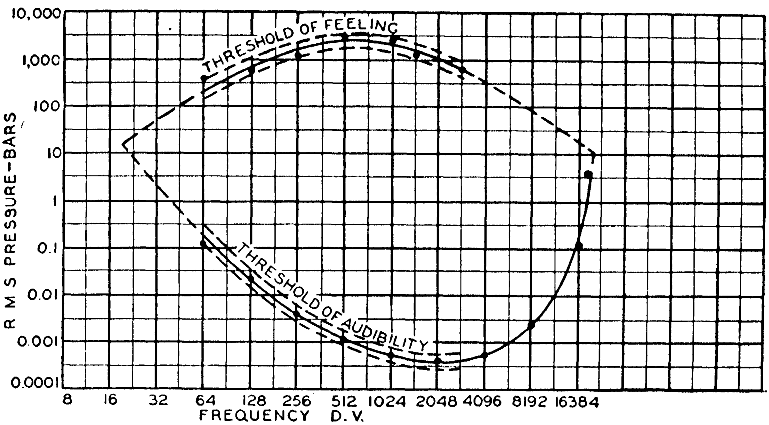


FIGURE 70.

curves for normal ears, one taken by Wegel and one taken by Kranz, are shown in Fig. 71. The small peaks and valleys are probably related to the particular structure of the ears of the two observers.

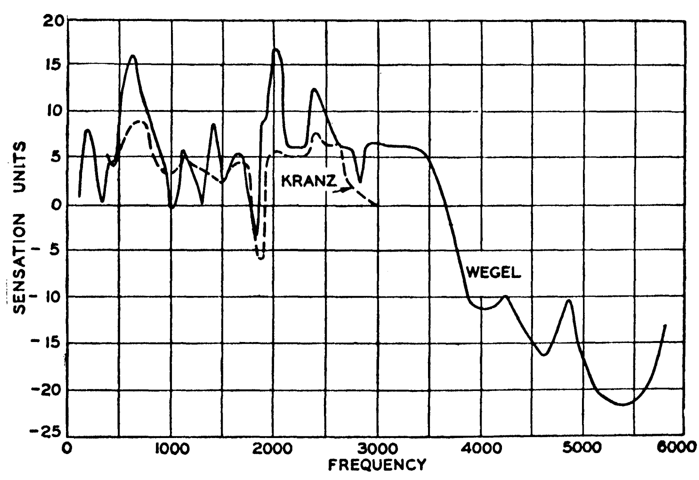


FIGURE 71.

An examination of the individual curves used to obtain the average values given above indicates that at some frequencies these curves depart as much as 20 db from the average. It is thus seen that each person has a hearing acuity which is peculiar to himself. Consequently, he learns to interpret the sounds about him with his particular hearing mechanism. The sensation produced by the same piece of music must be different for each person because of these individual variations.

The values given in Table XVI show what minute changes in the air pressure can be detected by the ear when these changes take place with a rapidity equal to that produced by voice sound waves. Since the atmospheric pressure is approximately 1,000,000 bars, it is seen that if the pressure is changed one-billionth of its total value, such a change is sensed by the ear. Tones producing 1 per cent variation in pressure are so intense as to injure the hearing mechanism.

Feeling Intensity vs. Frequency

When a tone exciting the ear is continuously increased in intensity, it finally reaches a loudness which produces the sensation of feeling. If the tone gets much louder than this value, it becomes painful. Although this threshold of feeling may have no relation to the auditory sense, it does serve as a practical upper limit for the intensities of tones which can be sensed by the ear. Measurements reported by Wegel [1] have indicated that this threshold of feeling can be as definitely determined as the threshold of audibility. The results of his measurements on forty-eight normal ears are plotted in Fig. 70. The same apparatus was used for this work as that used in determining the threshold of audibility. Both the curves for the threshold of audibility and the threshold of feeling have been extrapolated until they intersect. The feeling sensation in the middle range of frequencies is first a tickling sensation

[1] Wegel, R. L., "The Physical Examination of Hearing and Binaural Aids for the Deaf," published in *Proceedings of The National Academy of Sciences*, Vol. 8, No. 7, July, 1922.

and then it becomes actually painful as the loudness is increased. In the lower frequency range the sensation of feeling becomes milder until, at frequencies around 60 cycles, it is sensed as a flutter. As the frequency is still further decreased to the point where the two curves intersect, it is very difficult to distinguish between the sensation of feeling and the sensation of hearing. This same difficulty also exists at the upper limit.

Upper and Lower Pitch Limits

As pointed out by Wegel, this constitutes the first rational method of defining what is meant by the upper and the lower pitch limits of audibility. It is readily seen from these two curves that both the upper and lower limit of audibility on the pitch scale will be entirely dependent upon the particular intensity at which the measurements are made. When this intensity is such that it is just producing the sensation of feeling and also of hearing, the upper or lower limit of audibility is reached. The two curves may be considered as representing the limits of audition both on the pitch and the intensity scales. They are boundary lines separating those tones which can be sensed from the tones which cannot be sensed by an average normal ear. Parts of these boundary lines are dotted because accurate data are lacking for those particular regions. The two dotted lines on either side of the main boundary lines show the probable deviation of an observation made upon one particular person. In other words, one-half of such observations will lie within the two dotted curves as indicated.

A large amount of data has been taken by many observers on the upper and lower limits of audibility of pitch. An examination of most of this work indicates that not much attention was paid to the particular intensity used in such a determination. Also, the tones produced by whistles, by bowed strings, or by striking metal bars, the instruments usually used, were not pure tones. Also, it has been found that the variation in the sensitivity of the ears for high frequencies is very large among different individuals. Usually the number of individuals tested was too small to give a good average. The limits shown in Fig. 70 were chosen after consideration of all the available data and also in view of our own experiments in the laboratory. It is seen that the limits on the pitch scale chosen are from 20 to 20,000 cycles per second. These are average values. There is no doubt that some persons have a keen acuity for notes of high pitch and could possibly hear notes having a frequency much higher than 20,000 cycles per second. Also, it is possible that the auditory sense is stimulated in some persons by frequencies lower than 20 cycles per second. Organ pipes having a pitch lower than this have been constructed, but the sensation produced by them is probably due to the overtones rather than the fundamental tone being emitted.

The area enclosed by the two curves giving the threshold of feeling and the threshold of audibility is called the "auditory sensation area." To each point in it there corresponds a definite auditory sensation when the ear is acted upon by a tone having the frequency and the intensity indicated by the coordinates. Pure tones outside of this area produce no auditory sensation. It will be noticed that the scale of frequency and also the scale of intensity used in this chart are logarithmic. It is almost imperative that such scales be used when representing such large ranges. As will be seen from the discussions in the next chapter, the choice of such a scale is more in keeping with the way one perceives changes in pitch and intensity.

CHAPTER III

Minimum Perceptible Differences in Sound

There is a well-known law in psychology called the Weber-Fechner law which states that the increase of a stimulus necessary to produce a just discernible increase in the resulting sensation bears a constant ratio to the total stimulus. It is sometimes stated in the form that the magnitude of the sensation produced is proportional to the logarithm of the stimulus. If the same law applies to the hearing sensation, then the fractional increase in intensity, which is just perceptible as a change in intensity, should be a constant independent of the intensity. Similarly, the minimum fractional increase in frequency, which is perceptible to the ear as a change in pitch, should be constant. A number of different observers have made attempts to determine these ratios. The apparatus available for this work seriously limited its accuracy. Organ pipes, tuning forks, and sometimes falling steel balls hitting upon steel plates were used as sources of sound.

Minimum Perceptible Differences in Intensity

The recent work of Knudsen was more accurate than any of the previous observers, not only because of careful observations, but because he used the accurate tools described in the last chapter.

A schematic of his apparatus [1] for determining these two ratios is shown in Fig. 72. As indicated, the source of sound was a telephone receiver actuated by the electrical current from a vacuum tube oscillator. By means of the resistances

[1] Knudsen, V. O., "The Sensibility of the Ear to Small Differences of Intensity and Frequency," *Physical Review*, Vol. XXI, No. 1, January, 1923.

in the circuit, the output of the oscillator could be varied by any desirable measurable intervals. The circuit for measuring minimum perceptible changes in intensity is so designed that a motor-controlled key periodically changes the resistance *R* across which the receiver is shunted by any desired time intervals. The tone then emitted by the receiver will vary in intensity depending upon the speed of the motor which operates the key. Knudsen found that the best conditions for determining the minimum change were obtained when this key changed

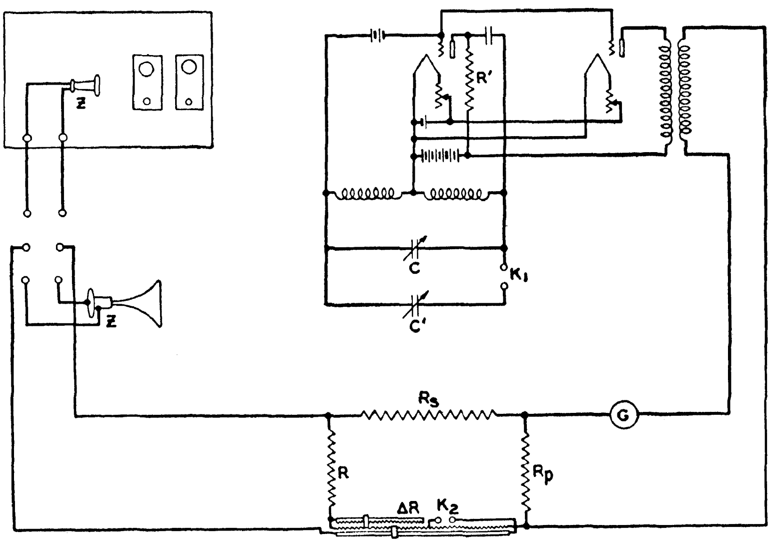


FIGURE 72.

the intensities at a rate of about fifty times per minute. The resistances are first adjusted so that the change in intensity is plainly perceptible. This change is gradually reduced until the just perceptible difference is determined.

Knudsen's apparatus, however, limited his observations to a frequency range of from 100 to 4000 cycles. It also limited the intensity range. To extend both these ranges Bell Telephone Laboratories took up the problem. The method used involves the principle of beats. Electrical currents from two oscillators producing slightly different frequencies are sent into a special telephone receiver. When the receiver is held to the ear beats are produced. The amount that the intensity of the tone fluctuates can be controlled by changing the relative magnitudes of the current from the two oscillators. At high intensities and low frequencies, it was necessary to use a special telephone receiver to avoid distortion. This receiver was of the moving coil type and would reproduce pure tones throughout the entire range in the auditory sensation area. The arrangement of the apparatus is shown in Fig. 73. The oscillator and the attenuator in circuit No. 1 were set so that the telephone receiver produced a tone having any desired frequency and sensation level. Oscillator No. 2 was then adjusted so as to produce a tone of slightly different frequency. As a result, beats were produced.

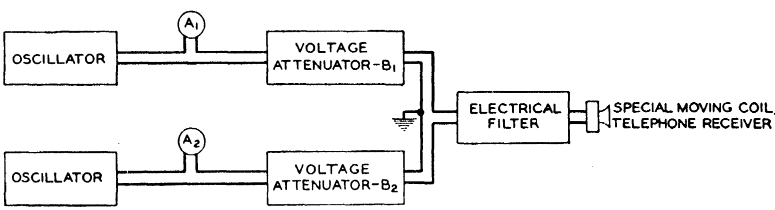


FIGURE 73.

Tests were made to determine the differential sensitivity of the ear for different rapidities of fluctuations of intensity. It was found that the ear was most sensitive when the number of beats per second was kept between 1 and 6. When the number became as low as 1 every 5 seconds or as high as 25 per second, then the change in energy necessary to be perceived was about three times greater than that obtained when the number of beats remained between 1 and 6. For this reason in the experimental work the rate of 3 beats per second was chosen as being best for perceiving small differences in intensity and all the results reported below were obtained by using this rate.

The minimum audible voltage of one attenuator, say, B_1, was determined, B_2 being set far below the minimum audible voltage. B_1 was then set at any desired value and B_2 adjusted until the observer signalled that he heard beats. The setting of B_2 was changed each time the observer signalled whether or not the tone seemed to fluctuate. After about twenty such judgments the operator was able to locate with considerable certainty the setting of B_2 for which the observer was just able to detect a fluctuation in intensity. If B_2 was set below this value the fluctuation in intensity was imperceptible. The

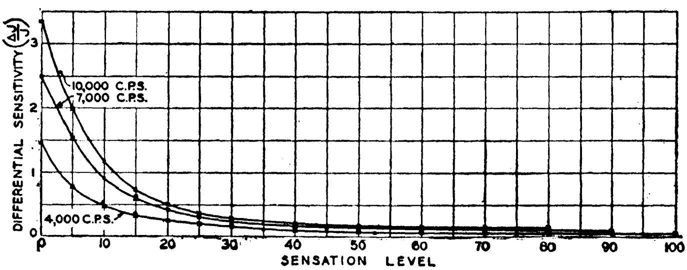


FIGURE 74.

r.m.s. voltage introduced into the receiver circuit by each oscillator could be calculated from the readings of the ammeters and the settings of the attenuators. At any frequency the r.m.s. alternating pressure on the ear drum is proportional to the voltage introduced into the receiver circuit. Complete series of measurements were made on twelve male observers at frequencies of 35, 70, 200, 1000, 4000, 7000, and 10,000 cycles per second and at intensities from weak tones near the threshold of audition to very loud tones near the threshold of feeling. The results of these measurements are shown in Figs. 74, 75, and 76. In the first two figures mentioned the abscissas represent sensation level and the ordinates represent differential sensitivity. In Fig. 76 the data are arranged to show the variation of differential sensitivity as the frequency changes.

For sensation levels above 50 db, the fractional change in the intensity which is just perceptible is between 5 per cent and 10 per cent. At a sensation level of 10 db the fractional increase must be 73 per cent to be perceptible. For frequencies as low as 60 cycles a change of 20 per cent is just perceptible

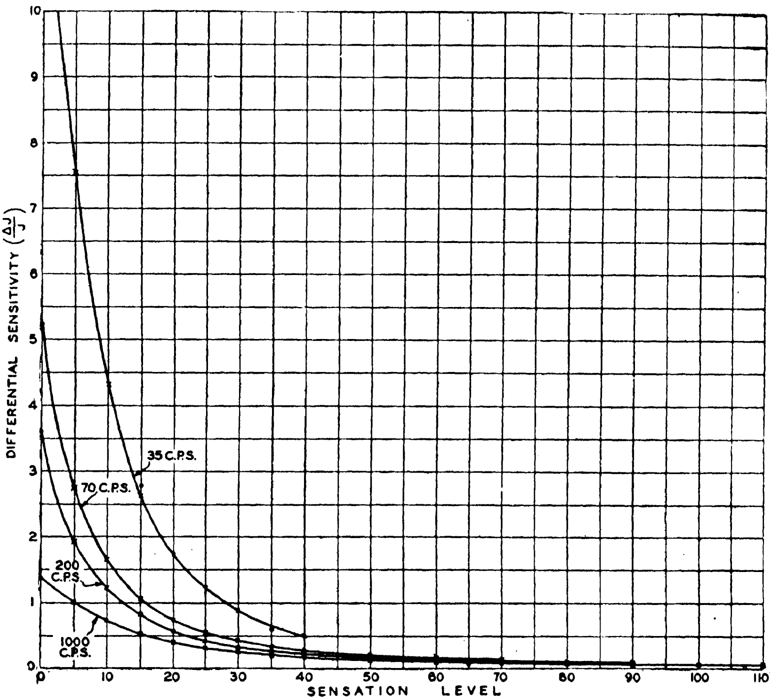


FIGURE 75.

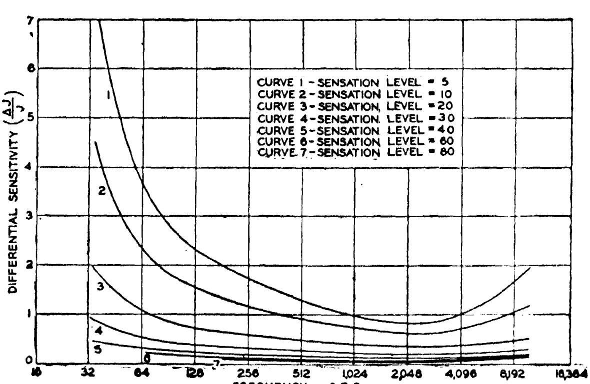


FIGURE 76.

at the high levels, and an increase as much as 200 or 300 per cent is necessary at levels as low as 10 db. This work was done

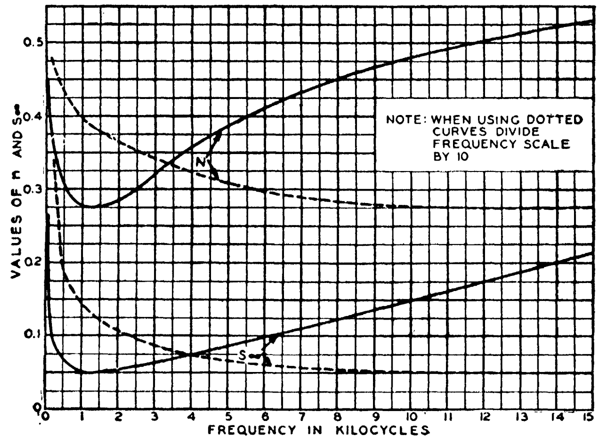


FIGURE 77.

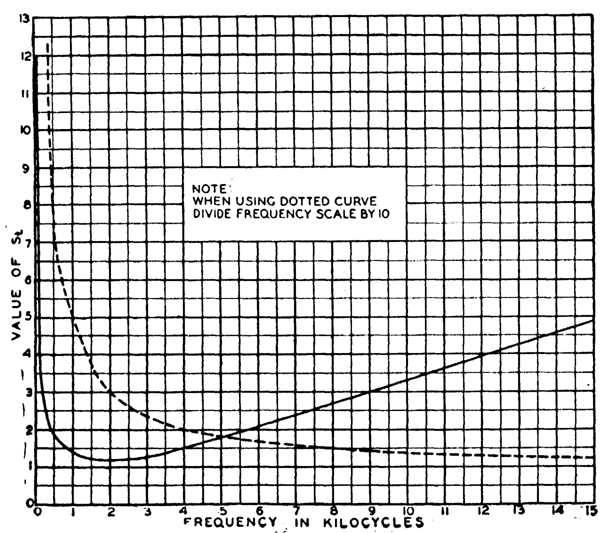


FIGURE 78.

by R. R. Riesz, who evolved the following formula for representing the results:

$$\frac{\Delta J}{J} = S_\infty + (S_t - S_\infty)10^{\frac{-\alpha_s n}{10}} \tag{1}$$

where

$$S_\infty = .000015f + \frac{126}{80f^{1/2} + f}$$

$$S_t = .3 + .0003f\frac{193}{f^{.8}}$$

$$n = \frac{24{,}400}{358{,}000f^{.8} + f^2} + \frac{.65f}{3500 + f}$$

and where $\frac{\Delta J}{J}$ (called the differential sensitivity) is the minimum fractional increase in the intensity that is just perceptible and α_s, the sensation level of the tone before the increase. It is evident from this formula that S_∞ is the differential sensitivity for high sensation levels and S_t the differential sensitivity for $\alpha_s = 0$ or at the threshold of audibility.

For convenience in calculation, the values of S_∞, S_t, and n are given in curves of Figs. 77 and 78. For example, consider the values for a 1000-cycle tone. The values of S_∞, S_t, and n are .051, 1.35, and .28, respectively. Then for this tone

$$\frac{\Delta J}{J} = .051 + 1.3 \times 10^{-.028\alpha_s} \tag{2}$$

Minimum Perceptible Differences in Frequency

To determine the minimum perceptible change in frequency, Knudsen used the same apparatus as was used to determine the minimum perceptible differences in intensity, except that a small capacity was added and subtracted periodically from the capacity in the oscillating circuit. By adjusting this added capacity to the proper value, any change in pitch that was desired could be obtained. The average values of $\frac{\Delta f}{f}$ which he obtained are shown in Fig. 79. For the higher and the lower ranges of pitch, the curves have been extrapolated beyond the observed data. For frequencies between 500 and 4000 the minimum fractional difference in frequency which is perceptible is .3 of 1 per cent. For the lower and the higher ranges of frequencies, it requires a greater fractional change in frequency to cause a perceptible change in pitch. Only a small amount of data is available which shows how $\frac{\Delta f}{f}$ varies with different sensation levels but it indicates that $\frac{\Delta f}{f}$ becomes larger in about the same way as $\frac{\Delta J}{J}$ becomes larger as the sensation level becomes lower. The data given in Fig. 79 correspond to a sensation level of 40 db.

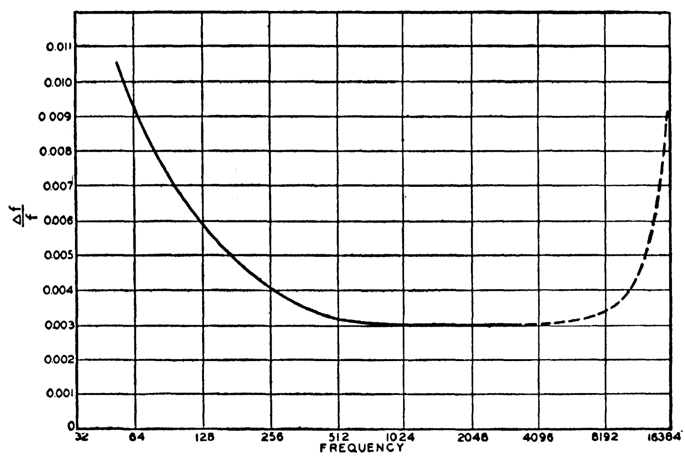


Fig. 79.—Minimum Perceptible Difference in Frequency.

Minimum Time for Tonal Perception

Another measurement which is of particular interest to psychologists is the minimum time a pure tone must excite the ear in order that it be sensed as a tone having a definite pitch. Here again the data available are rather discordant but the most probable values are given in Table XVII.

TABLE XVII

Freq.	Weak Tones		Medium Tones	
	Time (Sec.)	Cycles	Time (Sec.)	Cycles
128	0.0946	12.1		
256	....	....	0.06908	17.6
384	0.0627	24.08	0.0445	17.1
512	0.0579	29.64	0.04274	21.8

From such limited data it is difficult to make any safe generalizations but it is seen that the time is approximately independent of the frequency of the tone and is about one-twentieth of a second. Most of the unvoiced stop consonants have a duration less than this in ordinary speech so that apparently no sense of pitch plays a part in their interpretation.

Levels of Frequency (Pitch) and Levels of Intensity (Phonic Level)

In Part One, Chapter III, the sensation level of a sound reaching the ear is defined as the number of decibels above the average threshold for normal ears. The relation between sensation level and intensity of the sound is dependent upon the character of the sound so that for complex tones, no simple relation exists between the two quantities. As will be discussed in a later chapter, two sounds are equally loud when they produce the same magnitude of sensation. Experiments have also shown that there is no simple relation between loudness and intensity. For these reasons, the logarithmic scale of intensity is called "level of intensity" rather than "sensation level" or "loudness." The most logical relation then between the intensity level α and the intensity J is given by the equation

$$\alpha = A \log \frac{J}{J_0} \qquad (3)$$

where A and J_0 are arbitrary constants to be chosen to give the most convenient scale. If the bel is the unit used for representing the level and the microwatt per square centimeter for representing the intensity and also if the comparison intensity is taken as one microwatt then

$$\alpha \text{ (bels)} = \log_{10} J \text{ (microwatts)}.$$

If the db is used, then A must be put equal to 10. As used in the International Critical Tables the intensity level is called phonic level. According to this definition, the zero phonic or intensity level corresponds to the intensity of sound in a free plane wave when 1 microwatt of power flows through a square centimeter. It also corresponds to the intensity operating upon the ear drum when a pressure of approximately twenty bars is produced in the ear canal. As will be seen from the auditory sensation area chart, this zero level corresponds very nearly to the level which gives the maximum pitch range. Also, as will be evident from the data of Part One, Chapter III, it corresponds closely to the average speech intensity close to the mouth since the 10 microwatts of power flow through an area of about 10 square centimeters as it is radiated into the air.

For these reasons it seems logical to choose this level as a standard for comparing sound levels existing in any acoustic field and it is so used in this book. Using this terminology, the sensation level is the difference between the phonic levels of the tone at the given intensity and at the threshold intensity. The letter α will be used to designate the intensity level, α_0 being the particular value corresponding to the threshold. Then the sensation level α_s is given by

$$\alpha_s = \alpha - \alpha_0. \tag{4}$$

There is a definite relation between pitch as sensed by the ear and the frequency of vibration, so it seems reasonable to call the logarithmic scale of frequency a pitch scale; that is, the level of frequency is the pitch. In music, this level is determined by the position of the note upon the musical staff. For representing the quantitative relations in audition, then,

$$P = A \log f/f_0. \tag{5}$$

The most natural unit to use for measuring pitch is the octave which means that the base of logarithms to use is 2. Also if f is measured in kilocycles and the reference pitch is taken as 1 kilocycle, then

$$P \text{ (octaves)} = \log_2 f \text{ (kilocycles)} = 3.32 \log_{10} f. \tag{6}$$

The reference pitch chosen corresponds approximately to "high C" on the musical staff. The pitch steps on the chromatic scale are all equal to a semi-tone and correspond to a frequency ratio of $\frac{1}{12}$. For readily transferring from the musical notation to pitch numbers, the table below is given. As indicated these numbers are for international standard pitch and for the first octave above "high C."

Chromatic Scale

($a''' = 1760$ cycles per second)

Musical Notation	*Pitch Numbers*
c^3	0.066
c♯	0.149
d	0.232
d♯	0.315
e	0.398
f	0.482
f♯	0.566
g	0.649
g♯	0.732
a	0.816
a♯	0.899
b	0.982
c^4	1.066

For physicists pitch ($c^3 = 1024$) subtract .032 from each pitch number.

For each octave above the one represented add 1 and for each octave below subtract 1 from these numbers.

A pitch unit which is $\frac{1}{100}$ of an octave has been found convenient for practical use. Such a unit is logically called a centi-octave. It is approximately $\frac{1}{8}$ of a semi-tone.

To enable one to readily transfer from pitch to frequency or the reverse, a curve showing the relation expressed in equation (6) is shown in Fig. 80. For comparison purposes, the width of one full tone step on the major scale is shown on this chart. A curve showing the relation between intensity and intensity level is given in Fig. 81. Using these scales of pitch and intensity level, the boundary lines of the auditory

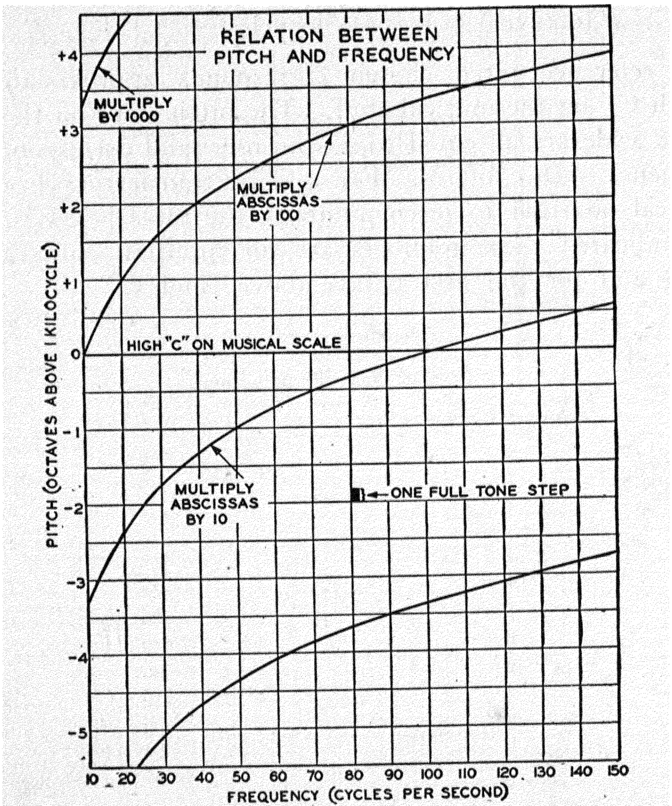


FIGURE 80.

sensation area have been replotted and are shown in Fig. 82. The lower curve gives the values of α_0 and the upper curve the values of α_m.

If the minimum and maximum audibility curves were plotted on an energy scale, the perceptible increment ΔE near the maximum audibility curve would be a million million times

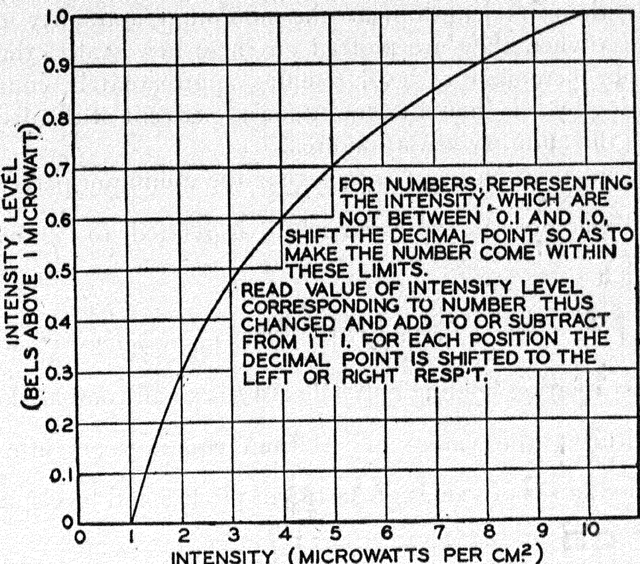


FIG. 81.—RELATION BETWEEN SOUND INTENSITY AND INTENSITY LEVEL.

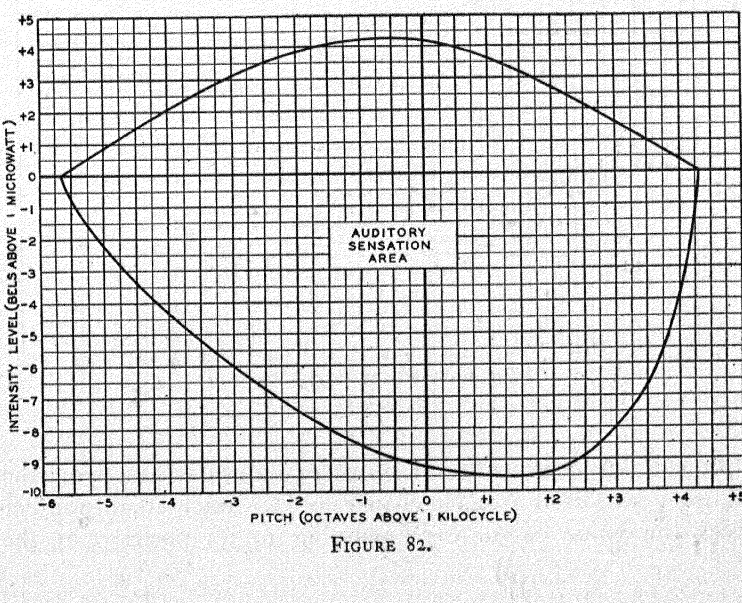


FIGURE 82.

larger than its length near the minimum audibility curve, whereas when they are plotted on these new scales, the perceptible increment in level remains approximately constant, changing by less than a factor 10 for 90 per cent of the distance across the auditory sensation area.

From equation (3) it is seen that the minimum perceptible level difference $\Delta\alpha$ expressed in db is related to $\frac{\Delta J}{J}$ by the equation

$$\Delta\alpha = 10 \log\left(1 + \frac{\Delta J}{J}\right) \doteq \frac{10}{\log_e 10}\frac{\Delta J}{J}, \tag{7}$$

the last relation holding only when ΔJ is small compared to J. Substituting the values of $\frac{\Delta J}{J}$ from equation (1) into this equation, a set of values of $\Delta\alpha$ for all pitches and levels can be

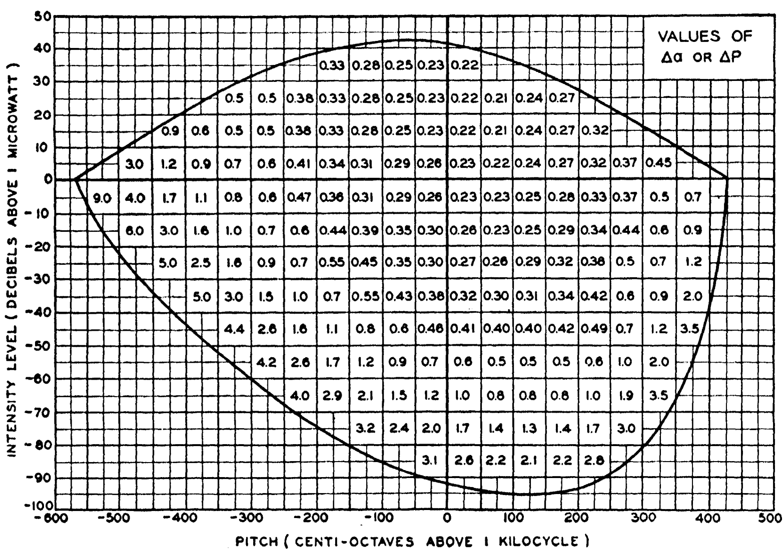


FIGURE 83.

obtained. A representation of these values is given on the auditory sensation area chart in Fig. 83. Each small number gives the value of $\Delta\alpha$ corresponding to its position in the auditory sensation area. It will be seen that the differential sensitivity for intensity varies from .21 to 9 db, depending upon the position in the auditory sensation area. An inspection of these numbers shows that one decibel represents about the change that is perceptible under average conditions, sometimes being a little more and sometimes a little less. For this reason it has proved to be a very convenient size. The fact that $\Delta\alpha$ is not the same at various positions shows that the Weber–Fechner law does not hold rigorously.

Similarly, ΔP expressed in centi-octaves is related to $\frac{\Delta f}{f}$ by the equation

$$\Delta P = 100 \log_2\left(1 + \frac{\Delta f}{f}\right) \doteq \frac{100}{\log_e 2}\frac{\Delta f}{f}, \tag{8}$$

the last relation holding only when Δf is small compared to f.

A comparison of the values of ΔP obtained by applying equation (8) to the Knudsen data with the values of $\Delta\alpha$ as shown in Fig. 83 shows that within the observational error and within the limited ranges of frequency and intensity explored, they are the same. Therefore, until more extended data are obtained for ΔP, it will be assumed that the figures shown on the chart of Fig. 83 can be taken as values of ΔP as well as values of $\Delta\alpha$. That this is so is not entirely a coincidence for the choice of decibels and centi-octaves for representing $\Delta\alpha$ and ΔP made this approximately true.

Number of Distinguishable Tones in Audition Range

The relations given above make it possible to calculate the number of pure tones which the ear can perceive as being different. For example, if starting at the minimum audibility curve, ordinate increments that are successively equal to the values of $\Delta\alpha$ at the sensation levels corresponding to the successive positions, are laid off along a constant pitch line, then the number of such increments between the curves for the threshold of audibility and of feeling is equal to the number of pure tones of constant pitch that can be perceived as being different in intensity. A rough estimate of the number of such tones can be made by an inspection of the chart shown in Fig. 83. More accurate values may be obtained as follows:

The number δN of distinguishable gradations in intensity for a small intensity level difference $\delta\alpha$ is given by

$$\delta N = \frac{\delta\alpha}{\Delta\alpha},$$

so that

$$N = \int_{\alpha_0}^{\alpha_m} \frac{1}{\Delta\alpha}\, d\alpha \tag{9}$$

which gives the number of distinguishable differences in intensity along any pitch line, α_0 and α_m being where the pitch line intersects the boundary curves of the auditory sensation area.

Using the relations given by equations (1) and (7), the value of this integral will be found to be

$$N = \frac{1}{nS_\infty} \log_e \frac{S_\infty 10^{\frac{(\alpha_m - \alpha_0)n}{10}} + S_t - S_\infty}{S_t}. \tag{10}$$

If the Weber–Fechner law held accurately, then S_t and S_∞ would be equal so that this expression would reduce to

$$N = \frac{\log_e 10}{10} \frac{\alpha_m - \alpha_0}{S_\infty} = \frac{\alpha_m - \alpha_0}{\overline{\Delta\alpha}} \tag{11}$$

as it should, for this states that the number of distinguishable tones is equal to the total number of db across the auditory sensation area divided by the number of db for one tone change.

Approximate values of N can be obtained directly from the chart of Fig. 83. The number of distinguishable tones in each line of 10 db change of level may be taken as ten times the reciprocal of the value of $\Delta\alpha$ corresponding to that line. Therefore, if the reciprocals of the numbers along any pitch line are added together, the sum will be one-tenth the total number of distinguishable tones.

Values of N, together with the values of the parameters necessary for its calculation, are shown in Table XVIII. It is interesting to note that for frequencies as low as 30 cycles the auditory sensations are very indistinct. It is seen that very little sense of loudness change is possible in this range. The hearing and feeling sensations are difficult to distinguish. There are more than ten times as many distinguishable changes at frequencies between 1000 and 2000 cycles as at a frequency of 60. In the last column of Table XVIII the average values of $\overline{\Delta\alpha}$ for the entire intensity range are given. They are obtained by dividing the numbers in the third column by the numbers in the seventh column.

TABLE XVIII

P	f	$\alpha_m - \alpha_0$	S_∞	S_t	n	N	$\overline{\Delta\alpha}$
− 500	31.25	30	.263	12.61	.448	3.11	9.65
− 400	62.5	65	.182	7.45	.415	34.2	1.90
− 300	125	91	.126	4.33	.386	93.8	.970
− 200	250	113	.087	2.70	.358	188.9	.598
− 100	500	127	.062	1.79	.318	299	.425
0	1,000	134	.051	1.37	.276	374	.358
+ 100	2,000	131	.053	1.07	.286	358	.366
+ 200	4,000	121	.074	1.75	.361	259	.467
+ 300	8,000	96	.128	2.86	.456	119	.807
+ 400	16,000	41	.240	5.18	.534	16.34	2.510

In a similar way the number of tones of different pitch can be obtained. The number of distinguishable differences δN in a small pitch interval δP is given by $\delta N = \frac{\delta P}{\Delta P}$. The total number of different gradations is then

$$N = \int_{P_1}^{P_2} \frac{1}{\Delta P} dP \tag{12}$$

where P_1 and P_2 are the pitch limits.

The number of perceptible differences in pitch N depends upon the path taken through the auditory sensation area.

There are three paths of particular interest, namely, (1) having a constant intensity level, (2) having a constant sensation level, (3) having a constant loudness level. Any of these can be obtained by graphical methods by plotting values of $\frac{1}{\Delta P}$ corresponding to each position in the auditory sensation area traversed and measuring the area under the resulting curve limited by P_1 and P_2. For example, in Fig. 84 are shown the paths for equal levels: the solid curved lines—sensation levels, the dotted lines—loudness levels, and the horizontal solid lines

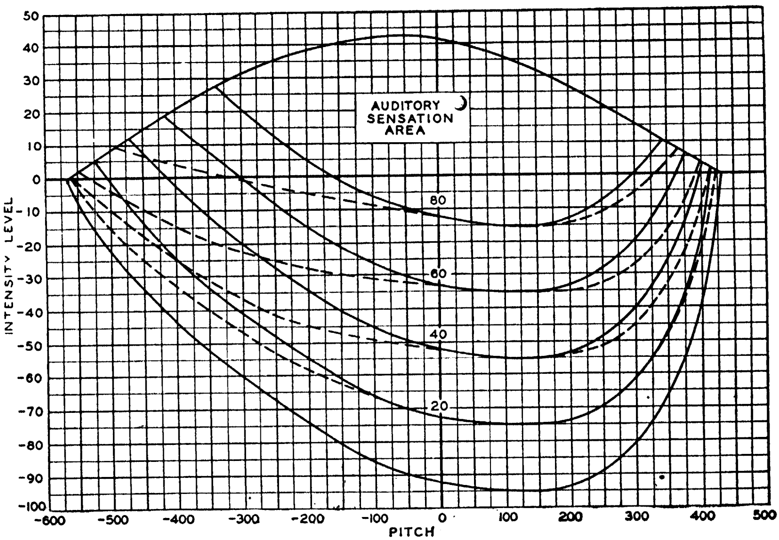


FIGURE 84.

—intensity levels. The dotted curves are taken from data to be discussed later.

The number of distinguishable changes in pitch as the tone changes along these lines is given in Table XIX. It is seen that in the process of changing the pitch of a pure tone from the highest to the lowest audible pitch, there are approximately 2000 perceptible gradations in pitch. The maximum number of perceptible gradations is obtained when the intensity level is kept constant at about zero level. As stated in Chapter I, there are about three times this number of rows of rods containing nerve endings in the basilar membrane, each row containing 4 or 5 rods terminating in 10 or 15 hair cells. (See Fig. 63.) If the tone is kept at a constant sensation level of 20 db while it is varied in pitch, there will be approximately 500 gradations in pitch. At lower sensation levels the number of gradations is still smaller. It is thus seen that at these low levels the position of maximum response must shift over more than ten rows of rods before the pitch change is perceived. This indicates that only a very few of the nerve endings in the stimulated area are activated at these low levels. As the intensity increases, more and more of them become activated.

TABLE XIX

Intensity Level		Sensation Level		Loudness Level	
α	N	α_s	N	L	N
− 60	590	20	520	20	460
− 40	1240	40	1270	40	1010
− 20	2050	60	1640	60	1690
0	2340	80	2180	80	2120

If an ordinate increment of $\Delta\alpha$ and an abscissa increment of ΔP be drawn in the auditory sensation area diagram, a small rectangle will be formed which may be considered as forming the boundary lines for a single pure tone. All tones which lie in this area sound alike to the ear. The number of such small rectangles in the auditory sensation area corresponds to the number of pure tones which can be perceived as being different. The number δN of such tones in a small rectangle $\delta P \cdot \delta\alpha$ is $\delta N = \dfrac{\delta\alpha \cdot \delta P}{\Delta\alpha\,\Delta P}$. The total number of such tones in the auditory sensation area is then given by

$$N = \int_{-565}^{+432} dP \int_{\alpha_0}^{\alpha_m} \frac{d\alpha}{\Delta P\,\Delta\alpha}. \tag{13}$$

The value of this integral was obtained by obvious graphical methods and was found to be approximately 540,000. One can obtain a number which is approximately correct for the number of distinguishable tones by the following simple procedure. Since $\Delta\alpha$ and ΔP are equal and given by the number on each square on Fig. 83, the number of tones in each square is 500 divided by $(\Delta\alpha)^2$ or by the square of the number in the center of each square. If the results for each square thus obtained are added together, the total will be the desired number of distinguishable tones.

One might well ask the question: How many complex sounds which are different can be sensed by the ear? At first thought, one might say that this number is represented by all the possible combinations of pure tones. Of course, such a number would be entirely too large, for some of these would sound alike to the ear, since the louder tones would necessarily mask the feebler ones. It is evident, however, that the number of such complex sounds will be very much larger than the number of pure tones.

Position Along the Basilar Membrane for Sensing Tones of Various Pitch

These data make it possible to calculate the position along the basilar membrane where the pure tones of different pitch are sensed. According to the theory of hearing discussed in Chapter I, a change in pitch corresponds to a change in the position of the maximum response of the basilar membrane. It seems reasonable to assume that the nerve terminals are uniformly distributed throughout the length of this membrane. Such an assumption is supported by the anatomical facts available. Consequently, it would be expected that for each perceptible step in pitch, the position of maximum response shifts the same amount. Since the calculation given above showed that there were approximately 1600 perceptible changes along the 60 db loudness level line, each step must correspond to .02 millimeter shift in the position of maximum response. The tone having the highest pitch, namely, 432, corresponding to 20,000 cycles, is sensed at the oval window and the one having the lowest pitch, namely, – 564, corresponding to 20 cycles, is sensed at the helicotrema.

Starting at the oval window the first step of .02 millimeter corresponds to a step of pitch from 432 to 430.5, since ΔP is 1.5 (see Fig. 83) in this region. In this way it will be found that the first 10 steps corresponding to a distance of one-fifth millimeter from the oval window end are concerned with pitches from 432 to 418 corresponding to frequencies from 20,000 cycles to 18,000 cycles. The next 10 steps go from 413 to 405; the next 10 from 405 to 394, and so on. In this way the position on the basilar membrane for each pitch can be obtained.

It is evident that this procedure is equivalent to the following mathematical process: Let l be the distance from the oval window end to the position of maximum response and l_0 the total length of the membrane and E the distance for each perceptible step. Then

$$l = E\int dN = l_0 \frac{\int_P^{430} \frac{dP}{\Delta P}}{\int_{-565}^{430} \frac{dP}{\Delta P}}. \qquad (14)$$

The denominator of this fraction is the number of perceptible steps and has been calculated above. The numerator can be obtained from the same graph that was used in calculating the denominator. In this way the curve shown in Fig. 85 was obtained. It was from this figure that the positions given in Fig. 66 were located.

It will be seen that a tone of zero pitch corresponding to a frequency of 1000 cycles is sensed in the middle of the membrane. The range used in sensing the speech sounds is from 7 to 28 millimeters from the oval window, which indicates that about two-thirds of the entire length is used for such purposes. When more accurate data for values of ΔP are available it will be interesting to compare the results calculated for these positions at different levels (sensation, intensity, or loudness).

If they are very different some interesting conclusions concerning pitch and loudness might result. This method of locating

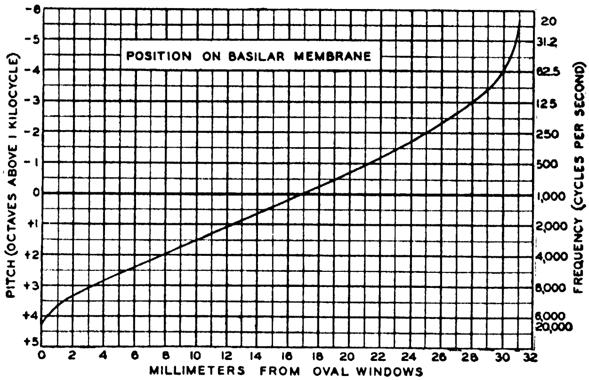


FIGURE 85.

the positions on the basilar membrane where tones of different pitch are sensed is due to Wegel and Lane.[1]

[1] Wegel, R. L., and Lane, C. E., "Auditory Masking and Dynamics of the Inner Ear," *Physical Review*, February, 1924.

CHAPTER IV

MASKING EFFECTS

IT is a common experience that when any sound is impressed upon the ear it reduces the ability of the ear to sense other sounds. If while a sound A is being impressed upon the ear, another sound B is gradually increased in intensity until the sound A can no longer be heard, the sound A is said to be masked by the sound B. When the ear is stimulated by a sound, particular nerve fibres terminating in the basilar membrane are caused to discharge their unit loads. Such nerve fibres then can no longer be used to carry any other message to the brain by being stimulated by another source of sound. Masking experiments appropriately chosen, then should enable us to determine what portions of the membrane are stimulated by any external sound.

A. A. Mayer [1] was one of the first to point out the experimental fact that low-pitched sounds had a masking effect different from that of high-pitched sounds. He stated that a tone of low pitch will completely mask one of higher pitch but that a tone of high pitch will not mask a tone of lower pitch. The apparatus which he used, however, made it very difficult to control the intensity or the purity of the tones used. On account of its importance, the problem of masking has been studied rather extensively at Bell Telephone Laboratories and the investigation is still being carried on.

Masking of Pure Tones by Pure Tones

The masking effect of one pure tone by another was determined by means of apparatus which was similar to that used in the determination of the acuity of hearing described in Chapter

[1] Mayer, A. A., *Phil. Mag.* 11, 500, 1876.

II. A damped telephone receiver was used for generating the pure tones. Connected to this receiver were two vacuum tube oscillators equipped with filters for eliminating any harmonics and with attenuators for supplying any magnitude of current. The attenuators were arranged so that by turning a dial the intensity level of the tone could be reduced very quickly from the maximum value to a value below the threshold. The intensity level for the threshold was determined both for the masked and the masking tones. The masking tone was then kept at a constant sensation level while the tones of other pitch were gradually increased in intensity until they were just perceptible in the presence of the masking tone. The level expressed in decibels that the masked tone was raised above its threshold value in the quiet is called the threshold shift.

The results of these measurements are shown in the curves of Fig. 86. The frequency of vibration of the masking tone is given by the number at the top of each chart and its sensation level by the number on each curve. The frequency of vibration of the masked tone is given by the abscissa and the threshold shift of the masked tone by the ordinate.

For example, in the fourth chart the masking effects of a tone having a frequency of 1200 cycles are shown. It is seen that the greatest masking effect is near 1200 cycles, which is the frequency of the masking tone. A tone of 1250 cycles must be raised to 46 db above the threshold to be perceived in the presence of a 1200-cycle tone which is 60 db above its threshold, or it must be raised to within 14 db of the masking tone before it is perceived. This corresponds to an intensity ratio between the tones of only 25. A tone of 3000 cycles, however, can be perceived in the presence of a 1200-cycle tone which is at 60 db when it is only 8 units above its threshold. This means that the intensity ratio between these two tones, under such circumstances, corresponds to 52 db or to a ratio of approximately 160,000 in intensity.

However, as the loudness of the masking tone is increased, all of the high tones must be increased to fairly large values before they can be heard. For example, the high frequencies

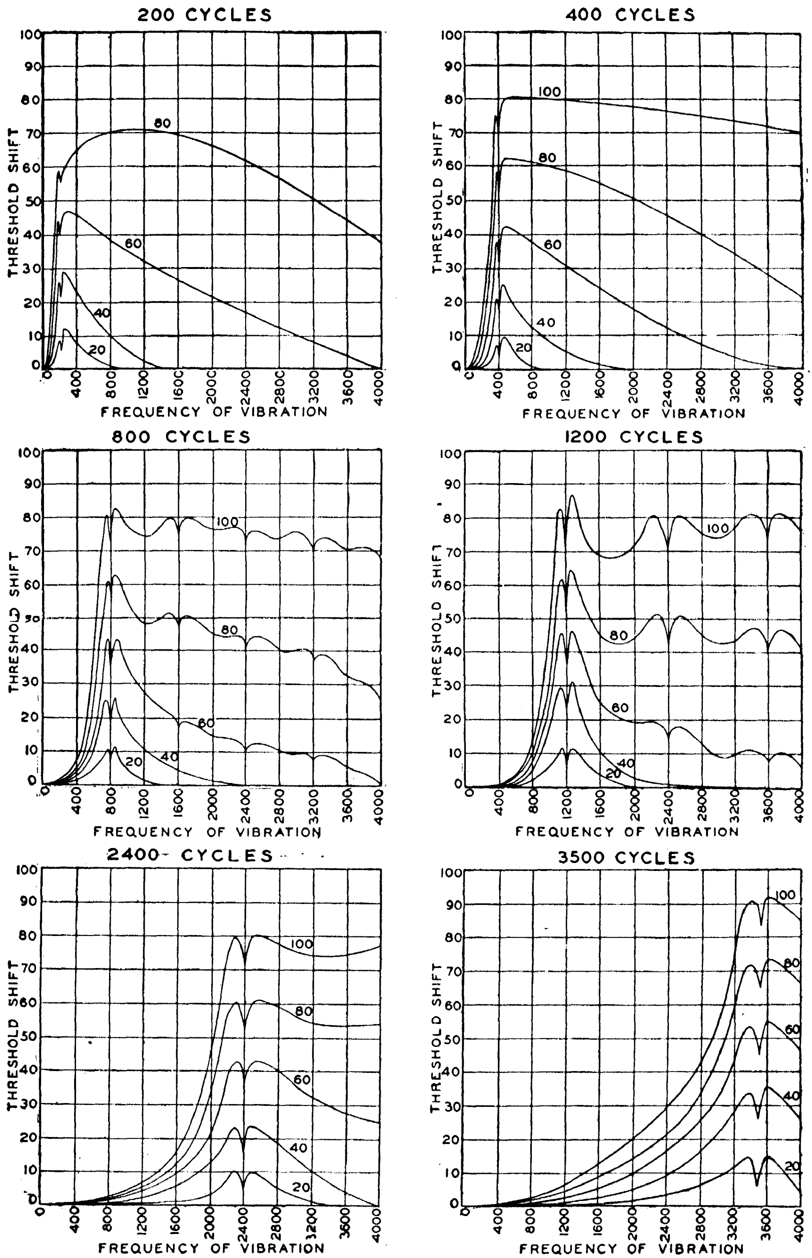

200 CYCLES
400 CYCLES
800 CYCLES
1200 CYCLES
2400 CYCLES
3500 CYCLES
THRESHOLD SHIFT
FREQUENCY OF VIBRATION

FIGURE 86.

must be raised 75 db above the threshold to be heard in the presence of a 1200-cycle tone having a sensation level of 100 db. But even for such large intensities for the masking tone, those frequencies below 300 are perceived by raising their loudness only slightly above the threshold value. It should be noticed that in all cases, those tones having frequencies near the masking frequency, whether they are higher or lower, are easily masked.

It is thus seen that Mayer's conclusion that a low-pitched sound completely obliterates higher pitched tones of considerable intensity and that higher pitched frequencies will never obliterate lower pitched tones is true only under certain circumstances. A low tone will not obliterate to any degree a high tone far removed in frequency, except when the former is raised to very high intensities. Also a tone of higher frequency can easily obliterate a tone of lower frequency if the frequencies of the two tones are near together. When the two tones are very close together in pitch, the presence of the masked tone is perceived by the beats it produces. This accounts for the sharp drop in the curves at these frequencies. A similar thing happens for those frequency regions corresponding to harmonics of the masking tone. In the charts for the 200- and 400-cycle masking tones these drops are not shown, inasmuch as they were small, but in an accurate picture they should be shown.

These results are plotted in a different way in Fig. 87. The abscissas represent the loudness of the masking tones, the frequency of which is indicated at the top of each of the charts. The amounts that the threshold of the masked tone is shifted are plotted as ordinates as in the previous figure.

For example, in the first chart the results are shown for a masking tone of 200 cycles. The curve marked 3000 indicates the masking effect of a 200-cycle tone upon a 3000-cycle tone. It is seen that the sensation level of the low-pitched tone can be raised to 55 db before it has any interfering effect upon the high-pitched tone. For higher levels than this it has a very marked effect.

It will be noticed that in nearly all of the charts the curves for different frequencies intersect. This leads to some rather interesting conclusions regarding the perception of a complex

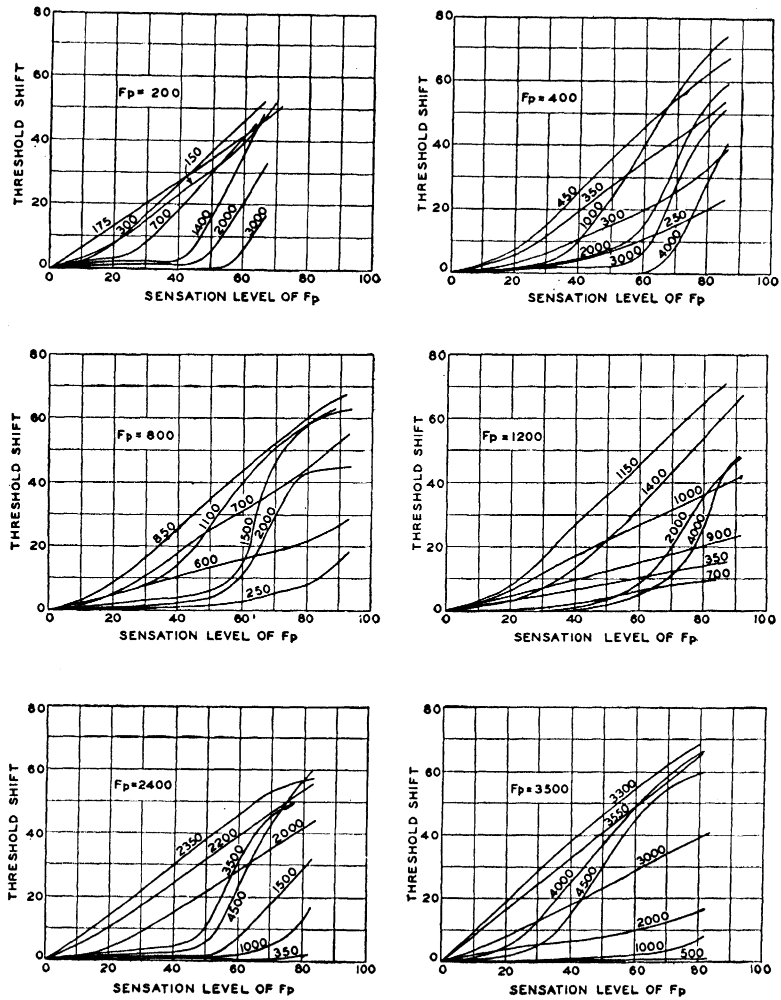


FIG. 87.—MONAURAL MASKING.

tone. If, for example, a complex tone had three frequencies of 400, 300, and 2000 cycles with levels of 50, 10, and 10, respectively, the ear would hear only the 400- and 2000-cycle tones as is evident from the masking curves for 400 cycles. It would be necessary to raise the 300-cycle tone an additional 6 db for it to be heard in the presence of 400 cycles at a sensation level of 50. However, if the complete sound were magnified 30 db without distortion so that the three tones had levels of 80, 40, and 40, respectively, then the 400- and 300-cycle tones only would be heard. Under such conditions, the 300-cycle tone could be attenuated approximately 8 db before it would disappear. These conclusions will be somewhat modified when all of the tones are sounding simultaneously, as the data were taken for two tones only, but the general picture given above will still be true. It follows that the sensation produced by a complex sound is different in character as well as in intensity when the sound is increased or decreased in intensity without distortion. In general, as the tone becomes more intense, the low tones will become more prominent because the high tones are masked. Due to the non-linearity of the ear transmitting mechanism, the low-pitched tones produce more subjective harmonics, harmonics in the sensed sound but not in the original pressure variations and for that reason increase in loudness faster than the high-pitched tones. It is a common experience of one working with complex sounds to have the low frequencies always gain in prominence as the sound is amplified. This phase of the subject will be discussed again in a later section.

The question naturally arises: "Does the same interfering effect exist when the two tones are introduced into opposite ears instead of both being introduced into the same ear?" The answer is "No." Curves showing the results of such tests are shown in Fig. 88.

For comparison the results for the case when the tones are both in the same ear are given by the light lines. Take the case of 1200 and 1300 cycles. It is rather remarkable that a tone in one ear can be raised to 60 db, that is, increased in intensity one million times, before the threshold value for the tone in the other ear is noticeably affected. If the 1300-cycle tone were introduced into the same ear as the 1200-cycle tone, its sensation level would need to be shifted 40 db, corresponding to a 10,000-fold magnification in intensity above its threshold intensity in the free ear before it could be heard. It is seen

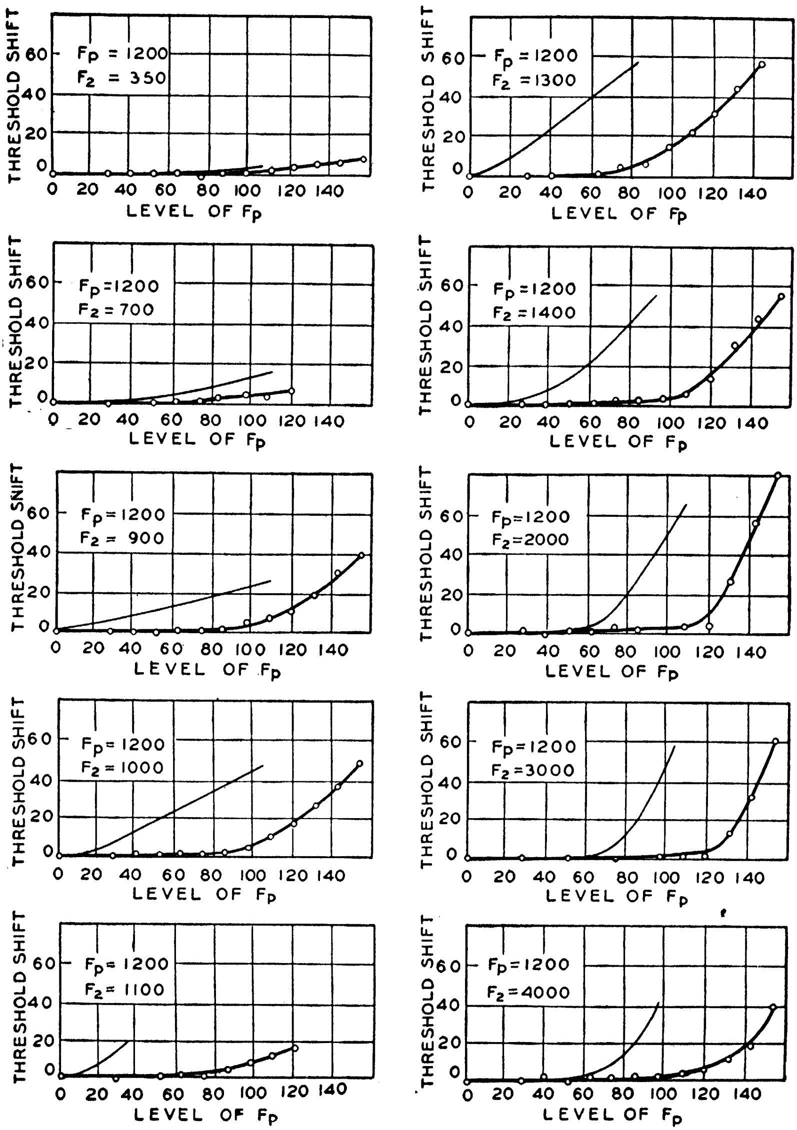


FIG. 88.—BINAURAL MASKING.

that if one set of curves is shifted about 50 db it will coincide with the second set. This strongly suggests that the interference in this case is due to the loud tone being transmitted by bone conduction through the head with sufficient energy to cause masking.

That this is the case is substantiated by experiments on persons having unilateral deafness. If the telephone receiver is held to the deaf ear of such a person and the intensity of the tone gradually increased, the threshold value will be reached when it has a sensation level of approximately 50 db. That the sound has been transmitted to the good ear by bone conduction is shown by the fact that under such circumstances the tone is greatly enhanced by placing the finger in the good ear. Another fact which is discussed in more detail later is that binaural beats are most pronounced when the intensity of the tones in the two ears has a sensation level difference of approximately 50 db.

These experiments indicate that when a receiver of the type used in these experiments is placed on the ear it communicates a certain amount of its vibration to the bones of the head. This vibration then is transmitted through the head to both ears, the intensity of the stimulation being approximately at a level of 50 db below that produced by the diaphragm of the receiver acting upon the air in the ear canal. It is claimed by some otologists that the vibration communicated to the cap of the telephone receiver used in practice is a great aid for a person having a certain type of deafness in transferring the speech vibrations to the end organ of hearing by means of bone conduction. It is seen from these experiments that in order for this to be true, the acuity of hearing by the air path must be at least reduced 50 db without in any way interfering with the acuity of hearing by the bone path. The value 50 db is entirely dependent upon the type of telephone receivers used. This value may be decreased or increased, depending upon the type of apparatus used in communicating the tones to the ear.

This suggests that interference of room noise to telephone conversation is not due principally to that which goes to the free ear, but to that which gets into the same ear to which the telephone receiver is being held, due mainly to leaks under the receiver cap. Even when the receiver is held very tightly to the ear, enough noise is conducted through the hard shell of the receiver and then to the air in the auditory canal to produce greater interference than that caused by noise coming into the free ear. This conclusion has been confirmed by direct experiments with telephone users.

Subjective Tones

The sharp dips in the curves of Fig. 86 at frequencies corresponding to multiples of the masking frequency require explanation. These dips suggest that they may be produced by harmonics of the masking tone. For example, the curves for the masking tone having a frequency of 1200 cycles and sensation levels of either 100, 80, or 60 db look like those which would be produced if the masking were caused by three tones having frequencies of 1200, 2400, and 3600. A careful analysis of the tone by means of the harmonic analyzer described in Part Two showed a frequency of only 1200 to be present. At the frequency corresponding to the dips, beats were plainly audible. This indicates that the masking tone creates harmonic frequencies in the ear.

As stated in Chapter I, these harmonic frequencies are due to the non-linear response of the hearing mechanism. The tones introduced by this non-linearity are called subjective tones. The magnitude and the frequency of such tones may be determined by using the principle of beats mentioned above. If while the masking tone is present, an exploring tone is changed in frequency and intensity until the beats are most prominent, then the intensity and frequency of such an exploring tone can be taken as the intensity and frequency of the subjective tone. It was found that there are three classes of subjective tones which are called harmonics, summation tones, and difference tones, respectively. When a single tone stimulates the ear, tones of the first class are produced. The frequencies of such tones are exact multiples of the frequencies of the stimulating tone. When two tones stimulate the ear, a series of subjective harmonics for each tone and also a series of difference and summation tones are produced. The subjective difference tones have frequencies which are equal to the differences obtained by subtracting the frequency of one tone from that of the other and also by subtracting the frequency of any harmonic from that of any other harmonic. Similarly, the summation subjective tones have frequencies which are obtained by taking sums instead of differences. When more than two tones stimulate the ear these three classes of tones are produced but the situation then becomes very complex. For example, it was found that the sensation levels of the first two subjective harmonic tones produced by a 1200-cycle tone having a level of 80, were 60 and 50 db, respectively. For a level of 60, the subjective harmonics were 20 and 15 db, respectively. For levels below 40, the subjective tones were undetectable.

The results of some experiments reported by Wegel and Lane are given in the chart of Fig. 89. One tone, called the primary tone, was held at a constant level while a second tone, called the secondary tone, was varied both in frequency and intensity. The resulting sensations are represented on the chart. For levels below the masking curves only the 1200-cycle tone can be perceived. For levels just above the masking curve and in the frequency range between 1200 and 2400 cycles, only the primary and the difference tones can be perceived. In other words, in this region, the presence of the secondary tone is detected by hearing the subjective difference tone between the primary and the secondary tones. For higher levels, the primary, the secondary, and the difference tones can be perceived. For example, if the secondary tone is held at a frequency of 1600 and at a level of 60 db, the ear will perceive three tones, namely, the 1200-, the 1600-, and the 400-cycle tones. For still higher levels for the secondary, very complicated mixtures of tones are perceived.

A careful analysis was made of the mixture of tones present in the ear when a primary tone of 1200 cycles at a sensation level of 80 db was present along with a secondary tone of frequency 700, and at the same level. The component frequencies were determined by introducing an exploring tone and determining the frequencies at which beats occur. If f_1 represents the primary, and f_2 the secondary, the frequencies found in the mixture were f_1, 1200 cycles; f_2, 700; $f_1 + f_2$, 1900; $f_1 - f_2$, 500; $2f_1$, 2400; $2f_2$, 1400; $3f_1$, 3600; $3f_2$, 2100; $2f_1 + f_2$, 3100; $2f_1 - f_2$, 1700; $2f_2 + f_1$, 2600; $2f_2 - f_1$, 200 (?); $4f_2$, 2800; $2f_1 + 2f_2$, 3800; $2f_1 - 2f_2$, 1000; $3f_1 + f_2$, 4300; $3f_1 - f_2$, 2900; $3f_2 + f_1$, 3300; $3f_2 - f_1$, 900. No attempt was made

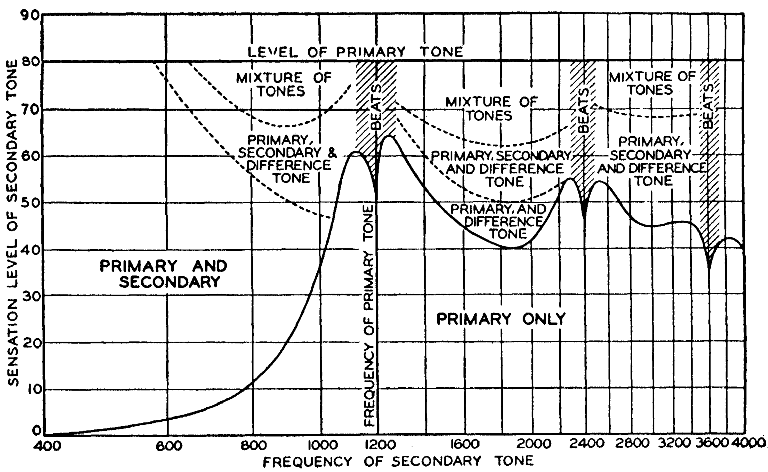


FIG. 89.—SENSATION LEVEL CAUSED BY TWO PURE TONES.

to determine their magnitudes, although approximate values can be obtained by measuring the intensity of the exploring tone at which the beats at each frequency are most prominent. Such measurements are rather difficult when so many tones are present. Except for the absence of frequency $4f_1$, this series is all that would be expected (see Appendix C) if the response of the ear were non-linear and represented by the equation,

$$x = a_0 + a_1p + a_2p^2 + a_3p^3 + a_4p^4. \tag{1}$$

In this equation x is the response of the mechanism of the middle ear; a_0, a_1, a_2, etc., are constants, and p is the pressure in the ear canal. While frequencies introduced by higher powers of the pressure were probably present, they were very faint and no careful search was made for them.

A study of the levels of a primary tone which is necessary to produce detectable subjective harmonics reveals some interesting and important data. To do this, the pure tone was held at a convenient level while the presence of the harmonic was determined by the beating effect produced by an exploring tone as described above. In this way the sensation level for tones of various pitches at which the second, the third, the fourth, and the fifth harmonic first appeared was determined. These data are shown in the curves of Fig. 90. For tones above zero pitch no subjective harmonics appear until the sensation

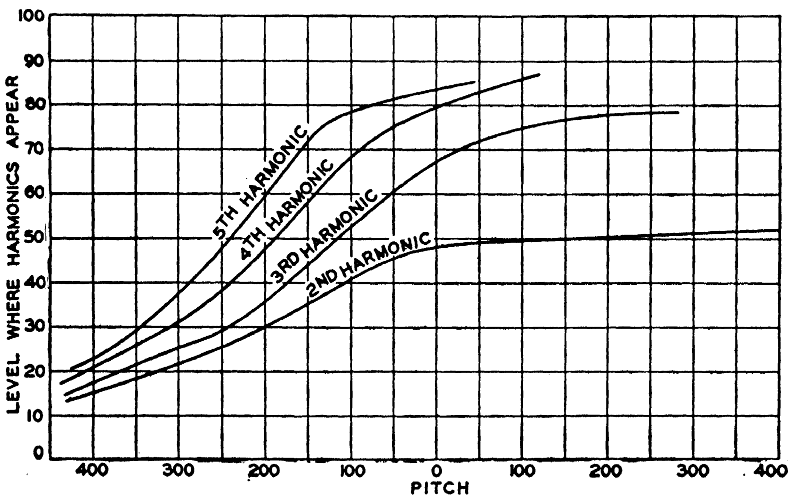


FIGURE 90.

level is 50 db, where the second harmonic just becomes detectable. In the low-pitched range the harmonics appear when the tone is very faint. For example, for a tone of pitch −4 octaves, which corresponds approximately to the frequency of the alternating current usually used in electrical lighting systems, the fifth harmonic appears before a sensation level of 25 is reached. The sensation levels usually used while listening to speech or music lie between 50 and 100 db, which shows that the frequency spectrum impressed upon the inner ear must be

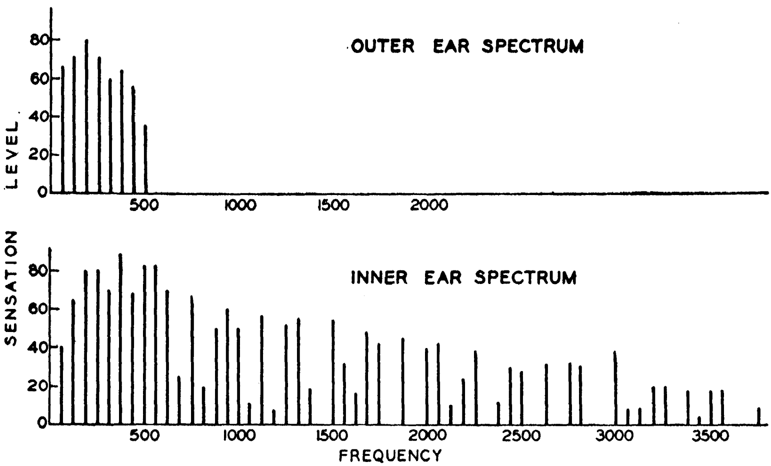


Fig. 91.—Loudness Spectrum. Dotted Lines Introduced by Overloading in Ear. (Loudness Equals 80 TU.)

very different from that impressed on the outer ear, the modifications being due to the non-linear transmission through the middle ear.

To illustrate this, an estimate was made of the inner ear spectrum which resulted from impressing an organ tone on the outer ear. For this purpose, the data given above on the non-linear properties of the ear were used. It was assumed that the summation and difference tones were of the same order of magnitude as the corresponding harmonics. The total amplitude of any one frequency component was obtained by taking the square root of the sum of the squares of all the contributions to this frequency component. In Fig. 91 the results of such a calculation are shown. The top chart gives the spectrum of the tone impressed upon the outer ear; the bottom one the estimated spectrum impressed upon the nerve terminals of the inner ear. It cannot be claimed that this is an exact representation of the inner ear spectrum, for the problem is too complicated for an exact solution, but it probably represents the main facts. A comparison of the two spectra in this figure illustrates the large differences which are produced in the process of transmission.

It is thus seen that the fact as noted by Mayer that low tones mask high tones much more readily than high tones mask low tones is due to the non-linear response of the hearing mechanism, which results in sending to the inner ear subjective harmonics that excite the nerve terminals which would otherwise be ready to receive the stimulation from the high tones. On the other hand, the high tones can only produce a stimulation in the regions near the maximum stimulation, that is, in regions where the higher tones are sensed. Therefore, little interference will be caused to the sensing of the low-pitched tones.

Calculation of the Form of Vibration of the Basilar Membrane

From the theory of hearing and the masking curves it is possible to find an approximate form of vibration of the basilar membrane. Let a primary tone of constant pitch and level be impressed upon the ear. Then to find the portions of the basilar membrane which are vibrating, an exploring tone is used. Those portions corresponding to pitches of the exploring tone where no threshold shifts are produced are not vibrating with amplitudes greater than those corresponding to the threshold. The other portions are vibrating with amplitudes which are measured approximately by the threshold shifts of the exploring tone. This statement requires modifications for those frequency regions close to the points where the primary tone or its harmonics are sensed. In these regions the presence of the secondary tone is recognized at a much lower level than would otherwise be the case. The amplitude caused by the exploring tone in these regions when its presence is just detectable is very much smaller than the amplitude caused by the primary tone. In these regions, however, a good estimate of the amplitude can be made from the shape of the rest of the curve representing the vibration form. The amplitude corresponding to the pitch of the primary tone is measured directly by the sensation level of the primary tone. Similarly, the amplitude of each of the subjective harmonics is measured by the sensation level of the exploring tone when the best beats are produced with the subjective tone.

For convenience in speaking of the amplitude of vibration of the basilar membrane, a new term, nerve sensation level, will be defined. At each position along the basilar membrane it has been seen that there corresponds a tone of definite pitch. When a sound is being sensed by the ear, the nerve sensation level corresponding to each position on the basilar membrane is the sensation level of the pure tone corresponding to that position which would produce the same amplitude.

Using an exploring tone in the manner described above, F. H. Graham of Bell Telephone Laboratories made a careful study of the nerve sensation levels [1] produced by an 800-cycle tone with the result shown in Fig. 92. Separate curves are shown for several different intensities. The peaks in the curve occur at positions corresponding to frequencies which are multiples of 800 cycles. The points marked (*o*) were obtained by the best beat method; the points marked (·) were obtained by noting the threshold shift. The scale at the bottom gives the position on the basilar membrane and the scale at the top denotes the frequency of vibration of the exploring tone.

In a similar way the nerve sensation level produced by three

[1] A nerve sensation level of 80 at 1200 cycles corresponds to an intensity level of -12. Consequently, the power in a free wave corresponding to this level is $10^{-1.2}$ microwatts. The amplitude of vibration in water corresponding to this is 3×10^{-7} centimeters which is about 10 times the diameter of the molecule. The amplitude of the basilar membrane for producing this sensation level is probably not more than 100 times this value. Consequently at the threshold the displacement is only a small fraction of the diameter of the molecule.

tones having frequencies of 1000 cycles, 1500 cycles, and 2000 cycles, respectively, at sensation levels of 80 db, were obtained, the results being shown in Fig. 93. The common difference tone produces the masking in the region corresponding to

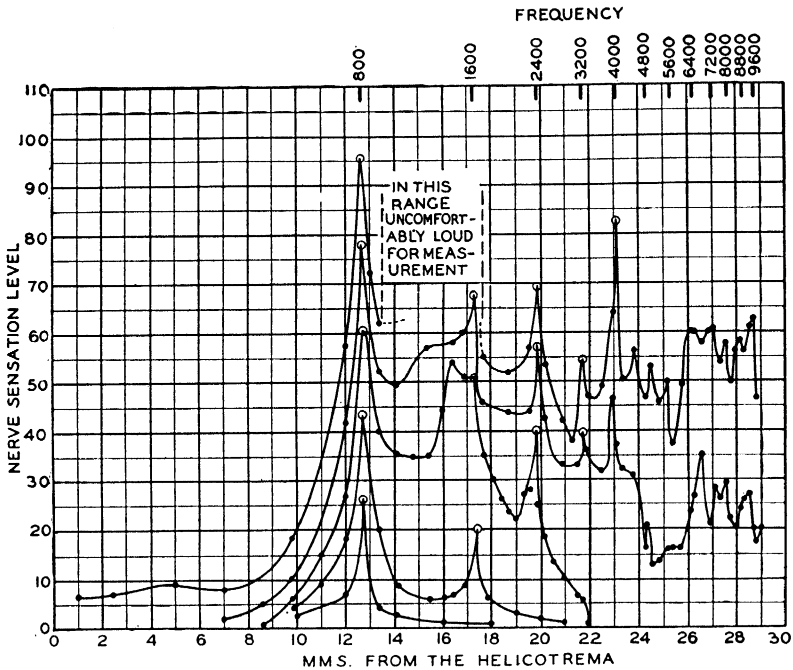


FIG. 92.—VIBRATION FORM OF BASILAR MEMBRANE PRODUCED BY A PURE TONE OF 800 CYCLES.

500 cycles. The subjective difference tone 2000 – 1000 reinforces the 1000 cycle tone. The summation tone 2000 + 1000, the second harmonic of 1500, and the third harmonic of 1000 units to give the subjective tone 3000 which was definitely measured as shown in the figure.

It is seen that except for frequency regions near those corresponding to the masking tone and its harmonics, the masking curves become the form of the vibration of the basilar membrane by a simple transformation of variables used in plotting, the frequency scale being converted into a distance scale along the basilar membrane, and the threshold shift scale into an amplitude scale. By keeping this in mind, the masking curves as shown in Fig. 86 will convey a good notion of the form of the vibration of the basilar membrane when excited by the pure tones indicated.

Figure 94 shows the form of the vibration of the basilar membrane when the intensity level of the exciting primary tone is – 32 db, corresponding to a pressure of one-half bar in the external ear canal. The dotted curve is the locus of the maxima for all frequencies which are impressed upon the ear at this intensity level. As pointed out in Wegel and Lane's paper,[1] the curves become less sharp as the frequency

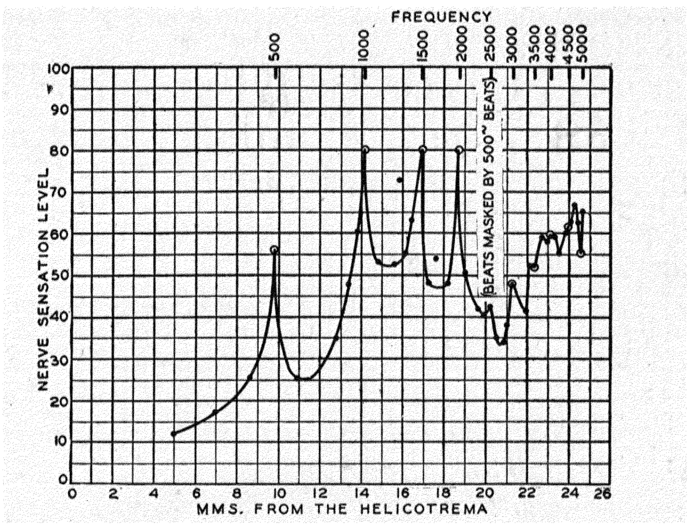


FIG. 93.—VIBRATION FORM OF BASILAR MEMBRANE PRODUCED BY THE THREE PURE TONES—1000 CYCLES, 1500 CYCLES, 2000 CYCLES.

is decreased. This is in agreement with what would be expected from the dynamical structure of the cochlea. At very low frequencies the stimulus may be conceived as due to a more or less bodily motion of the tectorial membrane along the basilar membrane.

[1] "Auditory Masking and Dynamics of the Inner Ear," *Physical Review*, February, 1924.

It is to be expected that similar curves at extremely high frequencies should become less definite, probably not by becoming flatter, but by having their maxima at or beyond the proximal end of the organ of Corti. This conclusion is arrived at principally from a consideration of the curves of absolute sensitivity of normal ears in which the sensitivity is seen to drop off very sharply at about 15,000 cycles.[1] This sort of an assumption is further substantiated by the fact that when plotted as displacement of the basilar membrane, sensitivity curves of abnormal ears in which the lesion can be reasonably well traced to degeneration of the nerves of the proximal end

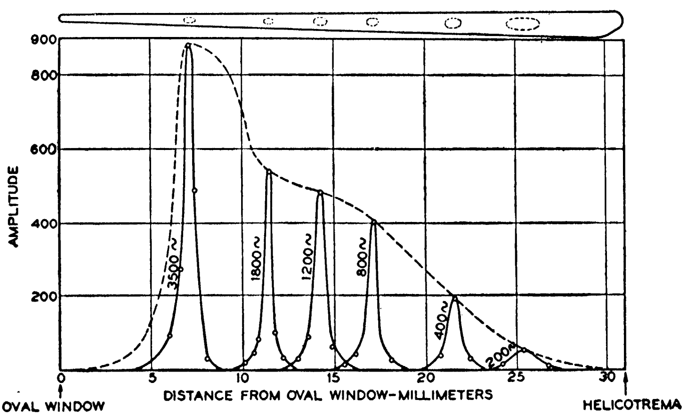


FIG. 94.—AMPLITUDE ALONG BASILAR MEMBRANE FOR DIFFERENT FREQUENCIES (R.M.S. PRESSURE .5 DYNES.)

of the basilar membrane, also indicate similar sharp cut-offs at frequencies much lower than 15,000 cycles, whereas no such abrupt cut-offs have yet been recorded at the low end of the frequency scale.[2] According to this theory of the action of the cochlea, it follows that as long as there is sensitivity in any of the nerves, even if it is only in a small region, the ear will be able to detect a tone of any frequency if it is sufficiently

[1] For original data see C. E. Lane, *Physical Review*, 19, 492, May, 1922.

[2] For example, see paper by Dr. E. P. Fowler and R. L. Wegel, "Audiometric Methods and Their Applications," *Transactions of the American Laryngological, Rhinological and Otological Society, Inc.*, 1922.

intense unless the necessary intensity is greater than the person can endure.

A plan view of the basilar membrane is shown drawn to scale at the top of the figure. Conjectured contour lines are drawn enclosing areas over which the amplitude is more than one-half that of their centers. The lengths of these areas are obtained from the curves shown in the figure and their widths by taking one-half the width of the membrane.

Masking Effects of Complex Sounds

The masking effects of complex sounds are what one would expect after considering the data on pure tones. When the masking sound is composed of two components, the masking is a combination of the masking produced by each component plus the effects due to the summation and difference subjective tones. In the previous section such curves were described for a complex tone with components having frequencies of 1000, 1500, and 2000 cycles and having the common sensation level of 80 db.

Masking curves for a musical sound having its components 60 cycles apart are shown in Figs. 95, 96, 97 and 98. The pressure exerted on the ear drum by each component is given by the lower curve on each chart. Figure 95 shows the effect when the components of appreciable size are above 3000 cycles; Fig. 96 when they are above 1500 cycles; Fig. 97 when they are below 1500 cycles; and Fig. 98 when they are below 500 cycles. No attempt was made to explore the detail of the curve near each component.

As stated in the chapter on "Noise," the threshold shift or masking curve of a complex sound may be taken as a measure of the annoying effect of the noise. It shows the reduction of the capacity of the ear to sense sounds in the presence of such a noise. Experimental tests have shown that the threshold shift of the speech sounds produced by noise is approximately the same as the average shifts produced on a 500-, a 1000-, and a 2000-cycle tone as shown in the noise audiogram.

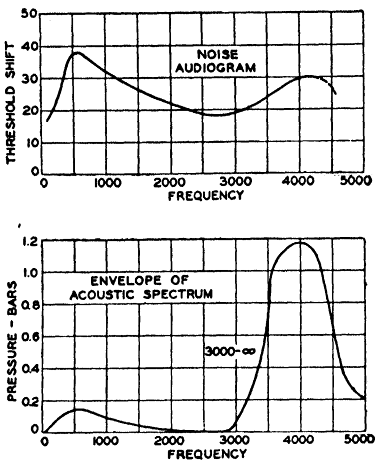


FIGURE 95.

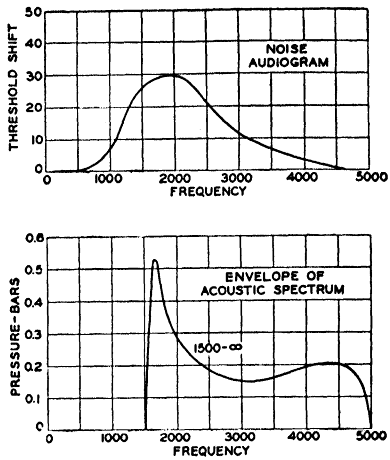


FIGURE 96.

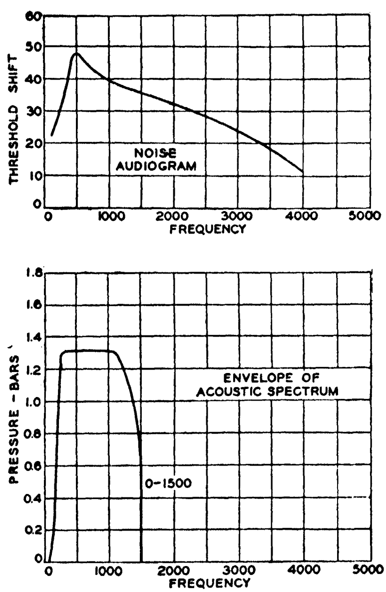


FIGURE 97.

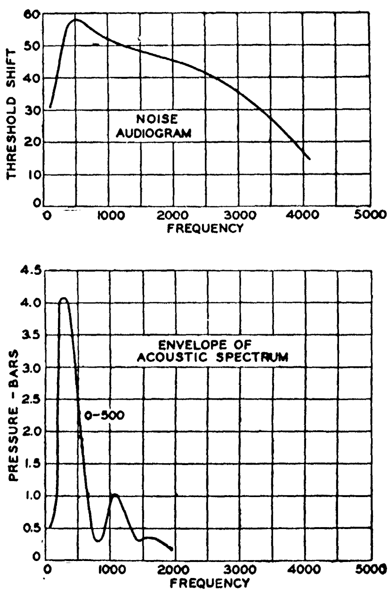


FIGURE 98.

The masking effect of noise in everyday life is very great. Table XX gives an idea of the noise effects produced in various familiar places. The amount of noise is measured by the threshold shift produced on the 3–A audiometer tone. Such shifts are approximately the same as would be produced for speech. These figures show how impossible it is to reach an audience of any size with an ordinary speaking voice unless the noise is kept at a low level.

TABLE XX

Amount of Noise Threshold Shift	Maximum Distances for Hearing Average Speech	Typical Place for Such Noise
0	1250 feet	Soundproof booth
10	395 feet	Country residence
20	125 feet	Quiet office
30	40 feet	Average office
40	12 feet	Noisy office or department store
50	4 feet	Railway train or automobile
60	15 inches	New York subway
70	5 inches	
80	1.5 inches	Boiler factory

CHAPTER V

Binaural Beats

When listening to a sound produced in the air normally both ears are used and certain sensations are dependent upon this fact. Changing the normal condition for such listening produces certain effects known as binaural effects. Some of these effects have already been mentioned because they were directly related to the material being discussed. These and other effects are described in this chapter and their bearing on the theory of hearing discussed.

Pure binaural effects are probably produced in the brain itself and on account of the difficulty of interpreting the sensations produced, there have been very large differences in results obtained by different observers. In general, it has been found that the two ears aid in locating the direction and distance of a sound source. This is particularly tıue of sources giving pure tones, for it has been found that persons having a total loss of hearing in one ear have great difficulty in locating the direction of such tones. Even when using both ears, however, it is very difficult to locate the source of pure tones when the pitch is high.

The Effect of Intensity and Phase upon the Apparent Location of Sound Images

If a sound is conducted from a source to the two ears by different acoustic paths, some very interesting effects are produced. If the apparatus is arranged so that the phase in one of the paths can be changed, the following sensations are experienced by most persons who try the experiment. When the difference in phase between the two paths conducting the sound to the ear is zero, then the source appears to be in the median plane, that is, directly in front of the individual. As the difference in phase increases, the sound image travels, apparently, along a circle toward the ear which is leading in phase. When the image appears to be opposite the ear, it suddenly seems to jump through the head to the other side, finally coming back again to the median plane.

Several methods of producing the phase difference have been used by various investigators. The results obtained by G. W. Stewart and his students seem to be more consistent than those obtained by any of the other observers. For a source of sound he used an instrument called the "Phaser."[1] This instrument consisted essentially of a rotating toothed wheel with two telephone bipolar receiver magnets placed radially close to the teeth and capable of being separated from each other by a variable known number of degrees of rotation of the wheel. The currents induced in the windings of these small magnets were conducted to two telephone receivers which were placed on the ears of the observer. It was thus possible to secure any difference in phase between the sounds at the two ears by moving the bipolar receiver magnets to different positions around the tooth wheel. By including selective networks in the circuit, tones of a fair degree of purity were obtained. The observer sat at the center of a circular scale and was asked to point in the direction from which the sound appeared to come.

A second method made use of two tuning forks placed at the end of rubber tubes, the sound being conducted to the ear by means of stethoscopes connected to these tubes. The tuning forks were slightly out of tune so that the difference in phase would change periodically. Under such circumstances, the image appeared to rotate around the head. By recording the time necessary for the image to rotate through a given angle, and comparing it with the time to produce a complete cycle, results were obtained which showed the relation

[1] *Physical Review*, May, 1920, p. 433.

between the phase shift at the ears and the angular displacement of the sound image.

In a third method the source of sound was two tuning forks which were electrically driven with the same frequency. The sound was conducted to the ear by means of stethoscope tubes as described above. Means were provided for changing the attenuation in the acoustic paths so as to keep the intensity at the ear constant while varying the phase. The phase difference was produced by choosing different lengths of the air path from the tuning forks to the ears.

In all these methods it was found that within the experimental error the angular displacement of the image was proportional to the difference in phase at the two ears, the constant of proportionality varying somewhat with frequency. The following formula represents the average results obtained by such experiments:

$$\frac{\Phi}{\theta} = 0.0034f + .8 \text{ (approx.)}$$

where Φ is the phase difference at the two ears, θ the angular displacement of the sound image from the median plane, and f the frequency of vibration. For example, the phase difference of a pure tone having a frequency of 500 cycles is always about 2.6 times the angular displacement of the sound image. When the phase difference has reached 180°, the image has been displaced 70°. As the phase difference increases, the image jumps to − 70° and then again returns to the median plane when the phase difference reaches 360°. For a pure tone of 60 cycles the phase difference is approximately equal to the angular displacement of the image.

If the phase of sound reaching the two ears is kept constant while the intensity is varied, an angular displacement in the sound image is also produced. However, the results are very much less definite and the variations from one individual to another are very great. Stewart concluded from his results that for any one individual the angular displacement of the image was proportional to the difference in intensity level at the two ears. However, his results show that for some individuals the angular displacement for a given difference in intensity level was twice as great as for another. As the frequency of the exciting tone approaches about 1000 cycles, the uncertainty of locating the image becomes greater, and for many persons there is no localization for the higher frequencies.

According to the theory outlined in Chapter I, the phase difference produced at the two ears is preserved in the composite nerve current going to the brain. The discharges from any particular nerve fibre occur at intervals which are exact multiples of the period of the sound wave. As the intensity of the sound becomes greater, this interval becomes less and approaches the period of the sound wave. The effect of all of the impulses from the individual nerve fibres is to produce a stimulation pattern in the brain which has a periodicity of the sound wave. The maximum stimulation at the brain centre is definitely related to the maximum pressure in the sound wave in front of the ear drum. The interval of occurrence between the two depends upon the time of transmission of the mechanical vibration from the drum of the ear through the middle ear and through the cochlea to the basilar membrane, and also the time of transmission of the nerve impulses from the nerve endings on the basilar membrane to the brain. The two maxima produced in the brain from the stimulation coming from the two ears will occur at approximately the same time if the phase of the sound vibration at the two ears is the same. However, when this phase is different there will be a corresponding time interval between the occurrence of the two maxima produced in the brain. It is undoubtedly the recognition of this time interval that enables us to recognize phase difference.

As pointed out by Hartley and Fry,[1] when the experiments are performed as indicated above, a new experience is produced which is different from that produced by a source of sound being actually present. If a sound source is placed in the air and moved about, the change in phase and intensity produced at the two ears must take place together in a certain way so

[1] *Physical Review*, December, 1921, p. 431.

that any experiments which produce sounds which keep either the phase or intensity constant, produce an experience different from that ordinarily obtained when listening to actual sound sources. The fact that the experiments on the change in the apparent image with change in phase give much more definite results than those obtained when the phase is kept constant and the intensity varied shows that the mind can more readily reconcile differences in intensity than differences in phase. Hartley and Fry give a series of calculated curves showing how the phase and intensity vary as a sound source is moved about the head. It is only when reproduced sounds have phases and intensities corresponding to points on these curves that localization as perfect as that obtained from an actual source would be expected.

Binaural Location of Complex Sounds

During the war the binaural location of complex sounds became very important. Its use made it possible to locate enemy submarines and aeroplanes. It was found that when two transmitters connected separately to two receivers were used in picking up the sound and transmitting it to the two ears, the direction of a complex source of sound could be located by the individual listening. To obtain a complete duplication of the binaural effect produced without such a transmission system, the two transmitters must be mounted on something equivalent acoustically to the head and at positions corresponding to the two ears. Then if the transmission system transmits faithfully the phases and amplitudes of the component sounds to the ears, the same auditory sensation will be produced as that obtained when the head is placed in the position where the artificial head carrying the transmitters is located. Any variation from this ideal transmission system will result in producing results which are different from those ordinarily produced by direct listening.

Experiments have shown that a considerable departure from this ideal may be made and yet a fairly good sense of localization be obtained. As was the case with pure tones, so with complex tones it was found that the phase was the controlling factor. For this reason, phase compensators were introduced into such a binaural transmission system so as to bring the apparent location of the sound directly in front of the observer. The amount of compensation necessary to do this indicated the position of the complex sound. The very fact that such compensators will not produce the proper shifts for all the components, indicates that the localization of such a compensated transmission system will be somewhat indefinite. When the source of sound is a combination of a few pure tones the compensated transmission system might very well produce a shifting of these components to different apparent positions with the result that either images in several directions are formed or confusion produced so that no localization is obtained. The former effect has been observed and described by Bowlker.

The location of a complex sound under ordinary conditions is very definite and may be made even by a person who is totally deaf in one ear. It is difficult to point out definitely all the factors which contribute to this ability. Certainly the reflections in the room under ordinary circumstances give considerable aid. In any event, the phases and amplitudes of the components must change in a prescribed way. Our experience with such sounds has given us an education so that we unconsciously know the way these changes take place as the source is moved into different positions. According to this view, it seems reasonable to expect that new complex sounds would be very much more difficult to locate than those with which we are ordinarily familiar.

Binaural Beats

If by means of two telephone receivers connected to vacuum tube oscillators tones of slightly different pitch are introduced into each ear, the sensations described below will be produced. Let us assume that one tone, called the "primary," is kept at a sensation level of 80 db in the right ear. When the sensation level of a second tone, called the "secondary," is at about 10 db in the left ear, then faint beats are observed. As the sensation level of the secondary is increased these beats become more pronounced, reaching a maximum at a level of 30 db. For higher levels the beats again become fainter until a level of 45 db is obtained, above which they are not heard. These beats, called "objective" beats, produce sensations which correspond in every way to those produced when the primary tone is reduced in sensation units about 50 db and introduced directly into the same ear as the secondary. These objective beats are undoubtedly due to physical interference in the left ear, the vibrations coming from the right ear by means of bone conduction. This conclusion is confirmed by the fact that the difference in intensity level for best beats increases when soft rubber pieces are placed under the receiver caps. It is important to remark here that this phenomenon is one which can easily be observed by anyone who tries it and beats are produced for all frequencies.

Keeping the primary again at a sensation level of about 80 db in the right ear it will be found that for a sensation level of 60 db for the secondary tone in the left ear beats again appear having a maximum effect when the tones in both ears have the same sensation level. These beats called "subjective" beats disappear when the secondary tone is at a level of 100. These subjective beats are entirely different in character from those called objective beats. Some persons cannot hear them at all, and others report results which are quite discordant. It is with these subjective beats that most of the experiments on binaural beats during the past century have been concerned. For this reason it is not surprising that there was so much disagreement between results reported. In a good many of these experiments, tuning forks or other sources of sound were used which made it difficult to prevent the sound from one source going into both ears and thus directly producing beats of the ordinary kind.

The work of Stewart [1] showed that seventeen out of the

[1] *Physical Review*, June, 1917, p. 502.

twenty-three were able to hear beats sufficiently distinct to report them. The work of Lane[1] showed that eighteen out of twenty-two were able to hear the beats sufficiently well to determine their period. The kinds of sensation reported by the different observers were greatly different.

A summary of the essential facts concerning subjective beats as given by Lane is as follows:

"(*a*) If two tones of equal intensities and nearly the same frequencies are simultaneously presented to opposite ears, the beat frequency can be recognized by about 80 per cent of the observers, provided the frequencies of the beating tones are less than 800 or 1000 cycles. For higher frequencies the beats cannot be heard.

"(*b*) If the beats are slow, the one outstanding phenomenon observed by all who recognized the beat is an alternate left and right localization of the sound, localization being on the side of the tone leading in phase.

"(*c*) Most observers who hear the slow beats experience a more or less vague notion of the localization travelling along some path through the median plane when the localization shifts from one side of the head to the other, but there is no good agreement among the observers as to the position of this path.

"(*d*) The passing of the localization through the median plane is generally more clearly defined during phase agreement than during phase opposition.

"(*e*) While all observers who heard the slow beats reported without any previous suggestions the existence of the alternate right and left localization, none reported any intensity maxima until questioned as to the existence of such maxima. However, when questioned, over 80 per cent of the observers reported maxima corresponding to one or more of the following three-phase relations: (1) phase agreement, (2) 30° or 40° before opposition and (3) 30° or 40° after opposition. There was no good agreement among the observers and several during the course of the experiments shifted from one phase relation to another in their report on the time of intensity maxima.

"(*f*) For fast beats the chief sensation is that of an intensity fluctuation of the sound located somewhere within the head.

"(*g*) For intermediate beats some reported a predominating sensation of motion, others of intensity fluctuation and still others seemed to experience both sensations about equally well and could direct their attention upon either at the sacrifice of the other.

[1] *Physical Review*, September, 1925.

"(*h*) Subjective beats are heard equally well for tones introduced into the ears by means of telephone receivers with and without receiver cushions or presented by means of rubber tubes.

"(*i*) So long as the two tones are of equal intensity, the hearing of subjective beats may be heard about equally well for all intensity levels of the two tones."

It is important to note that while the objective beats are readily observed by all, the phenomenon of subjective beats is quite indefinite and differs very greatly with individuals; also, that no subjective binaural beats are obtained for frequencies higher than 700 or 800 cycles.

It seems clear that the phenomenon of subjective beats is one which is produced in the brain and is closely associated with the binaural localization described in the last section. The term "beat" is hardly descriptive of the phenomenon since the sensation obtained is that of a wandering localization. It is only occasionally that we notice the intensity maxima and they are usually associated with positions of best localizations rather than positions of maximum loudness. Also, since the intensity and phase relations do not correspond to those orinarily experienced in locating a source of sound, the psychological reaction of the observers must be that of experiencing a new sensation.

Other Binaural Phenomena

The following experiment suggested by Dr. H. D. Arnold showed some very interesting effects. A high quality transmission system was provided with a filter system so that all the frequency components below 1000 cycles were sent into one channel and delivered to the left ear. Those above 1000 cycles were sent into another channel and delivered to the right ear. When speech was transmitted over such a system there was apparently no distortion produced, although if either one or the other of the two receivers were taken away the speech sound was very distorted and it was hard to recognize what was being said. When both receivers were used, the speech seemed to be good quality and no difficulty was experienced in following what was being said. Apparently, in this case, the brain was able to combine the sounds obtained from the two ears to complete the proper picture. However, when music was transmitted, a different situation resulted. This was particularly true when listening to music from the piano. In this case the tones appear first in one ear and then in the other ear depending upon the pitch. This causes confusion and gives a very weird sort of sensation. When listening to sounds which have frequencies fairly well scattered in both the ranges below and above 1000 cycles, the sensation produced was about the same as that obtained by combining the frequencies into the same ear. When the sounds were predominating in either one or the other ear, localizations were produced first at one ear and then at the other as described above.

CHAPTER VI

Methods of Testing the Acuity of Hearing

The kinds of hearing tests that are needed may be classified into four groups according to their purpose as follows:

1. Industrial, or those made to determine the fitness of a candidate for employment. In certain types of work it is particularly important that a prospective employee meet a definite requirement for acuity of hearing. Tests made in the army and the navy for various branches of service are conspicuous examples of this kind of test.
2. Educational, or those made to determine the degree of hearing of school children in both public schools and schools for the deaf to determine the proper educational methods.
3. Clinical, or those made to assist the physician to make a proper diagnosis of the cause of deafness.
4. Research, or those made to determine new facts about both normal and abnormal hearing.

Different methods are appropriate for these different purposes. Unless the particular purpose is kept in mind, there is apt to be confusion when discussing the merits of a method of testing hearing.

It is highly desirable that a single scale, independent of the method of testing and of general application to all the purposes mentioned above, be used for representing the degree of hearing. Therefore, before describing the apparatus developed for these various purposes, it is well to describe such a scale of hearing which is coming into use by otologists and which may be used for all of these purposes. It will then be shown how the results obtained from the commonly made voice, watch tick, acoumeter, coin click, and tuning fork tests can be expressed in hearing loss units on this scale.

The sensation units used in describing the intensity levels of sounds are convenient and logical for defining the degree of deafness. The hearing loss is measured by the threshold shift from the average threshold intensity level for normal ears. In other words, the sensation level of a tone which can just be heard by the person being tested is the hearing loss (H.L.) expressed in sensation units. Expressed mathematically,

$$\text{H.L.} = 10 \log \frac{I}{I_0} \tag{1}$$

where I is the necessary intensity of the sound for hearing by the person being tested and I_0 the normal threshold intensity. This equation gives the hearing loss for any sound—pure tones, musical sounds, watch tick, voice, tuning forks, etc.

In a paper by Fowler and Wegel,[1] a hearing scale was proposed which has been objected to by some otologists because it is dependent upon the threshold of feeling as well as the threshold of hearing. On this scale the per cent hearing loss is the hearing loss in sensation units divided by the number of sensation units between the threshold of hearing and the threshold of feeling for an average normal ear. It is undoubtedly the best answer to the practical question as to what is the per cent hearing loss, and is very useful in expressing general results.

It is sometimes convenient to give a figure which represents the average per cent loss of hearing. It seems reasonable and logical to choose for this the fractional part of the normal auditory sensation area corresponding to tones which cannot be properly sensed by the person being tested. This is approxinately [2] equivalent to the ratio of the number of distinguish-

[1] "Audiometric Methods and Their Applications," published in *Transactions of the American Laryngological, Rhinological, and Otological Society, Inc.*, 1922.

[2] This would be exactly equivalent if the values of $\Delta\alpha$ and ΔP in Fig. 83 were equal for all positions within the inclosed area of this figure.

able tones in the hearing range of the person being tested to the number of distinguishable tones in the normal hearing range.

The charts shown in Fig. 99 illustrate the meaning of this definition of average hearing loss. The charts give the results

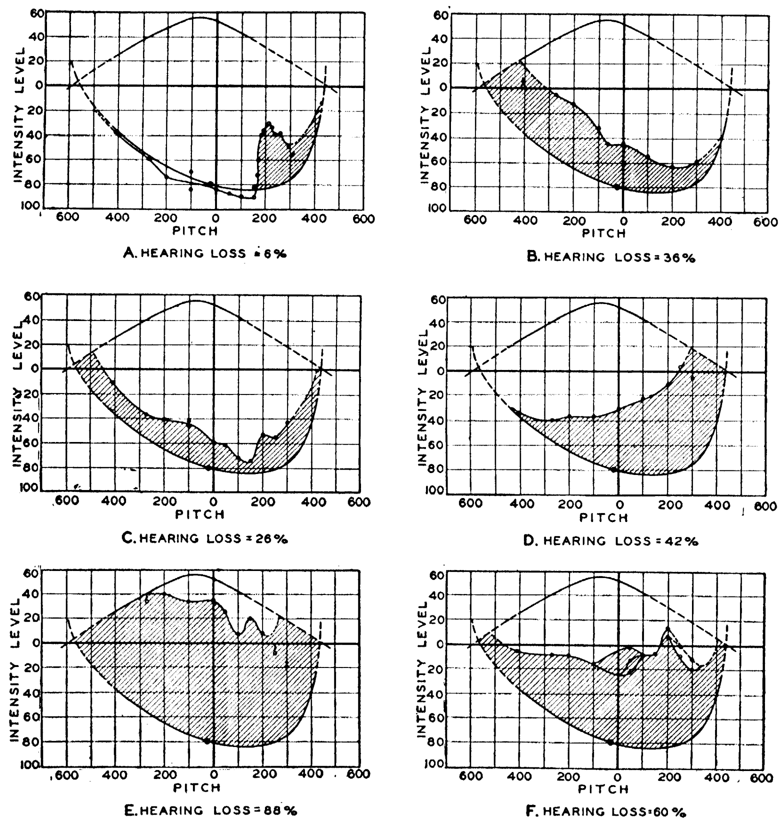


FIG. 99.—AUDIOGRAMS FOR TYPICAL CASES OF DEAFNESS.

for six cases of deafness. The line separating the light from the dark areas is the threshold curve for the deafened person.

It is frequently more useful to the deafened person to know the per cent hearing loss for speech rather than the average per cent loss which is obtained in the manner just described. This is readily done if speech sounds are used in making the test.

As is shown later, it may be calculated from the audiogram. For most purposes it is desirable to express the hearing loss in sensation units rather than per cent hearing loss.

Conversion of Hearing Loss in Sensation Units to Per Cent Hearing Loss

The hearing loss expressed in sensation units can be converted into per cent hearing loss by multiplying the former by a factor K which varies with the type of sound used in the test. It is obvious from the definition given above that this factor is equal to 100 divided by the number of sensation units between the normal threshold of hearing and the normal threshold of feeling for the particular sound used. Experiments made in Bell Telephone Laboratories have indicated that for speech this factor is .83; for the test tone of the 3–A audiometer it is equal to 1.0; for a watch tick, coin click, or acoumeter it is approximately 1.5. For pure tones the value of K is dependent upon the pitch and is given in Table XXI. The question marks indicate that the values of K for tones having either 32, 48, or 12,800 cycles per second have not been accurately determined.

TABLE XXI

FACTORS FOR CONVERTING HEARING LOSS EXPRESSED IN SENSATION UNITS INTO PER CENT HEARING LOSS

Vibration Frequency	K	Vibration Frequency	K
32	3.3 (?)	800	.77
48	2.2 (?)	1,024	.76
64	1.5	1,600	.77
100	1.28	2,048	.77
128	1.09	3,200	.81
200	.96	4,096	.86
256	.91	6,400	.94
400	.83	8,192	1.22
512	.79	12,800	1.42 (?)

Relation between Sensation Units Hearing Loss and the Maximum Hearing Distances for Speech

Let us now consider the relation between the hearing loss expressed in these sensation units and that expressed in the usual voice test method. In this latter method the tester pronounces words while slowly moving away until the patient just fails to interpret them. The degree of hearing obtained from such tests usually is expressed by a ratio of two numbers. For example, $\frac{6}{40}$; the numerator is the distance that the patient can hear and the denominator the distance that a person with normal hearing can hear. If the room in which such tests are made is acoustically treated so that no sounds are reflected from the walls, ceiling, or floor, the intensity of the sound decreases as the inverse square of the distance. In a room with numerous drapings and a heavy carpet, this condition is approximately fulfilled. Laboratory tests have indicated that for an average speaker calling numbers with an intensity corresponding to the musical notation *mf* (mezzoforte) and having the lips at a distance of $1\frac{1}{2}$ inches from the ear, the intensity is 90 db above the normal threshold of audibility or under such conditions the speech has a sensation level of 90 db. These tests have shown also that the intensity is 80 db above the intensity necessary for a 50 per cent correct interpretation of called numbers. For a voice corresponding to *pp* (pianissimo) the sensation level is about 15 decibels lower and for a voice corresponding to *ff* (fortissimo), about 15 decibels higher. Approximately the same results are obtained when using either a *pp* voice or a loud whisper. An average whisper is about 15 db lower than a loud whisper.

Using these data, the inverse square law, and the definition of hearing loss given above, it is possible to calculate for each intensity of calling the maximum distance for hearing and interpreting called numbers by a person of known hearing loss. The necessary equations for doing this are developed in Appendix D. In this appendix it is shown also that the maximum distance at which the normal ear can interpret called numbers is 40 feet for the average whisper, 222 feet for the loud whisper or *pp* voice, 1250 feet for the *mf* voice, and $1\frac{1}{3}$ miles for the *ff* voice. At first thought, these numbers seem to be unreasonably large to conform with our every-day experience, but it must be remembered that we are usually immersed in a continuous noise. This is especially true in the large cities. One needs only to recall the common experience while in the country on an early morning of hearing the cock crow and other familiar sounds at a distant farmhouse to realize that the voice will reach these long distances in a very quiet place.

As mentioned before, the noise in the usual city office is sufficient to shift the threshold of hearing 20 or 30 sensation units without causing any annoyance. The hearing distances corresponding to the four intensities of the voice which would be obtained in such an office might be as small as 15 inches, 7 feet, 40 feet, and 222 feet instead of the larger distances given above. It is seen, therefore, that a speech test made in such an office would not differentiate between ears that are normal and those having a 30 db loss in hearing. The importance of making hearing tests in soundproof booths is thus evident.

The maximum distances at which persons having different amounts of hearing loss can interpret these four intensities of voice were similarly calculated and are given in Table XXII.

It is generally recognized that it is almost impossible to obtain accurate results by means of the usual speech methods because of the uncertainties of controlling the intensity of the voice, and the noise and acoustic conditions of the testing room. However, for rough work these tests are useful and this table should aid in the interpretation of the results obtained from such tests. It is evident that there are two ways of varying the intensity of the speech sounds arriving at the ear of the patient, namely, by varying the intensity of calling and using a fixed distance, or by varying the distance and using a constant intensity of calling. It is seen from Table XXII that when the first method is used, if the tester calls at a distance of 15 inches and the patient cannot interpret an average whisper, his hearing loss is greater than 30 db; if he cannot interpret a loud whisper or *pp* voice, it is greater than 45 db; if he cannot interpret a *mf* voice, it is greater than 60 db; and finally, if he cannot interpret a *ff* voice (shouting intensity) it is greater than 75 db. It is thus seen that a rough measure-

TABLE XXII

MAXIMUM DISTANCES IN A QUIET PLACE FREE FROM REFLECTIONS FOR INTERPRETING CALLED NUMBERS BY PERSONS HAVING VARIOUS AMOUNTS OF HEARING LOSS

Hearing Loss Sensation Units	Average Whisper	Loud Whisper or *pp* Voice	*mf* Voice	*ff* Voice
0	39.5 feet	222 feet	1250 feet	1⅓ miles
5	22.2 feet	125 feet	704 feet	3950 feet
10	12.5 feet	70 feet	395 feet	2220 feet
15	7.0 feet	39.5 feet	222 feet	1250 feet
20	4.0 feet	22.2 feet	125 feet	704 feet
25	2.2 feet	12.5 feet	70 feet	395 feet
30	15 inches	7.0 feet	39.5 feet	222 feet
35	8.5 inches	4.0 feet	22.2 feet	125 feet
40	4.7 inches	2.2 feet	12.5 feet	70 feet
45	2.7 inches	15 inches	7.0 feet	39.5 feet
50	1.5 inches	8.5 inches	4.0 feet	22.2 feet
55	0.8 inch	4.7 inches	2.2 feet	12.5 feet
60	....	2.7 inches	15 inches	7.0 feet
65	....	1.5 inches	8.5 inches	4.0 feet
70	....	0.8 inch	4.7 inches	2.2 feet
75	....	....	2.7 inches	15 inches
80	....	....	1.5 inches	8.5 inches
85	....	....	0.8 inch	4.7 inches
90	....	....	....	2.7 inches
95	....	....	....	1.5 inches
100	....	....	....	0.8 inch
110		May be reached by speaking tube		
115				
120		Totally deaf		

ment of the hearing loss in sensation units is obtained by varying the intensity of calling at a given distance.

Probably more accurate results can be obtained by varying the distance, using a constant intensity of voice. For distances smaller than 6 inches the results will be unreliable due to the uncertain reflections between the head of the listener and the head of the speaker and the difficulty of making accurate measurements of the distance. This limits the maximum loss of hearing that can be measured by this method. The noise conditions and extreme distances limit the minimum loss that can be measured. For an average whisper this practical range is from a hearing loss of 40 to 20; for the loud whisper or *pp* voice from 55 to 20; for the *mf* voice from 70 to 30; for the *ff* voice from 85 to 45; the noise conditions limiting the lower value in the first two cases and the size of the room in the last two. When a greater accuracy than 10 or 15 db is desired it is necessary to calibrate the voice rather than rely upon the tester's judgment as to which intensity of voice he is using. This is done by determining the normal hearing distance or its equivalent for the particular intensity used. If d_0 is such a distance and d the distance that the patient can hear the same voice, then the H.L. (hearing loss) is given by

$$\text{H.L.} = 10 \log \left(\frac{d_0}{d}\right)^2 = 20\,(\log\ d_0 - \log d). \qquad (2)$$

Since the normal hearing distances are so large, indirect means must be used in their determination except for intensities smaller than the average whisper. By combining both methods a fairly large range of hearing can be covered. It depends, however, upon the ability of the tester to change accurately the intensity of calling from one type to the other, an accomplishment that is seldom realized.

When this method is used by a tester who is somewhat over-anxious that the patient show an improvement in hearing it might give quite erroneous results. If he should use the *pp* voice and find the hearing distance 1 foot before treatment, and then in the test after treatment, in his anxiety, raise the intensity of his voice to that corresponding to *mf*, he would find the hearing distance had increased to 5 feet, even though there were no change in the patient's ability to hear. The increase from 1 to 5 feet would sound like a big improvement to the patient.

As shown in Appendix D, *an improvement in hearing for the same patient of* 1 *foot for an average whisper is the same as an improvement of* $5\frac{1}{2}$ *feet for the pp voice, 32 feet for the mf voice, and 178 feet for the ff voice.* Equation (4) of Appendix D shows that if the hearing distance is expressed as a fraction of the normal distance this fraction will be the same for all intensities of voice used. This shows that the common practice of giving results of voice tests as fractions of the normal distance has a logical basis. As will be seen later, there is no such logical reason for expressing the results of tuning-fork tests as fractions of the normal time. As indicated in Table XXII, the intensity range of hearing speech is 120 db, corresponding to a distance variation of 1,000,000 to 1. The range from 0 to 60 db loss or 50 per cent loss for speech corresponds to a distance variation of 1000 to 1. This emphasizes the fact that variation of distance alone is entirely inadequate to cover the complete intensity range for hearing.

Reduction of Watch Tick, Acoumeter, and Coin Click Tests to Sensation Units Hearing Loss

The watch-tick test is probably more familiar to the average person than any other hearing test, because it is so commonly used in physical examinations. The distance for hearing is determined in a similar way to that used in the speech test and the results are recorded similarly as a ratio. These results can be reduced to sensation units hearing loss by the method outlined for voice tests.

In this case, however, reflections from objects in the room, especially from the head of the patient, produce very marked irregularities. This is due to the fact that the components of sound from the watch are mostly in the high frequency region around 2000 cycles per second. At such high frequencies more definite reflection patterns are produced by the objects in the room than for the low frequencies. The dotted line in

Fig. 100 shows a reduction curve which was determined experimentally for a watch tick and illustrates very well the irregularities which commonly exist in a room even when well damped. The solid curve in this figure is based on the inverse square law. This curve is drawn from equation (3) of Appendix D, the abscissas representing values of 20 log d, the value of y_0 being 20 log d_0. It is thus seen that if d_0, the normal hearing distance for the kind of sound used in the test, is determined, the value of y_0 can be read directly from the curve. From the distance d that a patient hears the same sound, the value of y is likewise obtained. If the values of

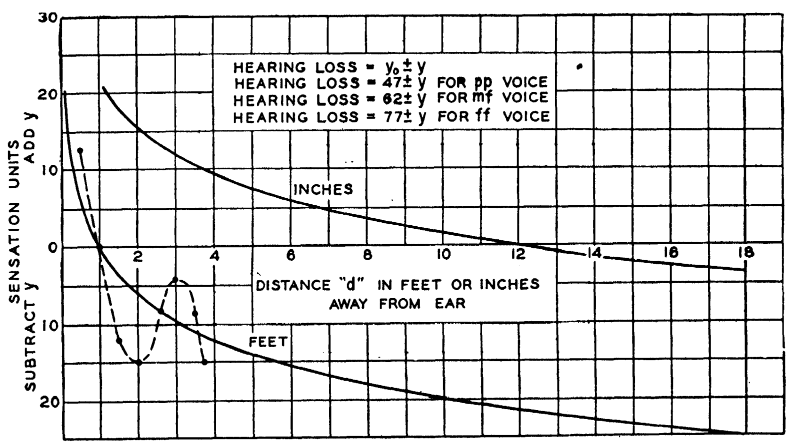


FIG. 100.—CHART FOR REDUCING RESULTS OF SPEECH, WATCH TICK AND ACOUMETER TESTS TO SENSATION UNITS OF HEARING LOSS.

y and y_0 thus obtained are on the same side of the zero line, their difference gives the hearing loss; if on opposite sides their sum gives the hearing loss. For example, the distance d_0 for a Bristol watch is about 6 feet, corresponding to a value of y_0 equal to 15.5. If a patient can just hear this watch at a distance of 4 inches corresponding to a value of y equal to 9.5, then the hearing loss for the watch tick is 25 db.

In Fig. 101 are shown the relative amplitudes of the component frequencies in a watch-tick sound. From this it will be seen that the watch tick is essentially a test of acuity for frequencies in the neighborhood of 2000 cycles per second. The sensation levels of the sound from several types of watches when held in contact with the ear were determined to be in a range from 40 to 70 db. An average for good grade watches

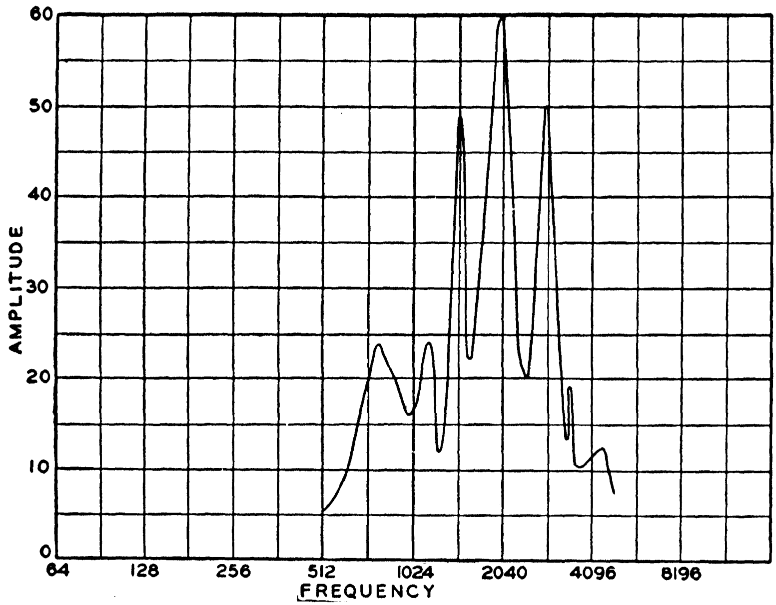


FIG. 101.—FREQUENCY-AMPLITUDE ACOUSTIC SPECTRUM FOR BRISTOL WATCH TICK.

is 45 sensation units. This means that a patient who can just hear such a watch tick has a 45 db hearing loss for frequencies near 2000 cycles per second.

The results of acoumeter or coin click tests can be reduced in the same way. They are open to the same objections as mentioned for the watch tick. This test is also essentially a test for high frequencies.

Reduction of Results of Tuning-fork Tests to Sensation Units Hearing Loss

In the standard method for determining the acuity of hearing at different pitches, use has been made of a series of tuning forks having various vibration rates through the audible range.

The fork is given a standard blow of some sort to set it into vibration. It is then held as close to the ear as possible, preferably with the flat part of the prong directly facing the auditory meatus. The time t in seconds from the striking of the blow until the patient no longer hears the sound is observed. A comparison of this time to that t_0 required for normal hearing gives a measure of the hearing by air conduction. It is well known in dynamics that the time difference $t_0 - t$ is proportional to the logarithm of the ratio of sound intensities created corresponding to each time. Since the hearing loss expressed in sensation units is also proportional to the logarithm of this ratio, it follows that

$$\text{H.L.} = \Delta(t_0 - t) \tag{3}$$

where Δ is the constant of proportionality. The constant is dependent upon the damping of the tuning fork and is the change per second in the intensity level of the sound produced. Any experimental method which will measure this rate of change will be suitable for determining the constant Δ. It is seen that the hearing loss is obtained by multiplying the decrease in time for hearing the fork by a constant of the fork. If it is desired to reduce the results to per cent hearing loss, the following relation is used,

$$\text{per cent hearing loss} = K\Delta(t_0 - t), \tag{4}$$

the factors K and Δ having definite numerical values for each fork used. For illustrating the kind of constants one might expect to obtain, the values of Δ for three groups of tuning forks are given in Table XXIII. For convenience, the values of K from Table XXI are also given. The first group of forks is from a set used in Bell Telephone Laboratories. The second set is used by Dr. E. P. Fowler. The last one is a 500-cycle standard fork used by Dr. Douglas Macfarlan. For example, when Dr. Fowler's 256 fork is used, the per cent hearing loss is given by

$$\text{per cent hearing loss} = 1.25\ (70 - t), \tag{5}$$

where t is the air conduction time for the patient. It is evident from the formula that a person having more than $87\frac{1}{2}$ per cent loss will not hear the fork at all.

TABLE XXIII

TYPICAL CONSTANTS FOR TUNING FORKS

Rate in *dv*	Damping Constant Δ	Time t_0 in Seconds for Normal Air Conduction	Factor K for Reducing Hearing Loss to Per Cent Hearing
	Bell Telephone Laboratory Forks		
24	.30	75	4. (?)
48	1.9	51	2.2 (?)
64	1.61	41	1.5
100	1.75	30	1.28
200	.93	110	.96
400	.46	140	.83
500	.59	135	.81
800	.87	112	.77
1000	1.19	71	.76
1200	1.14	69	.76
1800	2.29	44	.77
2000	2.41	45	.77
	Dr. E. P. Fowler's Forks		
128	1.08	65	1.09
256	1.38	70	.91
512	1.31	95	.79
1024	1.70	40	.76
2048	2.17	20	.77
	Dr. D. Macfarlan's Standard 500 Cycle Fork		
500	2	55	.81

$$\text{Hearing loss} = \Delta\,(t_0 - t)$$
$$\text{Per cent hearing loss} = K\Delta\,(t_0 - t)$$

It is useful to notice that if the hearing loss of the tester is known, the hearing loss of the patient can be found as follows: Set the fork into vibration by any means, not necessarily using a standard blow. Hold the fork to the patient's ear in the standard fashion. Start the stop watch when the patient signals he no longer hears the tone. Hold the fork to your own ear in the same standard way. Stop the watch when you cease to hear the tone. Then the reading of the watch in seconds, multiplied by the constant of the fork Δ, gives the difference in db between your hearing loss and that of the patient.

This method has the advantage that the results are not dependent upon the initial blow given to the fork, but it has the disadvantage that it depends upon the hearing of the tester, and any noise in the room affects the hearing of the tester much more than that of the patient, provided the former is much more acute than the latter.

For example, suppose that you know that the tester's hearing loss is 20 db at 1024 cycles per second, and also, when using the technic described above and Dr. Fowler's 1024 fork, the time difference is found to be 25 seconds. Then the hearing loss of the patient is

$$1.70 \times 25 + 20 = 62.5$$

or his per cent hearing loss is

$$62.5 \times .76 = 47.5 \text{ per cent.}$$

It would be a great advantage to those who use tuning forks if the makers of such forks for otologic purposes would furnish values of the damping constants as well as the frequency of vibration. Then the results of tuning-fork tests could be reduced to a common basis, as indicated above.

In order to increase the speed and accuracy of making hearing tests which are equivalent to those which have been described, new types of measuring instruments have been developed which are called audiometers. For making speech tests, a phonograph type of audiometer has been developed. In this instrument records which have been made by special electrical processes are used for reproducing the speech sounds. This insures a definite volume of voice, so that accurate comparisons are possible. The recorded speech sounds are transformed into their electrical equivalents by means of an electromagnetic reproducer. The electrical waves thus created are carried by means of wires to a receiver which is held to the patient's ear. The intensity of the sound sent into the patient's ear is controlled by electrical resistances in the electrical circuit. This phonograph type audiometer is now available in three different forms, according to the purpose for which it is to be used. For making a survey of the hearing of a large

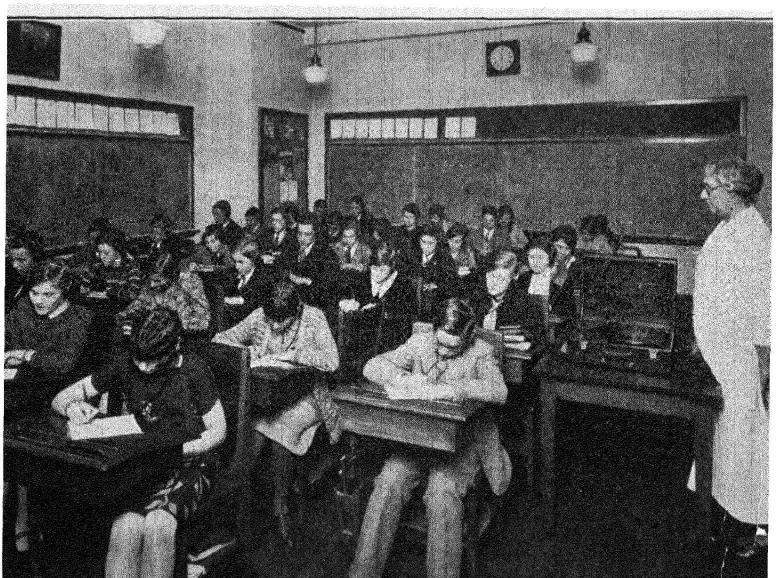


FIGURE 102.

group to find those who have a deficiency for hearing speech, it is furnished with the turntable unit, special records upon which the recorded speech sounds continually decrease in intensity as the record revolves, and multiple units containing 8 receivers per unit. A master sheet is furnished, which gives the hearing loss corresponding to the intensity of each of the numbers on the record. In this form no batteries or attenuators are required. An instrument of this sort was tried out at one of the public schools in New York City, and it was found that with a single phonograph unit and 5 multiple units that is, 40 receivers, the pupils could be tested at the rate of about 100 per hour. The entire school of about 1000 pupils was tested in three days, and a number determined for each pupil, which represented his hearing ability for speech. In Fig. 102 a picture of this instrument in operation in a schoolroom is shown. Through the efforts of the American Federation of Organizations for the Hard of Hearing about 250,000 school children had their hearing tested with instruments of this type last year (1927). These tests indicated that from 8 to 12 per cent of the children have defective hearing. It was through the cooperative efforts of this organization and Bell Telephone Laboratories that this instrument was developed.

In Fig. 103 a close-up view of the turntable unit and the tray unit is shown. For purposes of testing more accurately the hearing of one individual at a time, an attenuator is interposed between the turntable unit and the receiver, and a record is used which gives a constant speech level. For use in schools for the deaf or for testing patients who are very hard of hearing, an amplifier unit is interposed between the turntable unit and the attenuator box. By means of the attenuator box, the

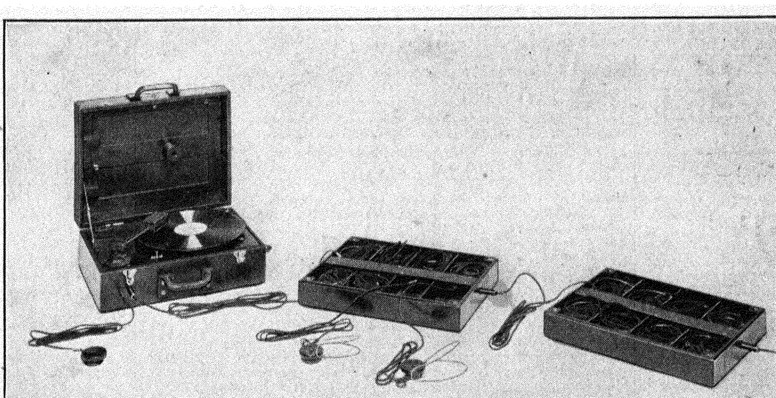


FIGURE 103.

speech intensities being delivered to the ear of the patient can be varied 10 billion-fold by turning the dial from one end of the scale to the other, this intensity variation corresponding to a distance variation of more than 100,000. This makes it possible to test all degrees of hearing from normal hearing to total deafness. Not only does it have this increased range over that obtainable in the ordinary speech test, but it has a definite intensity level for the speech and is not subject to variations in the acoustic properties of the room.

The 3–A audiometer was developed to take the place of such tests as the watch-tick tests, the acoumeter tests, or those designed to make a very quick test of the general hearing level. In this instrument a tone having components throughout the entire speech frequency range is electrically generated and delivered to a receiver to be held on the ear of the patient. The volume of the tone is controlled by the same attenuator unit used in the other audiometers. It reads directly, either in db loss or per cent hearing loss, since it was found that for this tone there were approximately 100 db between the threshold of hearing and the threshold of feeling for the normal ear. A picture of this instrument as used for measuring noise was shown in Fig. 58. When used for measuring hearing the offset receiver shown is replaced by the usual head receiver. This instrument has been found to be particularly useful in schools for the deaf. It enables the teachers to grade the degree of hearing of the child very quickly, and thus aids them in deciding upon the kind of methods to be used in teaching him. It is also useful in making a quick test of the hearing of large groups when they are tested one at a time. This 3–A audiometer is also useful when making tests of bilateral deafness, for it can be used in place of the Bárány noise apparatus. For this purpose it is superior to the latter device, because the volume can be placed at exactly the proper level to sufficiently mask the hearing in the good ear without in any way interfering with the hearing in the bad ear. The proper level for doing this depends upon the relative difference in the hearing of the two ears.

A modification of this instrument so that the generator derives its power directly from the alternating current usually supplied for lighting purposes has recently been made. This modified instrument is called the 5–A audiometer. A set similar to this buzzer-type instrument, but having a very limited intensity range, was developed some time ago by Seashore. It has been used mainly by psychologists in testing variations in the acuity of hearing of persons having no noticeable defect in the hearing mechanism.

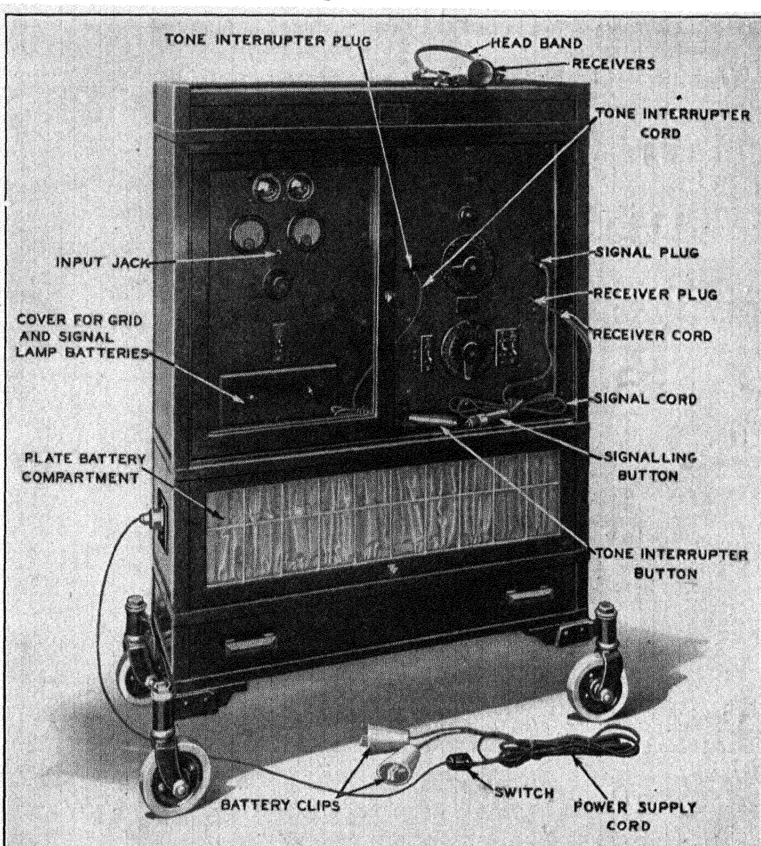


FIGURE 104.

For replacing the tuning-fork test several types of audiometers have been developed by Seashore, Dean and Bunch, Knudsen and Jones, Kranz, Bell Telephone Laboratories, and others. These instruments consist essentially of a generator for producing alternating currents of various frequencies, an

Audiogram blanks are furnished with the instrument, which makes it possible to show graphically the hearing loss at each frequency.

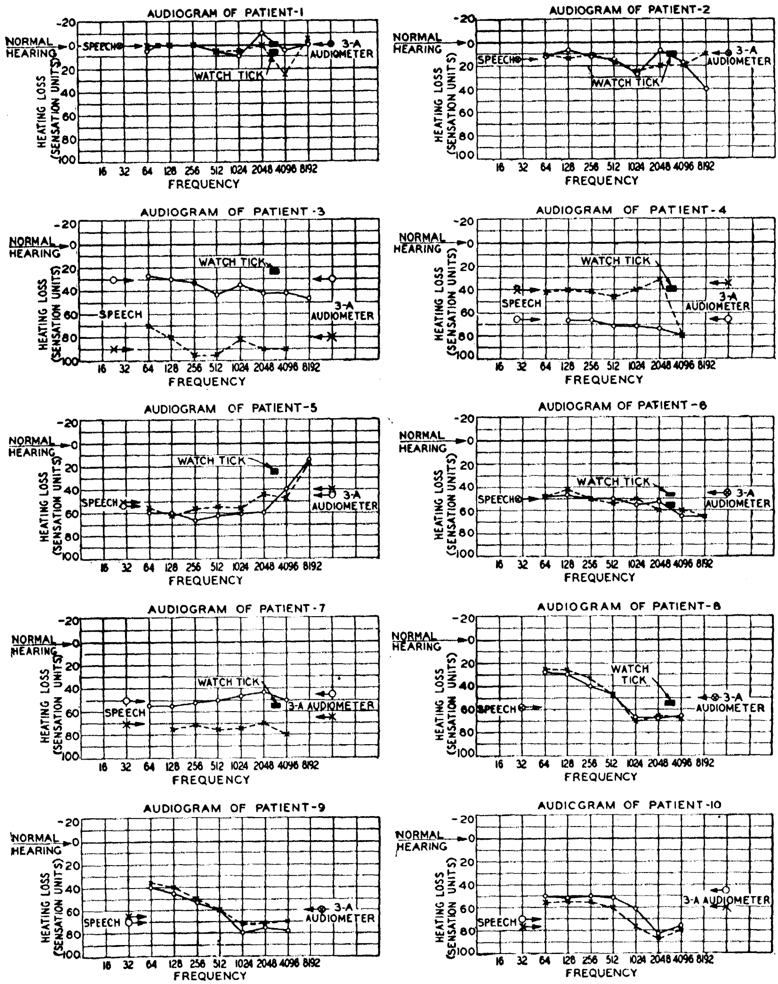


FIGURE 106.

The 1–A audiometer is similar in principle to the 2–A audiometer, but has a much greater range in frequency and intensity. Both of these instruments are equipped with a jack, Audiogram blanks are furnished with the instrument, which makes it possible to show graphically the hearing loss at each frequency.

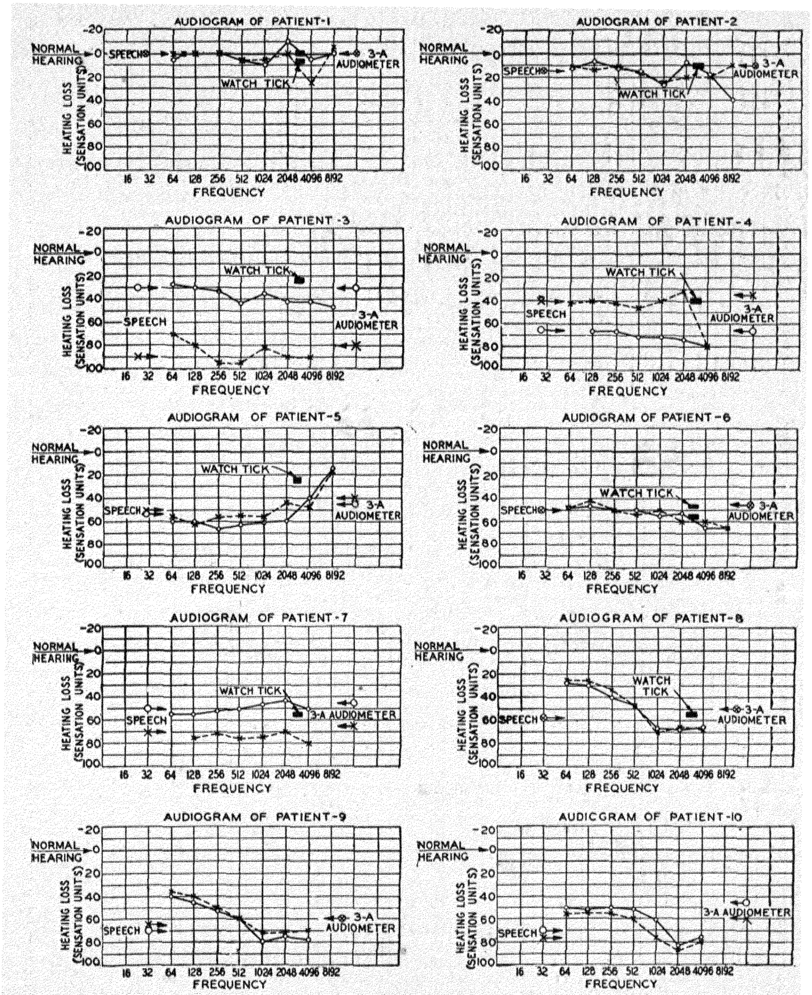


FIGURE 106.

The 1–A audiometer is similar in principle to the 2–A audiometer, but has a much greater range in frequency and intensity. Both of these instruments are equipped with a jack, so that auxiliary equipment may be used with them. The otologist making the test frequently desires to talk to the patient either before or during the time that the pitch-range test is being made. For this purpose a talking set has been provided. By plugging the end of the cord of the talking set into the jack of the audiometer, it is possible to talk directly to the patient through the receiver which he is using during the test. By turning the attenuating dial, the volume of the speech thus delivered to the patient's ear can be varied in intensity from high values to the threshold of audibility. Also, the phonograph audiometer may be used as an auxiliary to both the 1–A and the 2–A audiometers by plugging the cord from the turntable unit of the phonograph audiometer into the appropriate jack. Then the speech waves coming from the phonograph record will be delivered directly to the receiver of the patient's ear. The speech volume may be controlled by the attenuator dial. A mark on this dial can be determined so that readings on the scale from this mark will give directly the patient's hearing loss for speech.

To show the comparative results obtained by different methods of testing, the degrees of hearing of ten persons were tested by four different methods, namely, the phonograph audiometer, the 3–A audiometer, the standard speech test, and the watch-tick test methods. The results obtained are shown in Table XXIV.

Audiograms obtained by the 1–A and the 2–A audiometers for the persons tested are given in Fig. 106. On these audiogram charts the hearing losses for speech and for the 3–A audiometer tone are shown at the left and right, respectively. The black rectangle represents the watch tick results. A comparison of the results shown in columns III and IV of Table XXIV shows that the phonograph audiometer gives results which are in good agreement with the reduced results from the standard speech tests. The differences are well within the observational error.

The watch tick results are just what one would expect from the amplitude frequency characteristic of this type of sound.

It is interesting to note that if patients 3 and 5 were given a watch-tick test only, they would be considered to have the same amount of hearing, but further tests showed that patient 3 could hear speech at 20 feet, while patient 5 could hear it only 2 feet away. It is evident from the audiograms why this is

TABLE XXIV

I Ear Number	II Distances for Interpreting Numbers		III Hearing Loss Cal. from II		IV Hearing Loss Phonograph Audiometer	V Distances for Watch Tick Test		VI Hearing Loss Cal. from V		VII Hearing Loss 3-A Audiometer
	Voice A $Y = 58$	Voice B $Y_0 = 65$				A $Y_0 = 10$	B $Y_0 = 15$			
1—R	800 ft. (cal.)	1800 ft. (cal.)	0	0	0	40 in.	70 in.	0	1	0
1—L	800 ft. (cal.)	1800 ft. (cal.)	0	0	0	20 in.	36 in.	6	5	0
2—R	140 ft. (cal.)	320 ft. (cal.)	15	15	15	10 in.	31 in.	12	7	10
2—L	140 ft. (cal.)	320 ft. (cal.)	15	15	15	15 in.	64 in.	8	2	10
3—R	20 ft.	20+ ft.	32	?	30	3 in.	$5\frac{1}{2}$ in.	22	22	30
4—L	3 ft.	6 ft.	48	49	35	......	1 in.	..	39	35
5—L	1.7 ft.	3.7 ft.	53	54	45	2 in.	4.5 in.	26	23	40
5—R	1.8 ft.	3.7 ft.	52	54	50	3 in.	6.5 in.	22	21	45
6—L	2 ft.	3.5 ft.	52	54	50	......	contact	..	55	45
6—R	1.5 ft.	4 ft.	54	53	45	......	.5 in.	..	45	45
7—R	.9 ft.	2 ft.	57	52	50	......	contact	..	55	45
8—L	1.2 ft.	3.7 ft.	56	54	60	......	....	..	..	50
8—R	1.2 ft.	3.7 ft.	60	54	60	......	contact	..	55	50
4—R	No test				65	......	....	..	..	65
9—L	8 in.	13 in.	62	64	65	......	....	..	..	60
9—R	5 in.	8.5 in.	65	68	70	......	....	..	..	60
7—L	8 in.	15 in.	61	63	70	......	....	..	..	65
10—R	3.5 in.	9 in.	69	68	70	......	....	..	..	45
10—L	2 in.	6 in.	74	71	75	......	....	..	..	60
3—L	No test				90	......	....	..	..	80

true. Patient 5 hears the high frequencies very much better than the low frequencies. An examination of the various audiogram charts indicates that the 3-A audiometer gives a good criterion of the general hearing level. However, results obtained by it are not definitely related to the hearing loss for speech, although there is a general agreement. A notable exception is the lack of agreement in the results obtained by patient 10. The 3-A type audiometer gave 45 db for the right and 60 db for the left ear. The phonograph audiometer gave the hearing loss for speech as 70 db for the right and 75 db for the left ear. An examination of the audiogram for patient 10 shows why this should be. The most important frequencies for speech interpretation—that is, from 500 to 2000—are considerably below the general level, the low frequencies being the

TABLE XXV

HEARING LOSS FOR SPEECH

Ear Number	Calculated from Audiograms	Observed Average	Ear Number	Calculated from Audiograms	Observed Average
1—L	2	0	6—L	55	50
1—R	2	0	6—R	48	50
2—L	20	15	7—L	48	50
2—R	16	15	7—R	73	70
3—L	39	31	8—L	61	58
3—R	89	90	8—R	60	58
4—L	40	40	9—L	68	65
4—R	72	65	9—R	72	70
5—L	52	50	10—L	75	75
5—R	61	55	10—R	66	70

highest. This is also true for patients 8 and 9. For patient 5 the speech frequencies are on about the same level as the low frequencies, but very little loss is shown for the high frequencies. In this case it is probable that the 3-A audiometer tone was heard because of components between 2000 and 8000 cycles per second.

As will be seen from the experiments discussed in Part Four the important frequencies for recognizing speech are between 500 and 2000 cycles per second. The simple method of taking an average of the hearing losses at 512, 1024, and 2048, for determining the hearing loss for speech gives results in good agreement, with observations on these ten patients. In Table XXV is shown a comparison of results thus obtained. The observed values are averages of the figures given in columns III and IV of Table XXIV. The per cent hearing loss is obtained by multiplying the figures by .83. Until more accurate methods are devised, this simple procedure will be useful in calculating from the audiogram the per cent hearing loss for speech, which figure is of greatest interest to the patient.

It is thus seen that by the use of the hearing loss scale in db, it is possible to express the results of the different methods of testing upon a common basis so that they may be directly compared.

PART FOUR

The Perception of Speech and Music

CHAPTER I

The Loudness of Sounds

When a sound of any character is impressed upon the ear the magnitude of the sensation produced is called the loudness of the sound. It is related to the intensity of the sound but the relationship is very complicated, depending upon the character of the sound. Two sounds which produce equal intensities at the ear are not generally recognized as being equally loud. Let two sound sources of different character be adjusted so that they sound equally loud. Then let the intensity level of each of the sounds be raised the same amount. In general they will no longer sound equally loud.

For these reasons, neither intensity level nor sensation level can be taken as a measure of loudness. Some scale must be used so that the loudness of one sound as indicated by the number assigned to it will be the same as that of any other sound having the same number on the loudness scale.

It has been suggested that the unit of loudness be chosen to be the least perceptible increment in intensity. As seen from Part Three, Chapter III, the fractional perceptible increase varies through a wide range depending upon the sensation level and pitch of the tone. Although the relations would be simpler if this fraction were constant so that a logarithmic scale could be used, nevertheless, a scale could be built with the threshold intensity as zero loudness and with each perceptible increment above this intensity designated as one additional unit of loudness. The numbers on such a scale would then be the number of distinguishable gradations in intensity for any sensation level $\alpha - \alpha_0$ or if L represents the loudness on such a scale.

$$L = \int_{\alpha_0}^{\alpha} \frac{1}{\Delta\alpha}\, d\alpha. \qquad (1)$$

The difficulty with such a scale is that when dealing with tones of different pitch, equal loudness numbers do not correspond to tones which sound equally loud. For example, a tone of pitch -4 octaves and at a sensation level of 40 db sounds equally loud to a tone having a pitch level of zero and which is at a sensation level of 70 db; the loudness numbers on the above scale corresponding to these two tones are 20 and 125, respectively. This difficulty can be avoided if some sound be chosen as a reference for comparison, for example, a pure tone having a pitch level of zero. The loudness of any other sound would then be represented by the loudness of this reference standard when it is adjusted to sound equally loud to the sound being measured. However, after choosing such a standard, the scale based upon equal detectable increments loses its significance and is apt to be misleading. For these reasons, it seems better to choose the sensation level of a pure tone having the zero reference pitch, that is, corresponding to a frequency of 1 kilocycle per second, as a measure of its loudness. The loudness L expressed in decibels of such a reference tone is related to the intensity level α, the intensity I expressed in microwatts, and the pressure variation near the drum of the ear p expressed in bars, by the formula

$$L = \alpha + 92 = 10 \log I + 92 = 20 \log p + 66. \qquad (2)$$

The loudness of any other sound, whether a pure tone having a different pitch, a musical tone, or any other complex sound, is measured by the loudness of the reference tone which sounds equally loud as judged by an average normal ear.

Loudness of Pure Tones

Those who have tried to make measurements on the loudness of sounds having different pitches realize the difficulty in obtaining from different individuals judgments which are consistent. The average normal ear as used above is hypothetical but nevertheless very important. It implies that before any observed value is sufficiently reliable so that it may be duplicated by another experimenter, the results from a large number of observers must be obtained. This is true of most of the work described in this part of the book.

Some pioneer work on the relative loudness of pure tones using organ pipes as sources of sound was done by Sabine.[1] The sensation level of each tone was determined from the time necessary for the intensity of the sound to decrease to the threshold after the source was cut off. MacKenzie [2] of Bell Telephone Laboratories also made measurements of relative

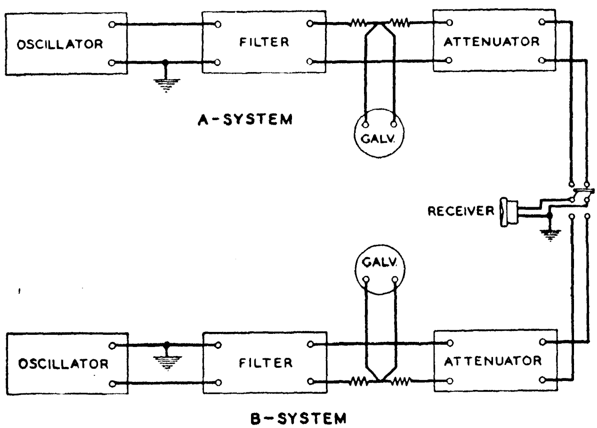


FIG. 107.—SCHEMATIC DIAGRAM OF APPARATUS.

loudness using an instrument called the Alternation Phonometer. Due to the rapidity of alternation of the tones being compared (twenty-five per second) there is some question as to the general application of his results.

The most comprehensive work on the relative loudness of pure tones is that reported by Kingsbury of Bell Telephone Laboratories. He found that reliable average values were obtained by using as observers eleven men and eleven women. Figure 107 shows a schematic diagram of the apparatus which

[1] "Collected Papers on Acoustics," p. 130.
[2] *Physical Review*, October, 1922.

he used. A 700-cycle tone was used as a standard of comparison. The attenuator in the A system was arranged so that the tone of this frequency generated by the receiver could be brought to any intensity level. Three independent settings of A for the threshold of intensity were then made. The comparison tone was then generated by system B, and similarly, three measurements on the threshold of the tone of this system were obtained. Next, the experimenter set the A attenuator at one of the selected comparison levels and allowed the observer to adjust the B attenuator until when listening alternately to the two tones they seemed equally loud. The attenuation settings and the deflections of the meters were then recorded for both systems. This process was repeated for the other fixed levels of the tone A until three independent determinations had been made for each level. The order of taking the fixed levels of the A tone was made as random as possible and two successive determinations were seldom made at the same level. When the comparison of the two frequencies was finished, the A and the B thresholds were once more secured.

The work was then repeated, using a tone of different pitch. In this way the sensation levels of twelve tones which appeared to have the same loudness as the reference 700-cycle tone were obtained for eight selected levels of this tone. In Table XXVI the results of these measurements are given. Each value is the average for sixty-six observations, three independent measurements being taken for each of twenty-two persons. In the first and second columns the frequency in cycles per second and the pitch in centioctaves of the comparison tone are given. In the third column the values of sensation levels in db for tones which sound equally loud to a 700-cycle tone, which is 9.3 db above the average threshold, are given. Similarly, each of the remaining columns gives the values for equal loudness when the 700-cycle comparison tone is at the value indicated opposite 700.

According to the definition of loudness given above, any particular value opposite 1000 gives the loudness of all the tones

TABLE XXVI

LOUDNESS OF PURE TONES

Frequency	Pitch	Sensation Level							
60	−406	8.4	12.5	15.7	19.2	24.8	29.9	37.3	41.9
80	−364	6.6	11.4	14.9	19.6	25.1	31.7	36.1	43.9
150	−273	9.9	15.2	20.8	25.6	32.5	39.7	47.8	52.9
200	−232	12.7	16.4	23.2	25.9	36.5	44.8	54.4	62.2
340	−156	9.4	18.0	25.1	33.8	41.7	50.7	61.6	73.5
440	−118	9.3	17.7	25.7	33.8	42.2	52.5	62.9	75.3
700	− 51	9.3	19.3	29.3	39.3	49.3	59.3	69.3	79.3
1000	0	11.7	20.7	31.1	40.7	52.1	60.9	69.8	78.5
1500	+ 58.5	11.6	21.4	32.5	42.5	52.5	61.7	71.3	79.2
1900	+ 92.7	11.9	22.4	36.3	45.7	56.7	65.0	73.7	80.2
3200	+168	9.9	18.5	31.6	43.1	52.8	61.1	70.0	77.0
4000	+200	8.9	22.1	31.6	44.0	53.2	61.1	68.9	
Loudness....	....	10.4	20.7	32.1	42.4	52.4	61.5	70.5	78.8

corresponding to the values in that particular column. For example, the loudness of all the tones giving the values in the sixth column is 40.7 db. Values of loudness given in the row corresponding to 1000 cycles are necessarily subject to an observational error. To obtain more accurate values, the average for tones of frequencies 700, 1000, 1500, 1900, 3200, and 4000 was used. This procedure is only justified because the data indicate that for tones in this range equal loudness changes correspond to equal sensation level changes. That this is true is another reason for choosing the type of loudness scale indicated above. The loudness values obtained from the averages of these six tones are given in the bottom row of this table. Using these values, curves can now be drawn showing the relation between loudness and sensation level for the various frequencies. In Fig. 108 such curves are shown. It is seen that the tones of low pitch increase in loudness much faster than those of high pitch.

All of the values given above were obtained by direct comparison with the 700-cycle tone. The question arises: "Will two tones having the same loudness, as indicated by such tests, sound equally loud when directly compared?" Kingsbury tested this point for the two tones of frequencies 200 and 3200 cycles using eleven observers. The results were

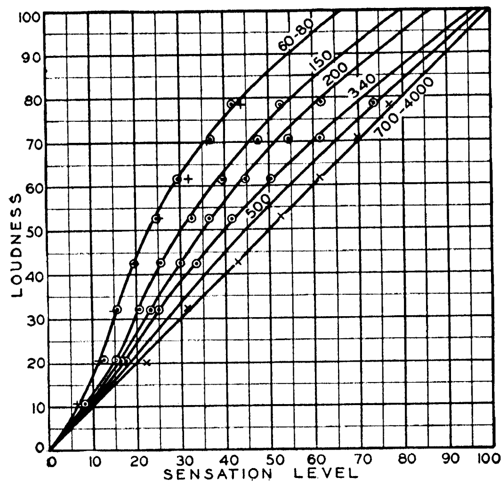


FIG. 108.—LOUDNESS OF PURE TONES.

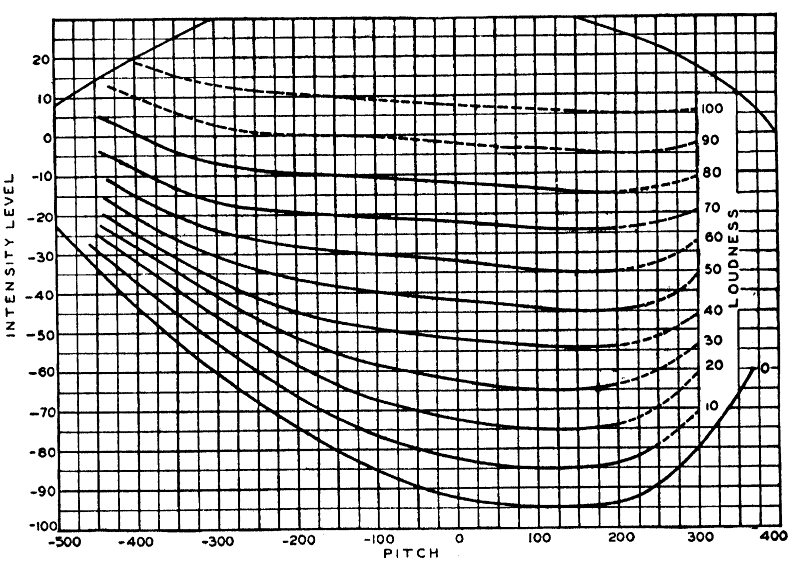


FIG. 109.—CONTOUR LINES OF EQUAL LOUDNESS FOR PURE TONES.

found to agree within the observational error, so it seems reasonable to conclude that two pure tones having the same number on the loudness scale chosen will sound equally loud as judged by an average normal ear.

In Fig. 109 contour lines of equal loudness for pure tones are shown. These lines were derived from the data shown in Table XXVI. They are useful in determining intensity levels from loudness balances or loudness from intensity level measurements. For example, a tone having a pitch of − 3 octaves will have a loudness of 50 db at an intensity level of − 30 db. A tone having the same loudness at three octaves higher in pitch has an intensity level of − 42. The dotted portions of the curves go beyond any experimental data and are only the author's estimate of how they should go.

Loudness of Complex Sounds

The comparison of the loudness of complex sounds which are different in character is also very difficult, but reliable averages can be obtained if a sufficient number of observers and trials are used. The 3–A audiometer mentioned in Part Two, Chapter II, and Part Three, Chapter VI, is useful for making loudness measurements of complex sounds. It was calibrated in terms of the standard 1000-cycle tone by making loudness balances with seven observers. The average results are shown in the curve of Fig. 110. Then to determine the loudness of any sound, the 3–A audiometer is adjusted so that the tone from its receiver has the same loudness as the sound being measured; from the dial reading and the relation expressed in the curve of Fig. 110, the loudness can then be determined.

In a similar way the loudness values at different sensation levels of four different complex sounds designated as A, B, C and D were determined and indicated in Fig. 111. The sound A consisted of repetitions of the sentences "Joe took father's shoe bench out" and "She was waiting at my lawn." These sentences were selected because they contain all the fundamental speech sounds which are important from a loudness standpoint. They are used in the laboratory for testing the efficiency of telephone apparatus. To insure that the loudness of this speech would remain at any desired level, it was recorded on a phonograph and reproduced by means of an electro-

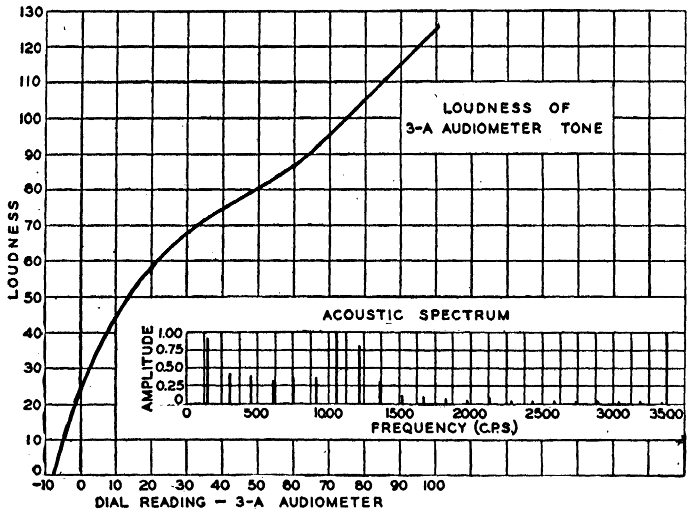


FIG. 110.—LOUDNESS OF 3-A AUDIOMETER TONE.

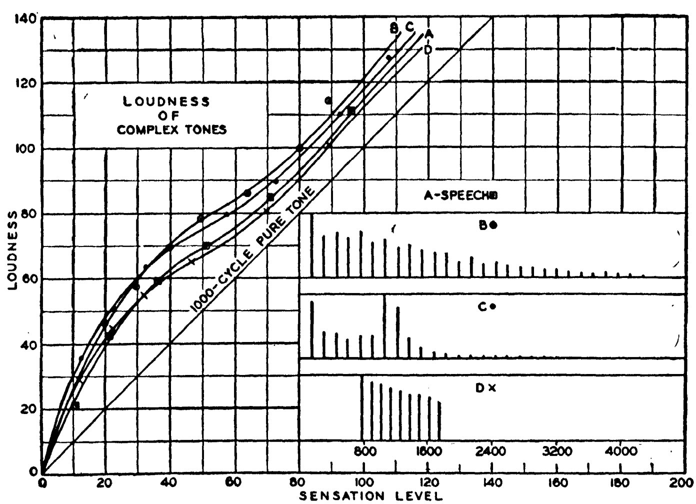


FIG. 111.—LOUDNESS OF COMPLEX TONES.

magnetic reproducer and telephone receiver. By means of an attenuator in the electrical circuit, the level of the speech coming from the telephone receiver could be adjusted to any value.

Acoustic spectra of the sounds B, C, and D are also given in the figure. As a general rule those sounds which have a large number of components increase in loudness at a faster rate with an increase in sensation level than those with a

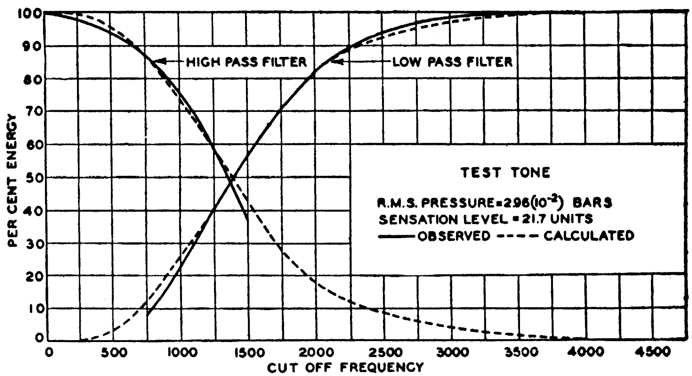


FIG. 112.—PER CENT OF UNFILTERED ENERGY EQUIVALENT IN LOUDNESS TO FILTERED ENERGY.

smaller number of components. Also those sounds having most of the energy in the low-frequency regions increase in loudness faster than those with the energy in the high-frequency regions.

Measurements were made to find the effect upon the loudness of the sound B when certain of its components were eliminated. This sound was created by means of an electrical buzzer which was connected to a telephone receiver. The circuit was arranged so that by means of electrical filters (see Figs. 134 and 135), any of the components of this sound could be eliminated. Tests were made to determine the effect on the loudness of eliminating certain frequency regions at three sensation levels, namely, 22, 43, and 70 db, respectively. The results of these tests are shown in Figs. 112, 113, and 114. The horizontal axis gives the cut-off frequency of the filter used. In the curve labeled "High Pass Filter" results are shown for the case when all frequencies below the cut-off frequency point are eliminated; in the curve labeled "Low

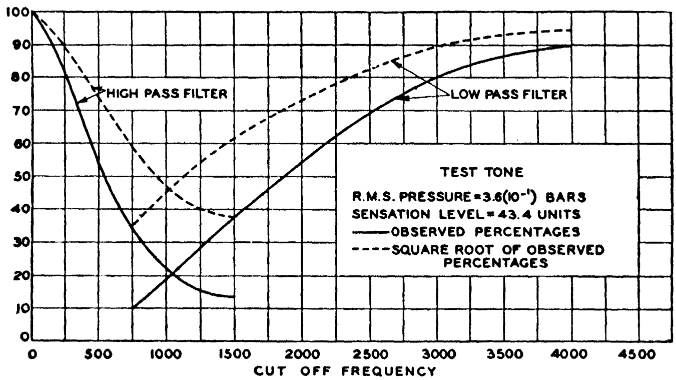


FIG. 113.—PER CENT OF UNFILTERED ENERGY EQUIVALENT IN LOUDNESS TO FILTERED ENERGY.

Pass Filter" for the case when all the components having frequencies above the cut-off frequency are eliminated. The vertical axis gives the per cent of its initial value to which the

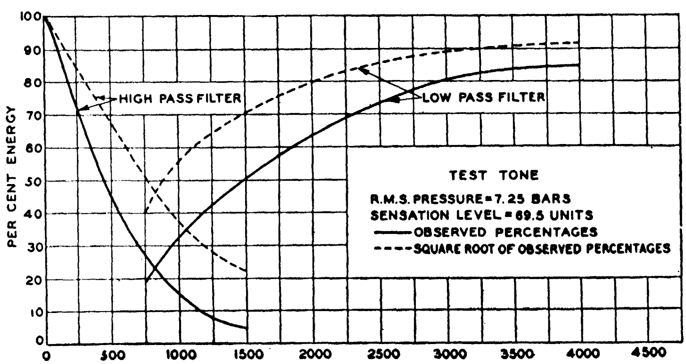


FIG. 114.—PER CENT OF UNFILTERED ENERGY EQUIVALENT IN LOUDNESS TO FILTERED ENERGY.

intensity of the unfiltered tone can be reduced before it has the same loudness as the filtered tone. For example, the point (1500, 50) on the low pass filter curve in Fig. 114 indicates that when all the components above 1500 cycles are eliminated, the loudness of the filtered sound is so reduced that the unfiltered sound must be reduced 50 per cent in intensity or 3 db to sound equally loud.

In telephone work a large amount of loudness balancing is done with a source of sound consisting of speech which has been transmitted through various kinds of systems. Experiments

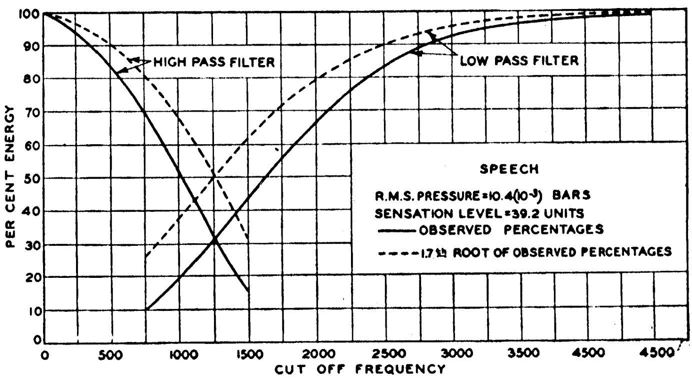


FIG. 115.—PER CENT OF UNFILTERED ENERGY EQUIVALENT IN LOUDNESS TO FILTERED ENERGY.

similar to those described for the test tone were therefore made with speech, the speech being transmitted through the high quality telephone system shown in Fig. 127. The results are shown in Figs. 115 and 116.

The frequency at the intersection point of the curves in each case has an important meaning. The sound composed of only those components above this frequency will appear to have the same loudness as a sound with only those components below this frequency. At low intensities the stimulating forces will be confined to the lower part of the basilar membrane in one case and to an upper part for the other case. Since equal loudness sensations are produced, it is presumed that the number of nervous impulses sent to the brain are the same for the two cases. This seems to indicate that at such intensities the number of impulses is proportional to the intensity and that the law of superposition holds. At higher intensities, however, the stimulation for the filtered tone having the low frequencies is not confined to the lower half of the basilar membrane but due to the subjective tones extends into the upper half also. Similarly, for the other filtered tone due to the difference tones, the lower half of the basilar membrane is partially stimulated as well as the upper half. When both tones produce stimulation simultaneously, then, on account

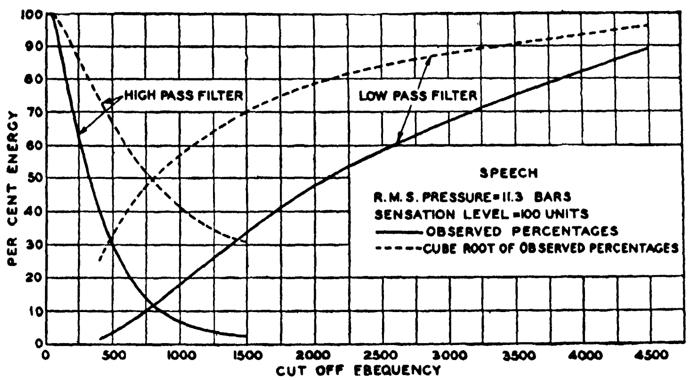


FIG. 116.—PER CENT OF UNFILTERED ENERGY EQUIVALENT IN LOUDNESS TO FILTERED ENERGY.

of the non-linearity of the various parts of the mechanism there is an interaction between the tones so that the law of superposition no longer holds.

The tone with the lower range of frequencies increases in loudness faster than the one with the upper range as the intensities of both are increased. Therefore, the dividing frequency for equal loudness shifts to the lower frequencies as seen in Figs. 112, 113, and 114. For example, for the test tone the dividing frequency shifts from 1300 for a sensation level of 22; to 1100 for a sensation level of 43; and to 800 for a sensation level of 70. For the two higher levels the intersection point is about 25 per cent instead of 50 per cent. In other words, the intensity of the unfiltered tone must be reduced to one-fourth its value to have the same loudness as either part of the filtered tone. The addition of one filtered tone to the other is equivalent from a loudness standpoint to raising the intensity of either one separately to four times its original intensity.

It has been stated above that the elements of the middle ear taking part in the transmission of the sound have a non-linear characteristic which accounts largely for the subjective harmonics. The data indicate that a linear relation exists between the intensity and the number of nervous impulses for very low intensities, but for higher intensities this relation does not hold. This non-linearity together with that produced in the middle ear, accounts for the effects just described. When the two filtered tones act together, no more energy can be expended in a given time than if each acted separately on different ears, but in the former case the stimulated energy is distributed differently along the basilar membrane in such a way that more energy is applied at those positions where a large number of nerve fibres are stimulated.

To obtain further light on this point loudness balances were made with two ears vs. one ear listening to the sound C. To do this the sound was adjusted in each receiver so that when the sound was thrown on alternately in the right and left ear, it produced the same loudness. The sounds in the two receivers were then produced simultaneously. It was found that the intensity of the sound when listened to binaurally must be reduced α db to sound equally as loud as the same sound listened to monaurally. The number of nervous impulses reaching the brain when listening binaurally must be twice that when listening monaurally. Consequently, this gives a means of determining the number of db the sound must be raised in sensation level in order to double the number of nervous impulses. Tests made with seven observers using the buzzer tone from the 3–A audiometer gave average values of α for various sensation levels as shown below:

Sensation Level	α = *Difference in Decibels*
29	3.1
46	5.3
64	10.0
81	9.0
99	8.2

It is seen that difference in loudness under these conditions is approximately the same as that corresponding to intersection points in the filter experiments described above. In Fig. 117 the results of both types of tests are given and a smooth curve

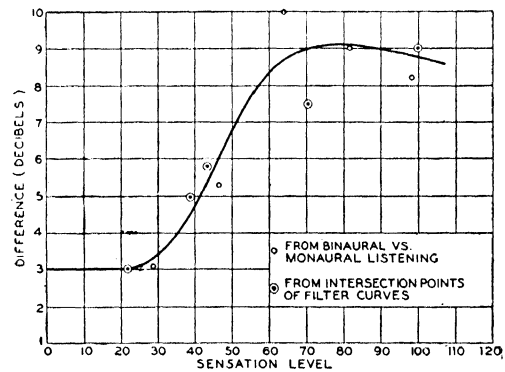


FIGURE 117.

drawn to best fit the data. It must be remembered that such data are very difficult to obtain on account of the wide variation in the judgment of different observers. This accounts for the large departures from the curve. The function proportional to the number of nervous impulses stimulated at the different sensation levels can be deduced from the relation expressed in this curve. Assuming the number of such nervous discharges to be 500,000 per second when the sensation level of sound was 100, then a reduction of 8.8 db reduces this number to 250,000; a further reduction of 9 db reduces it to 125,000; and so on. In this way the curves given in Fig. 118 were constructed.[1]

[1] Since this was written further experimental work has been done using pure tones as the source of sound. Although the results obtained by the author were similar to those given in the table, other observers obtained widely different results. Consequently, conclusions based upon this type of reasoning are seriously open to question. However, such curves as those shown in Fig. 118 are very interesting and important, but until further experimental data are secured they must be considered only as hypothetical.

Formulation of an Empirical Equation for Computing Loudness Losses

If the percentages of the total energy passed by the filters are plotted as a function of the cut-off frequency, curves which intersect at 50 per cent are obtained. Also, the sum of the two ordinates corresponding to any abscissa will be 100 per cent. The curves will be the same irrespective of the absolute value of the total energy. In the case of the experimental curves on loudness, an exponent of the observed percentages can be chosen so that at the intersection point the ordinate raised to this power will be one-half. If the other observed percentages are raised to the same power and a curve is made of the resultant figures, it is found that the sum of the ordinates from the two curves corresponding to any chosen abscissa is approximately unity. The dotted curves shown in the figures

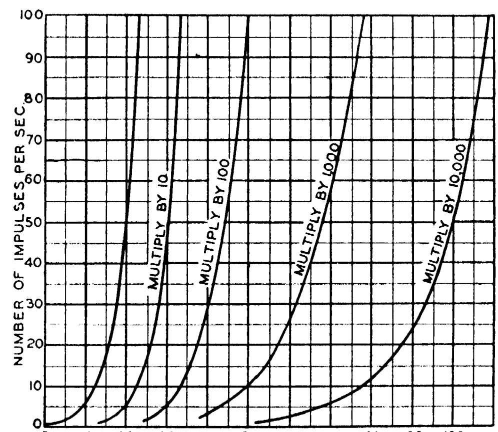


FIG. 118.—RELATION BETWEEN NERVOUS IMPULSES PER SECOND AND SENSATION LEVEL OF SOUND.

mentioned above were obtained in this way. If we desire to adopt the idea that each frequency component in the external sound wave contributes an integral amount to the resultant loudness in such a way that the component parts can be summed to give the resultant loudness, the fractional loss can be empirically represented by an equation of the type

$$y^b = \sum_{k=1}^{k=n} \frac{(W_k E_k)^b}{(E_{k_0})^b}. \tag{3}$$

The summation is taken over all the components. The quantity y is the fractional decrease in the undistorted sound necessary to make it have the same loudness as the distorted sound. E and E_0 are the energies in each component before and after the distortion. The weight factor W and the exponential constant b can be determined from the experimental data.

For no distortion, that is, when $E_k = E_{k_0}$,

$$y^b = 1 = \Sigma_1^n W_k^b. \tag{4}$$

For the filters used in these experiments E_k may be considered zero in the attenuated range and equal to E_{k0} in the unattenuated range of frequencies, so that for the low pass filter

$$y_L^b = \Sigma_1^{k'} W_k^b \tag{5}$$

where k' is the unattenuated component.

Similarly for the high pass filter,

$$y_H^b = \Sigma_{k''}^n W_k^b \tag{6}$$

where k'' is the last unattenuated component.

For any two complementary filters,

$$y_L^b + y_H^b = \Sigma_1^{k'} W_k^b + \Sigma_{k''}{}^n W_k^b = 1. \tag{7}$$

This equation must hold regardless of the weight factor function W_k. This means that the sum of the ordinates for the two curves corresponding to any abscissa must be unity. If this is not true, the empirical equation assumed is not adequate. Also at the intersection point $y_L = y_H$, so that

$$y^b = \tfrac{1}{2} \quad \text{or} \quad b = \log \tfrac{1}{2} / \log y \tag{8}$$

which is sufficient to determine the value b.

The value of y is related to the value of α given in Fig. 118 by the relation $y = 10^{\alpha}$ so that

$$b = \frac{3.01}{\alpha}. \tag{9}$$

An empirical formula similar to equation (3) modified to fit a continuous spectrum such as speech is

$$y^b = \int_0^{\infty} (WE/E_0)^b df. \tag{10}$$

For the filter experiments this reduces to

$$y_L{}^b = \int_0^{m_1} W^b df \quad \text{and} \quad y_H{}^b = \int_{m_2}^{\infty} W^b df \tag{11}$$

where m_1 and m_2 are the cut-off frequencies for low pass and high pass filters, respectively.

It is seen from these equations that the weight factor can be obtained by taking the slope of either of the high or low pass filter curves. Thus for computing the loss of loudness of speech coming from a telephone receiver due to attenuating certain frequency regions, the formula

$$y^{1/3} = \int_0^{\infty} (WE/E_0)^{1/3} df \tag{12}$$

can be used when the sensation level is in the important intensity range used in practice, that is, from 65 to 100 db. The function $W^{1/3}$ is the slope of either of the dotted curves shown in Fig. 116 and y is the fractional energy reduction in the undistorted speech required to make it equal in loudness to the distorted speech.

In this range of intensities a change in loudness is approximately the same as a change in sensation level. In other words, a curve showing loudness as a function of sensation level has a slope of approximately 45° in this range (see Fig.

111) for most sounds including various kinds of distorted speech. Consequently if α is a function which gives the loss in db at each frequency due to the introduction of some piece of apparatus in the circuit, then the loudness loss $\bar{\alpha}$ is given by

$$10^{-\frac{\bar{\alpha}}{30}} = \int_0^\infty G(f) \cdot 10^{-\frac{\alpha}{30}} df \qquad (13)$$

where $G(f)$ is a weighting factor depending upon the type of circuit into which the apparatus is introduced. If the system

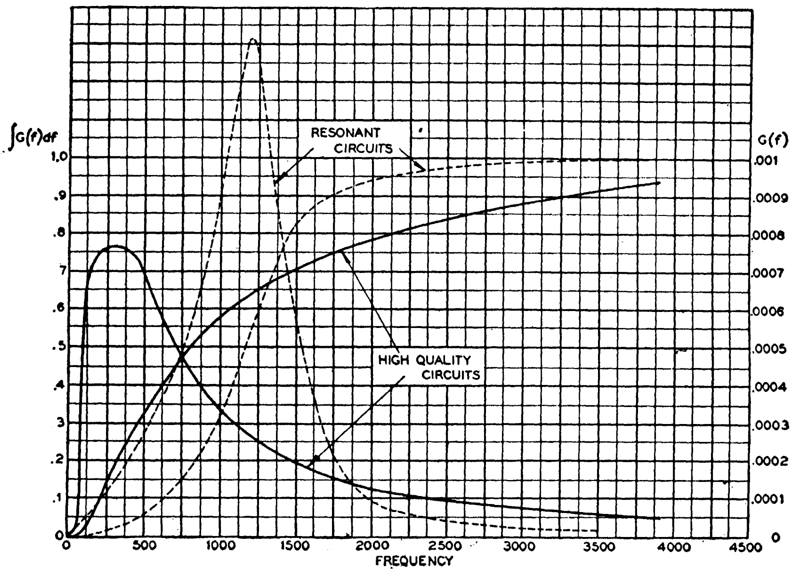


FIG. 119.—CURVES FOR COMPUTING LOUDNESS LOSSES.

reproduces speech perfectly, the function $G(f)$ is the slope of the dotted curve in Fig. 116. A curve similar to this was also obtained for a transmission system having resonant elements such that frequencies near 1200 cycles were reproduced much more efficiently than those either higher or lower than this value. The $G(f)$ functions for these two cases are shown in Fig. 119. The two corresponding curves for $\int_0^f G(f)\, df$ are also given. These aid in the calculation of the above integral, which must be done by graphical methods. For this purpose the variables f and α are changed to x and y by the relations

$$x = \int G(f)df \quad \text{and} \quad y = 10^{-\frac{\alpha}{30}}.$$

Then

$$\bar{\alpha} = -\ 30 \log_{10}\left[\int_0^1 y\, dx\right]. \tag{14}$$

The values of x are obtained from Fig. 119 and the values of y from the loss curves for the particular problem at hand. If these values are plotted on cross-section paper, then the area between the resulting curve, the X-axis, and the ordinates at 0 and 1 gives the value of the required integral.

Comparison of Observed and Calculated Values

In order to test this formula the loss in loudness was computed when different types of resonant networks were intro-

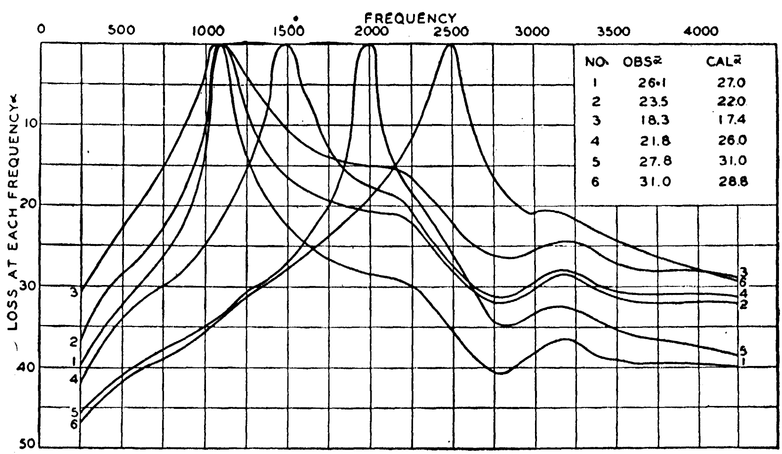


FIG. 120.—LOSS CURVES OF RESONANT SYSTEMS AND EFFECTIVE LOSS IN LOUDNESS OF SPEECH AT LOUDNESS LEVELS BETWEEN 70 AND 100 UNITS.

duced into an otherwise distortionless system. The losses at each frequency for six different resonant systems are shown by the six curves in Fig. 120. The table gives the calculated and observed loudness losses for these systems. The observed values are averages taken by several observers.

When the experimental measurement of the effective loss produced by such resonant networks is made at lower levels, the losses are smaller. Using the weighting factor functions derived from the experimental data for these lower levels taken with the filters, the curve on Fig. 121 was calculated. The averages obtained by several observers are shown by the circles. These data were obtained with the resonant system No. 1, having the response characteristic shown in Fig. 120. It is seen that the loss in sensation units at the low intensity levels is only about one-half that at the higher levels.

To illustrate how these relations can be used in telephone engineering, it is shown in Appendix E that if a condenser of

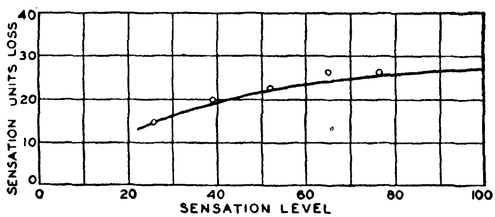


FIG. 121.—EFFECTIVE LOSSES FOR RESONANT SYSTEM NO. 1.

2-microfarad capacity is connected across a long transmission line near its middle, the loudness loss caused to the reproduced speech being transmitted is 9.3 db when the high quality circuit is used and 12.3 db when the resonant circuit is used.

In applying these formulæ for loudness losses it must be remembered that they are empirical in origin and hence are limited to only that class of data from which they were derived and to comparatively small loudness changes. Although they are adequate for most practical purposes, they are not satisfactory from the standpoint of understanding the fundamental elements which determine loudness and how they operate when a loudness judgment is made. Attempts to follow these processes and to develop methods for calculating loudness which have a general application have not yet been successful.

CHAPTER II

The Recognition of the Pitch of Musical Tones

As used in the musical sense, the pitch of a tone is the position on the musical scale to which the tone belongs, the high pitch being high and the low pitch being low on the musical staff. When referring to a single pure tone, the pitch, in this book, has been given a numerical value, namely, that given by equation (6) of Part Three, Chapter III. When the tone has a large number of components, as is usually the case with tones from musical instruments, it still retains the quality of pitch and the position on the musical scale can be determined. The numerical value of the pitch of a complex musical tone is the same as that of a pure tone which is judged to have the same pitch. The pitch of any musical tone can then be determined experimentally by comparing it to a pure tone which can be adjusted so that it seems to have the same pitch. If the device producing the pure tone is calibrated so that its frequency of vibration is known, its pitch is determined from equation (6) mentioned above.

Using this criterion, then, it is important to inquire what types of musical tones will have the same pitch. It is well known that for pure tones the frequency of vibration determines completely the pitch of the tone. However, when there are several components of different frequency in the tone, it is not so obvious which one will determine the pitch.

An experimental investigation was made using the musical tones whose spectra are shown in Figs. 51, 52, 53 and 54. The pitches of these musical tones were determined by comparison with a pure tone. This comparison tone was produced by a telephone receiver which was connected to a vacuum tube oscillator. Its pitch was adjusted to any desired value by making the proper settings on the oscillator. The musical tones were transmitted through the high-quality telephone system, into which were introduced electrical filters. By means of this circuit any portion of the spectrum could be eliminated.

The judgments of pitch and quality of the musical tones were made by three persons familiar with music. The results of the tests with the musical sounds mentioned are given in Table XXVII. In every case they agreed unanimously in the statements recorded in the fifth and sixth columns. The musical tones were produced at sensation levels between 70 and 80 db. In the third column of this table the letter F refers to the fundamental and the numbers refer to the overtones, thus (F & 1–6) means that the fundamental and the first six overtones were eliminated. It is seen that the vowel ah sung at a pitch d is affected only slightly in pitch or quality when the fundamental and the first two overtones are eliminated. Even with the fundamental and the first six overtones eliminated, the pitch still very definitely corresponds to the pitch of a pure tone with the frequency of the fundamental, namely, 145 cycles per second. The harmonic analysis of this filtered tone shows no frequencies below 1000 cycles per second. Eliminating all of the overtones above the sixth changes the quality by about the same amount as eliminating the fundamental and first and second overtones. The data also indicate that if the fundamental and all of the upper and lower harmonics except the third, fourth and fifth, are eliminated, the remaining compound tone has the same pitch as the fundamental, although the quality of the sound is very different from that of the sound ah.

As indicated in the table, similar results were obtained for the vowel ā, sung at the pitch a. Other vowels were tried with similar results. In general, neither the quality nor the pitch of notes from a rich baritone or contralto voice is appreciably affected by eliminating the fundamental and the first two to three overtones. If, however, higher overtones are eliminated, the musical quality (in particular the richness) is

TABLE XXVII

Effect of the Elimination of Various Components on the Pitch and Quality of Various Musical Sounds

Source	Pitch	Eliminated Components	Eliminated Frequencies	Pitch Change	Quality
Voice–ah	d(145)	F	0–250	No change	Inappreciable change
		F & 1–2	0–500	No change	Small change
		F & 1–4	0–750	No change	Large change
		F & 1–7	0–1250	No change	Very large change
		F & 1–9	0–1500	Uncertain	Noise
		6–∞	1000–∞	No change	Small change
		3–∞	500–∞	No change	Large change
		F & 1–2 & 6–∞	0–500 & 1000–∞	No change	Very large change
Voice–ā	a(218)	F	0–250	No change	Slight change
		F & 1–2	0–750	No change	Sounds like ah
		F & 1–4	0–1250	No change	Small change
		F & 1–5	0–1500	No change	Between ah and ō
		6–∞	1500–∞	No change	Between ah and ō
		3–∞	750–∞	No change	Sounds like ō
		F & 1–2 & 9–∞	0–750 & 2000–∞	No change	Very weak ah
Piano	c(129)	F	0–250	No change	Small change
		F & 1–2	0–500	No change	Metallic
		F & 1–4	0–750	No change	Clanging
		For more harmonics eliminated the tone lost all musical character			
		5–∞	750–∞	No change	No brilliance
Piano	c″(517)	F′	0–750	No change	Small change
		F & 1	0–1250	No change	Metallic
		All harmonics	750–∞	No change	Pure tone; Musical brilliance lacking
Violin	g′(388)	F	0–500	No change	Large change
		F & 1	0–1000	No change	Very large change
		F & 1–2	0–1500	Uncertain	Non-musical
		2–∞	1000–∞	No change	Violin quality gone
Clarinet	c′(259)	F	0–500	No change	Large change
		F & 1–2	0–1000	No change	Very large change
		F & 1–4	0–1500	No change	Non-musical
		7–∞	2000–∞	No change	Large change
		2–∞	750–∞	No change	Pure tone (no clarinet quality)
Organ pipe	c(129)	F	0–250	No change	Small change
		F & 1–2	0–500	No change	Large change
		F & 1–4	0–750	Uncertain	Noise
		15–∞	2000–∞	No change	Very small change
		5–∞	750–∞	No change	Small change
Organ pipe	c′(259)	F	0–500	No change	Large change
		F & 1–2	0–1000	Uncertain	Non-musical
		7–∞	2000–∞	No change	Small change
		2–∞	750–∞	No change	Sounds dull

noticeably affected and this is true even though the omitted overtones are all above the fifteenth. The high harmonics do not seem to be so essential for good quality in a soprano voice. An experimental test with only a few voices showed the rather unexpected result that the elimination of all the harmonic frequencies above 2000 cycles affected the musical quality of a bass, a baritone, or a contralto voice to a greater extent than the quality of a high soprano voice.

The table shows that the quality of the principal musical instruments is much more seriously affected by the elimination of the lower parts of their characteristic sound spectra than the quality of the sung vowels by a similar elimination. In any case such eliminations do not change the pitch, for this remains constant as long as the filtered sound can be recognized as a musical tone.

These results were confirmed in a very striking manner by using ten separate vacuum tube generators for producing the component frequencies. These generators were adjusted to give the frequencies 100 cycles to 1000 cycles at intervals of 100 cycles. They were all connected to a special telephone receiver and the currents regulated so that the pressure amplitude of the components of the sound emitted by the receiver were equal. By suitable switching arrangement any one of the components could be eliminated. When they were all impressed upon the receiver a full tone resulted which had a definite pitch corresponding to 100 cycles per second. The elimination of the 100-cycle component produced no noticeable effect. The elimination of any other single component had no effect upon the pitch and almost none upon the tone quality, although by careful listening, its introduction and withdrawal could be detected in most cases. Even with the first seven components eliminated, leaving only 800, 900, and 1000, the pitch corresponded to a frequency of 100. When only two components were left, they were heard as separate tones, the fundamental subjective tone at 100 being still plainly audible but much weaker than either component. Any three consecutive components were sufficient to give the tone a pitch corresponding to 100, as, for example, 200, 300, 400 or 600, 700, 800, etc. When four consecutive components were sounded, the fundamental subjective tone was very prominent. When all of the components were sounded, this fundamental seemed to be louder than the other components and dominated the tone.

The tests just described were made when the sensation level of the 700-cycle tone was at 90 db. When only three components, 700, 800 and 900, were used and the loudness of the combination was greatly decreased, it was found that the 100 cycle subjective tone disappeared when the sensation level of the combination was approximately 45 db. At this level the three tones were heard as separate tones, the 900-cycle one being the last to disappear as the loudness approached zero. When five or more consecutive components were used, the pitch seemed to remain the same for low values of the loudness even down to zero, although for these verv low values, it was very difficult to judge pitch.

If the components 200, 400, 600, 800 and 1000 were used, the pitch corresponded to 200 cycles, i.e., to the octave of the compound tone discussed above. This tone still had the same pitch when the 200- and the 400-cycle components were eliminated. Any two consecutive pairs gave the subjective tone 200, but only very weakly. Combination 300, 600, and 900 gave a harmonious sound, the listener having the tendency to hear the combination as separate musical tones.

From the results which have been described one might conclude that the pitch of a musical tone was determined by the common difference in the frequencies of the harmonics, rather than by the frequency of the lowest component. This conclusion suggested trying a combination of frequencies which are separated by a common difference, but which are not necessarily multiples of this common difference. The combination 100, 300, 500, 700 and 900 was tried and it was found to have no definite pitch, but sounded like a noise. However, one could distinctly hear the subjective tone at 200 cycles. Similarly, the combinations 100, 400, 700, 1000 and 100, 500, 900 and 200, 500, 800 were tried and found to have no definite pitch and to be entirely lacking in musical quality.

Further evidence of the above phenomenon was made possible by means of a carrier telephone system which was available in the laboratory. The technic of carrier telephony makes it possible to displace all the frequencies constituting a compound tone by the same absolute amount upward or downward. Thus, if the compound tone of ten components which has been described were transmitted through such a system when the carrier at the transmitting end differed from the carrier at the receiving end by 30 cycles, the frequencies received would be 130, 230, etc., up to 1030. It is found that such a shift destroys the musical quality which the original tone possessed. If the fundamental is very predominant this shift raises the pitch, but the inharmonic tones produced a harshness and the tone loses its musical character.

Structure of "Harmonic" Tones from Such Wind Instruments as the Bugle

In this connection it is interesting to examine the structure of those tones produced on wind instruments by changes in the blowing intensity rather than by changes in the length of the vibrating air column. When the air pressure blowing an organ pipe or horn is continuously increased, the pitch of the emitted tone corresponds first to the fundamental and then suddenly jumps to that corresponding to the first overtone and then to that corresponding to the second overtone, etc. As is well known, it is this effect that makes it possible to produce the different notes on a bugle. One might expect to find all the harmonics of the fundamental in each of these notes. However, this would be contrary to the observations we have just described. In fact, we find from a few experiments upon organ pipes that the overtones are all present in appreciable amount when and only when the pitch of the tone is that corresponding to the fundamental, but when the pitch corresponds to the first overtone, only those components which are multiples of the first overtone are perceptible. This is clearly shown in the sound spectra given in Fig. 122, obtained by

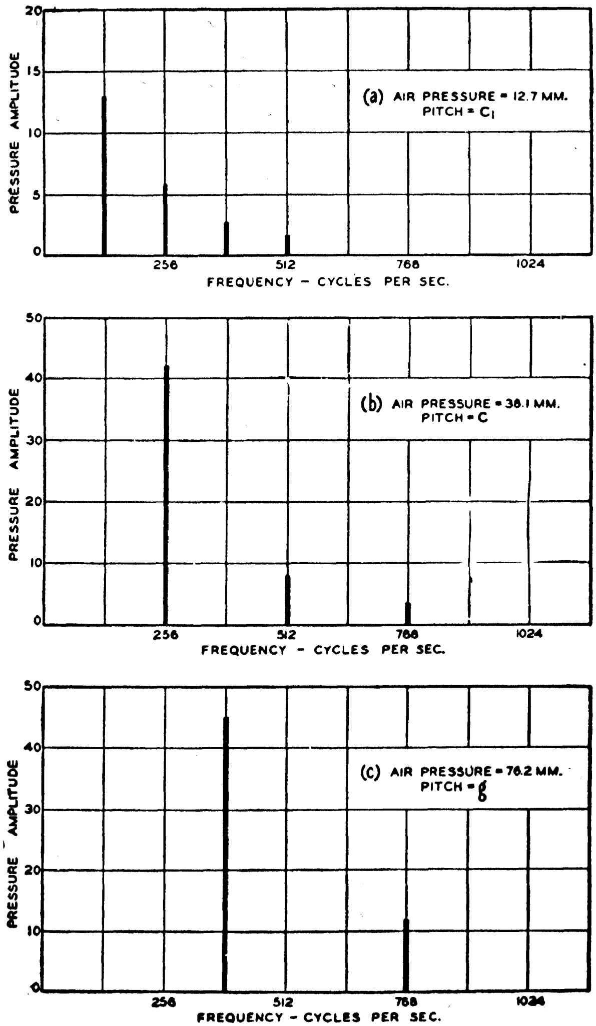


FIG. 122.—SPECTRA FOR ORGAN PIPE.

means of the harmonic analyzer. They represent the sound emitted by an organ pipe when it was blown with the various pressures indicated on the charts.

As shown in Part Three, Chapter I, during the transmission of a pure tone to the inner ear, a number of harmonics are introduced when the impressed tone is loud. Consequently, when such a tone is sensed on the basilar membrane, essentially the same nerve fibres are excited as when the impressed tone is complex. In the large majority of cases, then, there are a number of excited nerve regions on the basilar membrane located at positions corresponding to the harmonics when a musical tone is sensed by the ear. In this respect the loud pure tones produce similar effects to the loud complex tones. The relative intensity of stimulation at these various places varies with the harmonic content of the tone, but the positions remain fixed. It is very probable that the relative positions of these regions of stimulation, due to the harmonics—either objective or subjective—are the real determiners of the pitch. When the complex tones are at levels above 40 db most of the missing tones in the harmonic series are supplied as subjective tones. To illustrate this, the four charts in Fig. 123 have been drawn. The first chart represents the spectrum for the synthetic tone used in the experiments described above. If these components were transmitted through a system which was linear and which discriminated against the frequencies in the same manner as the ear mechanism, the spectrum which would arrive at the oval window would be that shown in chart (*b*). An estimate of the inner ear spectrum which is produced when the non-linearity is present is shown in chart (*c*). This spectrum was estimated by the methods described in Part Three, Chapter IV. In chart (*d*) is shown an estimated inner ear spectrum for the case when the first four components are eliminated from the synthetic 100-cycle tone. From these figures it is evident why this latter tone gives the same pitch and practically the same quality as when all the components are present.

At the very low intensities, however, a spectrum similar to that shown in chart (*b*) must be impressed upon the inner ear. At these low intensities the pitch remains the same even when the five lower components are eliminated. This seems to indicate that even though no stimulation was produced at the position corresponding to 100 on the basilar membrane, the pitch 100 is still recognized, due no doubt to the spacing of the other components at positions corresponding to those which would be produced by a fundamental of 100. It is also evident from these charts that the quality of the musical tone

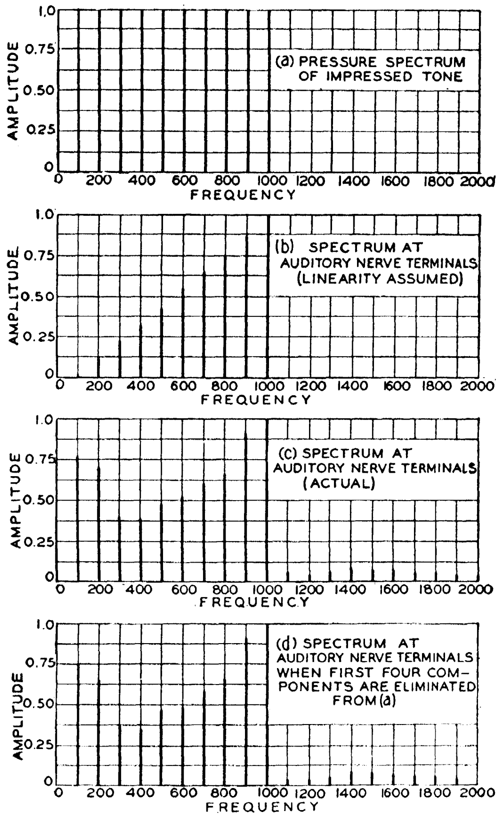


FIGURE 123.

must change as the intensity increases. This is a well-known fact and is important to remember when trying to reproduce music which gives the same effect as the original music.

Before leaving this subject it should be pointed out that the time pattern theory of hearing outlined in this book will also account for these experimental results on pitch and at the low intensities may be the main contributing factor. According to the nerve mechanism set forth in this theory the nervous discharge due to an impressed sinusoidal stimulus will always take place at the same phase of vibration, but not at every vibration. Consequently when the four tones having frequencies of 400, 500, 600, and 700 cycles per second act upon the ear, the impulses in the auditory nerve will be timed somewhat as follows. There will be certain fibres excited by the 400-cycle tone which will be firing every 4th vibration, certain ones excited by the 500-cycle tone firing every 5th vibration, certain ones excited by the 600-cycle tone which will be firing every 6th vibration and certain ones from the 700-cycle tone which will be firing every 7th vibration. These discharges will all unite to form impulses in the auditory nerve having a time interval of .01 second. Similarly a number of combinations will unite to give an impulse at $\frac{1}{2}$, $\frac{1}{3}$, $\frac{1}{4}$, $\frac{1}{5}$, $\frac{1}{6}$, and $\frac{1}{7}$ of this interval. There will be discharges at other time intervals, but the number of fibres causing them will be considerably less than at the particular ones given above. It may be that the recognition of these time intervals by the brain aids in the recognition of pitch.

CHAPTER III

Methods of Measuring the Recognition of Speech Sounds

In order to make a quantitative study of the effect of various kinds of distortion upon the average person's ability to recognize the sounds of speech, it is necessary to have an accurate method of measuring his ability. Many such methods have been proposed and used; they all are based upon the general method of pronouncing speech sounds into one end of a transmission system and having observers write the sounds which they hear at the receiving end. The comparison of the called sounds with those observed shows the number and kind of errors which are made. The differences arise in the technic of making such tests and in the type of speech material used. The system tested may be the simple system used in most conversations; namely, the air between the mouth and the ear in a room, or it may be a telepone system, or it may be reproduced speech from a phonograph. In any case this method gives a quantitative measure of the understandability of the speech which arrives at the ear.

Different types of speech material may be used as the testing material. It must be representative of speech and suitable for making tests. These two aims at times come into conflict and the best compromise must be made. If the fundamental speech sounds are taken as a unit, the per cent of correct letter sounds received is called "letter articulation," from which the use of the terms "vowel articulation," "consonant articulation" and "sound articulation"—the articulation for a particular fundamental sound—becomes obvious. If the syllable is used as a unit, the per cent correctly received is called the syllable articulation. If a sentence is used as a unit and is considered correctly recognized if the main thought is grasped by the observer, the per cent of such correctly received sentences is called "intelligibility."

In general, in this book where "intelligibility" is used, it refers specifically to the results obtained with the test sentences described below.

Several different types of articulation testing lists have been constructed and used. The one which was used for making the tests mentioned in this book is known at present in the Bell Telephone Laboratories as "The Standard[1] Articulation Testing List." Most of the articulation data discussed in this book were obtained with these lists. Only the very simple syllable forms are used, namely, those composed of the form CV (consonant-vowel), VC (vowel-consonant), and CVC (consonant-vowel-consonant). All the possible combinations[2] of the English speech sounds given in Table I were constructed, using these syllable forms. These syllables were arranged into lists of 50 each, such that in each list there were 5 syllables of the form CV, 5 of the form VC, and 40 of the form CVC. Also, each group contained approximately the same number of each of the fundamental speech sounds. To take care of all of the possible syllable combinations required 174 such lists or 8700 syllables. When using such lists in a test any possible combination of the speech sounds may occur so there can be no tendency to memorize the syllables, such as happens if words or peculiar types of syllables are used.

These syllables are written on cards which are shuffled each time before they are used so that the order in which they are pronounced is entirely haphazard. To illustrate the technic of articulation testing, a sample list is given in Table XXVIII. In the first column the syllable is given in its phonetic form. A key word showing how each syllable is pronounced is given in

[1] The word "standard" as used here has no significance except as a name for this particular set of lists.

[2] In the original lists, the sounds th as in (thin) and th as in (then) occurred only one-half as often as the other sounds and oi and ew were omitted.

the second column. The syllables are pronounced into the system at the rate of one every three seconds.

TABLE XXVIII

SPEECH-SOUND TESTING LIST. LIST NO. 160

	Speech sound	Key word		Speech sound	Key word
1	ha	ho(t)	26	gōb	go + b
2	hā	hay	27	shōl	shoal
3	wá	wa(g)	28	ros	rus(t)
4	wi	wi(th)	29	jod	ju(g) + d
5	vou	vow	30	bok	buck
6	ár	air	31	zīk	z + (d)ike
7	ez	e(bb) + z	32	bīch	buy + ch
8	ūsh	you + sh	33	kīth	ki(te) + th
9	an	on	34	gīt	gui(de) + t
10	id	(l)id	35	yif	y + if
11	jouv	jow(l) + v	36	sin	sin
12	moush	mou(nd) + sh	37	térm	term
13	rour	r + our	38	mérl	m + earl
14	zūth	z + (s)oothe	39	pérv	p + (n)erve
15	hūs	who + s	40	yēt	y + eat
16	chush	ch + (p)ush	41	bēl	b + eel
17	jum	j + (f)oo(t) + m	42	zēf	ze(al) + f
18	thup	th + (s)oo(t) + p	43	weng	whe(n) + ng
19	fuch	foo(t) + ch	44	kev	k + ev(er)
20	wóng	wa(ll) + ng	45	háng	hang
21	chóth	cha(lk) + th	46	pág	p + (r)ag
22	tój	ta(ll) + j	47	yās	y + ace
23	kóg	k + aug(er)	48	dāp	d + ape
24	fōn	(tele)phone	49	yang	ya(cht) + ng
25	dōs	dose	50	lan	l + on

In Table XXIX is given a sample of the records made by the observers. This table gives the results obtained when the list was transmitted over a system transmitting only frequencies as high as 1250 cycles per second. The correct word is written opposite the syllables which were recorded incorrectly. The errors for each of the fundamental sounds were then taken from this original sheet and recorded on an analysis sheet as shown in Table XXX. This table gives the average results of eight observers and two callers for the same system. As seen from Table XXX the average of the eight observers

TABLE XXIX

E.S. 269204

INDEXED
~~WORD~~
ARTICULATION
40 %

TRANSMISSION BRANCH
ARTICULATION TEST RECORDING SHEET

TITLE OF TEST 320311
CONDITION TESTED Low Pass Filter - 1250 ~
Attenuation = 5 napiers down

DATE 2-7-20 OBSERVER M. A.
TEST No. 11 CALLER H. E. D.
LIST No. 160

No.	OBSERVED	CALLED	ERRORS	No.	OBSERVED	CALLED	ERRORS
1	tan	térm	ér - a m - n	26	zip	thup	th - z u - i
2	zit	gĭt	g - z ĭ - i	27	kód	tó'j	t-k j - d
3	wa	wa'	a'-a	28	tish	chush	ch-t u-i
4	dāp	✓		29	yang	✓	
5	gōb	✓		30	zēt	zūth	ū - ē th-t
6	yis	yif	f-s	31	ref	ros	o - e s - f
7	māl	merl	ér - ā	32	jum	✓	
8	thin	sin	s-th	33	jó'g	kó'g	k-j
9	zĭp	zĭk	k-p	34	jad	jod	o-a
10	jouv	✓		35	tūth	hūs	h-t s-th
11	yāt	yās	s-t	36	id	✓	
12	thou	vou	v-th	37	ha	✓	
13	bĭp	bĭch	ch-p	38	fōn	✓	
14	ha'ng	✓		39	ko'th	cho'th	ch-k
15	mĭs	moush	ou - ĭ sh - s	40	rour	✓	
16	dăch	dŏs	ŏ - ă s - ch	41	an	✓	
17	kev	✓		42	bok	✓	
18	tig	pa'g	p-t a'-i	43	yēt	✓	
19	kĭs	kĭth	th-s	44	o'r	a'r	a' - o'
20	hā	✓		45	yēth	ūsh	y inserted ū - ē sh - th
21	weng	✓		46	wo'ng	✓	
22	dēl	bēl	b-d	47	kōv	pérv	p-k ér - ō
23	thich	fuch	f-th u-i	48	zēt	zēf	f-t
24	wif	wi	f inserted	49	lan	✓	
25	ez	✓		50	shēl	✓	

and the two callers gives 41.2 per cent for the syllable articulation. The letter articulation was 72.2 per cent. The consonant articulation was 65.8 per cent and the vowel articulation 83.4 per cent.

It has been found that when using these lists for testing systems which have an articulation of the order of 70 per cent, the probable error in the per cent articulation of an observation varies from ± 4 to ± 7 depending upon the crew (observer and caller). An observation used in this sense means the per cent articulation obtained by one crew using fifty of the syllables or one list. It has been a common practice to use ten lists instead of one, which reduces this error to approximately ± 2 per cent. If different callers and observers are used for each of the ten lists, it will be found that the probable error computed from the observer's chart will then be about twice the figure given above, namely, ± 4 per cent. In order to reduce the observational error to approximately 1 per cent, ten persons on the testing team are necessary. If each one of the ten calls to the remaining nine, an equivalent of ninety crews or ninety observations will result. If the average of these is taken, it will have a probable error which is not more than 1 per cent. It is obvious that the observational error computed in this way does not indicate the magnitude of systematic errors such as are due to memory, practice, familiarity with the circuit being tested, etc. The best method of obtaining articulation values which are comparable for several systems is to arrange the testing so that the team has the same practice for each system.

As would be expected, the articulation for a given sound depends somewhat upon its position in the syllable, that is, upon the letter sound preceding and following it. An analysis of results obtained with the standard lists showed that 42 per cent of the errors were initial consonants and 58 per cent were final consonants. The vowel errors for initial and final vowels were the same. Some consonants were recognized more easily after certain vowels than after others. However, the influence of a certain vowel upon the consonants as a class is the same as another vowel, so statistically a letter sound can be regarded

TABLE XXX

Summary Sheet — Average Errors above 2 %.

TRANSMISSION BRANCH

ARTICULATION TEST ANALYSIS SHEET

TITLE OF TEST *320311*

CONDITION TESTED *Low Pass Filter – 1250~* *Average of*

DATE *6-22-20* *Attenuation = 5 napiers down* OBSERVERS *8*

LIST Nos.______ CALLERS *2*

TEST Nos.______

Sounds Called	No. of Times Called	a	ā	á	e	ē	ér	i	ī	o	ō	o	u	ū	ou	% ERROR
		SOUNDS RECORDED AS														
a				10.2												13.1
ā																2.2
á		13.8			4.3				2.2			3.8				27.9
e			2.4	2.3						10.0						17.5
ē																2.9
ér		3.2	38.8	2.7	4.4					3.6	3.8					60.3
i													5.3			8.4
ī																4.2
o		46			24.2											31.9
ō																3.9
ó		5.4								2.2						10.2
u								26.1		4.8						34.6
ū													2.8			11.8
ou																5.6

Total number of sounds called______ Letter Articulation *72.2*

Total number of errors______ Word Articulation *41.2*

Vowel Articulation *83.4*

as independent of the other sounds from a recognition standpoint, provided the following conditions are fulfilled in making up testing lists:

TABLE XXX (Continued)

Summary" Sheet — Average Errors above 3 %.

TRANSMISSION BRANCH

ARTICULATION TEST ANALYSIS SHEET

TITLE OF TEST 320311

CONDITION TESTED Low Pass Filter – 1250 ~ Average of

DATE 6 – 22 – 20 Attenuation = 5 napiers down OBSERVERS 8

LIST Nos.______ CALLERS 2

TEST Nos.______

SOUNDS CALLED	SOUNDS RECORDED AS																						Omissions	Insertions	Error
	b	ch	d	f	g	h	j	k	l	m	n	ng	p	r	s	sh	th	t	v	w	y	z			
b			10.3																				4.5		21.2
ch						4.9		8.7							3.2	7.3	4.3	24.7							56.6
d	7.8				11.9		3.3																		35.2
f						3.3									26.0	11.7	19.0						3.9		63.8
g			30.5				3.9																		38.6
h													4.2												13.5
j			19.3		4.0																				29.8
k													3.8					33.7							48.0
l																									3.9
m											2.9														6.8
n										31.7															35.4
ng																									3.6
p		3.0						24.4										22.2							55.0
r									10.4														5.5		18.5
s				6.2												21.5	24.4	7.0							64.9
sh				3.6											27.3		13.0								53.2
th		3.4		9.6											21.1	8.3			7.0			10.9	3.1		70.7
t		7.3				3.4		19.5					6.2												42.4
v																	10.7					15.1			32.0
w																									1.0
y														9.3											15.8
z			6.2				4.2									6.2	8.8		7.9						39.8

No. of times each sound is called______ Letter Articulation 72.2

Total number of sounds called______

Total number of errors______ Word Articulation 41.2

Consonant Articulation 65.8

1. There shall be no context between sounds; that is, words and sentences in general are excluded.
2. The letter sounds shall occur approximately the same number of times.

3. The majority of possible letter combinations into the simple syllable forms shall be used.
4. The number of initial consonants must be approximately the same as the final consonants.

When these conditions are fulfilled, it is possible to calculate the syllable articulation from the consonant and vowel articulation.

When it is inconvenient to train the observers so that they will both pronounce and also write without error the phonetic syllables, then to determine the vowel and consonant articulation use may be made of simple word lists. The words in these lists must be chosen so that the memory effects are reduced to a minimum. The lists shown in Table XXXI were constructed to fulfill these conditions. In the vowel list the vowel is placed either between the two consonants b and t, or between b and k. With such lists the vowel must be interpreted correctly before the word can be identified. Similarly, in the second list the consonant is either followed by the vowel ī or preceded by wi. This makes it impossible to identify the word unless the consonant is heard. When using these lists the words are written on cards which are shuffled so that the order is different in successive tests. The word is marked right or wrong on the basis of the single letter sound and consequently the word "articulation" in this case becomes "letter articulation."

In research work which requires a large amount of testing, the standard syllabic lists are preferable for two reasons. They enable observers to obtain data at a faster rate, for it is readily seen that in pronouncing 100 syllables from the standard lists, 280 letter sounds are pronounced, while in pronouncing words from the word list only 100 letter sounds are pronounced, which are counted in the test. Therefore, considerably more time is required to obtain the same accuracy with the word lists as with the standard syllabic lists. Also the standard lists place each letter sound adjacent to another letter sound in a manner more nearly like that occurring in ordinary speech than obtained with the word lists.

TABLE XXXI

Vowel Word List (English Words)

Vowel Sound	English Words in the List					
á	bat	bat	back	back	....	4
ā	bait	bait	bake	bake	bake	5
e	bet	bet	beck	beck	....	4
ē	beat	beat	beak	beak	....	4
i	bit	bit	bit	bit	....	4
ī	bite	bite	bite	bike	bike	5
o	but	but	buck	buck	....	4
ó	bought	bought	balk	balk	....	4
ō	boat	boat	boat	boat	....	4
u	book	book	book	book	....	4
ū	boot	boot	boot	boot	....	4
ou	bout	bout	bout	bout	....	4
	Total number of words in list					50

The vowel articulation is the percentage of the vowel sounds correctly perceived.

Consonant Word List (English Words)

Consonant Sound	English Words in the List				
b	by	by	....	2	*Note.*—The h following w is silent in such words as whim, whip, etc. For greater accuracy the list should be doubled.
ch	which	which	....	2	
d	die	die	....	2	
f	fie	fie	whiff	3	
g	guy	wig	....	2	
h	high	high	....	2	
j	....	....	....		
k	wick	wick	....	2	
l	lie	will	....	2	
m	my	whim	....	2	
n	nigh	win	....	2	
ng	wing	wing	....	2	
p	pie	pie	whip	3	
r	wry	wry	....	2	
s	sigh	sigh	sigh	3	
sh	shy	shy	wish	3	
th	thy	thy	with	3	
th	thigh	thigh	thigh	3	
t	tie	tie	wit	3	
v	vie	vie	vie	3	
w	why	why	....	2	
y					
z	wiz	wiz	....	2	
	Total number of words in list			50	

The consonant articulation is the percentage of consonant sounds correctly perceived.

When only a limited amount of testing is done and that with a personnel who are untrained in writing the phonetic sounds, the word lists are very useful. This is particularly true when testing the relative quality of sets for aiding the deafened, for in these instances it is desirable that the observers be deafened persons and consequently they are usually untrained in observing and unfamiliar with phonetic markings.

As stated above, the intelligibility of a system transmitting speech is defined as the per cent of ideas expressed in the form of simple test sentences which, after transmission, are correctly understood by a number of observers. It is probable that the intelligibility is more directly related to the thing which it is desired to measure than is articulation. However, it is a much more difficult quantity to measure. It is evident that due to memory effects, a set of sentences can be used with the same personnel only a very few times. Also, the psychological factors become more prominent when using sentences than when using simple syllables.

However, a set of simple interrogative and imperative sentences was compiled so as to obtain the approximate relation between the articulation as determined by the standard lists and the intelligibility as determined by the simple test sentences. In making up these test sentences the list was designed to test the observer's acuteness of perception and to minimize demands upon his intelligence. The questions are of a self-evident nature, the answers being frequently implied in the questions. They vary in length from about five to twelve or more words, each sentence containing four or five "thought" words. These "thought" words must be correctly received in order to understand the idea of the sentence. Various topics covered by ordinary conversation are represented, including personal experiences and points of interest in politics, science, and commerce. An effort was made to eliminate duplicate ideas. In only a few instances were the ideas repeated and in those cases the manner of expressing them was varied. In this manner forty-nine lists of fifty sentences each were compiled. A sample of one of these lists is shown below.

INTELLIGIBILITY LIST

List I

1. Name a prominent millionaire of the country.
2. How large is the sun compared with the earth?
3. Why are flagpoles surmounted by lightning rods?
4. Give the abbreviations for January and February.
5. Name the tree on which bananas grow.
6. How often does the century plant bloom?
7. What description can you give of the bottom of the ocean?
8. Explain the difference between a hill and a mountain.
9. What is the chief purpose of industrial strikes?
10. Describe the shoes of the native Hollander.
11. Name some uses to which electricity is put.
12. What would cause the air to escape from a bicycle tire?
13. Where is more grain raised, in the East or the West?
14. Tell what is meant by an Indian Reservation.
15. For what invention is Thomas Edison noted?
16. Name a state which has no seacoast.
17. Write the Roman numeral ten.
18. Explain the difference between export and import.
19. Explain why a corked bottle floats.
20. What substance is a good conductor of electricity?
21. Explain why Indians were afraid of firearms.
22. Explain the purpose of fire drills.
23. At what time do ocean waves become dangerous?
24. What medicine would you take to remedy indigestion?
25. What knowledge is covered by the study of astronomy?
26. Name a good restaurant in this vicinity.
27. What is the importance of large windows in stores?
28. Explain why a giraffe eats the foliage of trees.
29. How are the pages of a magazine held together?
30. Explain why the name string-bean is appropriate.
31. Name a nearby city in which there is a shipyard.
32. Name a fruit which grows in bunches.
33. Which of our Presidents went to South Africa?
34. Why are wire springs used in beds?
35. Why are books bound in stiff covers?
36. Why did the home people conserve food during the war?
37. Name an insect that has a hard shell.
38. What symbol on the United States money stands for liberty?
39. What weapons did the Indians use in warfare?
40. In what kind of weather does milk sour?

41. What streets in this city have Dutch names?
42. How does turning a ship's wheel steer the ship?
43. What nation aided us in the Revolutionary War?
44. What are some personal characteristics of the people of Japan?
45. What candy is black and good for colds?
46. Name a famous Indian Tribe.
47. Why is this building lighted by reflected light?
48. Why are most lighthouses situated on rocks?
49. Give some ingredients used in soap.
50. Why is a house built of stone superior to others?

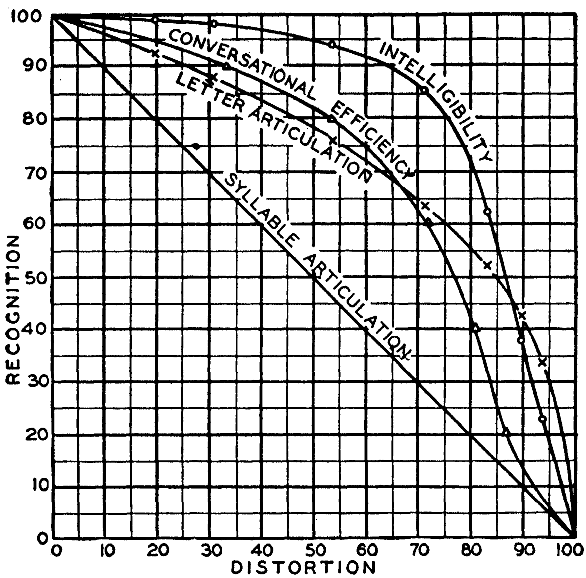


FIG. 124.—RELATION BETWEEN DISTORTION AND RECOGNITION.

In order to obtain an experimental relationship between the intelligibility and the syllable articulation, tests were made on eight transmission systems which gave syllable articulations varying from 5 per cent to 98 per cent. Six observers were used in the work and 1800 sentences were used to determine the intelligibility for each condition. Similarly, 1800 syllables were used to determine the articulation for each condition. The results obtained from these tests are shown in Fig. 124. In this figure distortion is taken as 100 minus the syllable articulation as obtained by the standard lists. This definition is arbitrary, but its choice helps to bring out the relationships existing between the various ways of measuring the recognition of distorted speech. The papers of the observers were corrected not only on the basis of syllables, but also on the basis of the letter sounds so that the letter articulation was obtained. This is also shown in the figure. After a testing crew has had some practice at listening to sentences and syllables, it becomes more efficient in recognizing the speech sounds. However, the data indicated that within the observational error the improvement is such that the points will still remain on the curve; that is, if the articulation shows an improvement of from 5 to 10 per cent the intelligibility will show an improvement of from 20 to 38 per cent as shown by the curve. It will be seen that for distortions greater than 80 per cent a change of 10 per cent distortion is equivalent to a change of approximately 40 per cent in the intelligibility, while an equal distortion change for distortions below 20 per cent corresponds to less than 1 per cent change in the intelligibility. For this reason these test sentences are useful for testing systems having very large distortions but are of little value for testing ordinary transmission systems.

In order to obtain a notion of how the time of transmitting an idea correctly over the system will vary with various amounts of distortion, the caller asked a question and the observer was instructed to reply orally. A record of the time was kept from the instant the caller began his sentence until a satisfactory reply had been received. Both caller and observer had been previously instructed to carry on a conversation over the system as they ordinarily would. The observer could ask the caller to have the sentence repeated, reworded, or any of its difficult words spelled until the caller was satisfied that his question had been understood. The ratio of the time required on the high quality system to that required on any other system is called for convenience the conversational efficiency of the latter system. Figure 125 shows the type of results thus obtained. It is seen that a system which is to transmit correctly at least 9 out of 10 test sentences must have at least 36 per cent syllable articulation. Such a system will have a conversational efficiency of 70 per cent, that is, only 70 per cent as many test sentences can be transmitted over such a system and be correctly understood in a given time as would be possible with an ideal system.

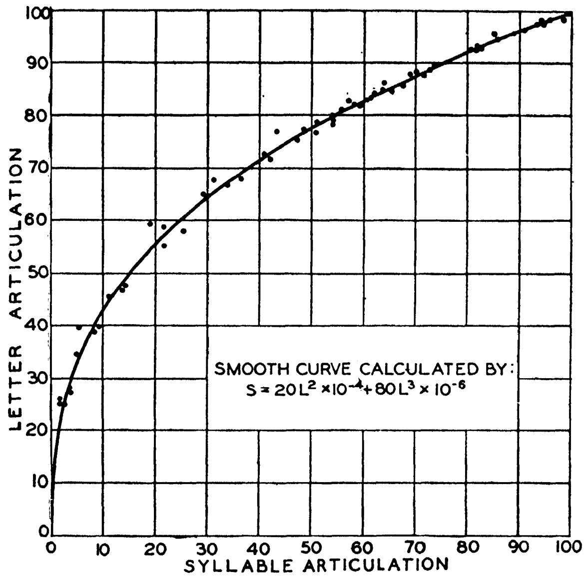


FIG. 125—RELATION BETWEEN LETTER AND SYLLABLE ARTICULATION.

Articulation and the Theory of Probability

Since the articulation is the number of successful trials out of 100 attempts at guessing the correct letter or syllable, it can be considered as a probability. If C, V, L, and S are the consonant, the vowel, the letter, and the syllable articulations, then $\frac{C}{100}$, $\frac{V}{100}$, $\frac{L}{100}$, and $\frac{S}{100}$ give the chance that a consonant, a vowel, a letter sound, or a syllable will be recorded correctly.

In the standard lists the chance of recording a syllable of the form consonant-vowel or vowel-consonant correctly is obviously then $CV \times 10^{-4}$. Similarly, the chance of recording correctly the syllable of the form consonant-vowel-consonant is $CVC \times 10^{-6}$. Since there are 20 of the first and 80 of the second type of syllables in each 100 syllables, in the standard lists the number of correctly recorded syllables when a list of 100 is pronounced is $20\,CV \times 10^{-4} + 80\,CVC \times 10^{-6}$ or the syllable articulation S_0 is given by

$$S_0 = 20\ CV \times 10^{-4} + 80\ CVC \times 10^{-6} \qquad (1)$$

If we are interested in the letter articulation L, then L can be substituted for both C and V and the above formula becomes

$$S_0 = 2L^2 \times 10^{-3} + 8L^3 \times 10^{-5} \qquad (2)$$

The validity of this procedure is dependent upon the conditions outlined on page 249 being fulfilled. For most of the lists of meaningless syllables which have been proposed for use it is justified. For a list which is made up so as to have syllable forms occurring with the same frequency as in written speech a different formula will relate L and S.

These relations have been confirmed by a large amount of experimental data. In Fig. 125 is shown a curve between L and S_0 as calculated from equation (2). The dots represent the experimental results. In these results the letter articulation was obtained from the same data as the syllable articulation, obtained with standard lists as explained in the beginning of this chapter.

[1] This relation was first pointed out by J. Q. Stewart.

CHAPTER IV

Effect of Changes in the Received Intensity of Speech Sounds upon Their Recognition

In order to study the effect of changes in intensity upon the recognition of speech it is necessary to obtain the speech sounds at varying degrees of intensity. One means of doing this is to vary the distance between the speaker and the listener. There are various objections, however, to using this method. As pointed out in Part Three, Chapter VI, it is impossible to secure the range of intensities required by varying the distance because of the large distance required. It would require distances larger than 1000 feet to reduce the intensity of the average voice to the threshold of audibility. Under such conditions, it is very difficult to control the interfering noises and also the reflections which produce distortion.

Consequently, a telephone system was constructed which reproduced speech with practically no distortion. It was arranged so that by means of distortionless attenuators, the intensity of the reproduced sounds could be varied through a very wide range. A schematic of this telephone system is shown in Fig. 126. As indicated, its essential elements are a condenser transmitter to receive the speech waves and transform them into the electrical form, an amplifier for magnifying the intensity of the electrical speech currents, an attenuator for controlling the intensity, an equalizing network, and a receiver for delivering the speech to the ear. The attenuator consists of a system of electrical resistances arranged so that the amplitude of the speech waves could be reduced in steps to as low as one-millionth of their maximum values. The equalizing network is an arrangement of resistances, condensers, and inductance coils, having a frequency selectivity which is the complement of that of the rest of the system. In other words, its introduction into the system compensated for the lack of even response in other parts of the system, particularly in the receiver.

Articulation tests were made with the standard lists described in Chapter III, using this high quality system when it was set to deliver various intensities from the threshold of audibility to very high values. In these tests two callers and eight observers were used and five lists called by each caller. Consequently, each of the points represents the results obtained from the records of 8000 syllables. The results are shown

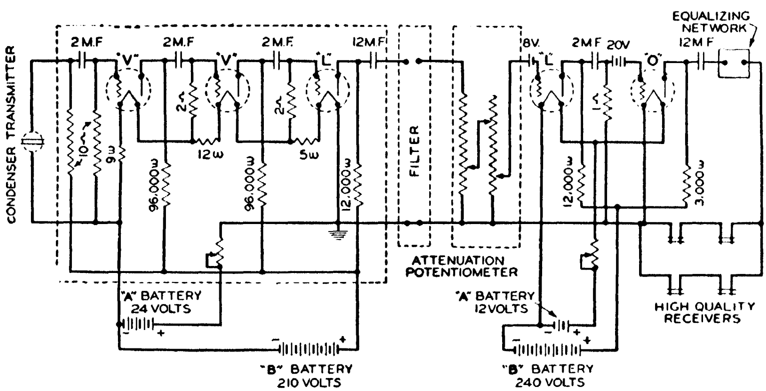


FIG. 126.—HIGH QUALITY TELEPHONE SYSTEM.

by the curve in Fig. 127. The differently shaped points correspond to data taken with different teams having different training. These data were taken at times separated by intervals as much as a year. The abscissas give the intensity level of average speech. As shown in Part Three, Chapter III, the zero on this scale is the intensity level existing at the ear when a speaker talks with average conversational intensity with his lips $\frac{1}{2}$ inch away from the ear. It is also the intensity level corresponding to a flow of speech energy in a free wave of one microwatt per square centimeter. As the various speech sounds are pronounced during a conversation, their intensities fluctuate about this average level. It is seen that the curve strikes the intensity axis at an intensity level of − 100 db. This point also corresponds to the threshold of audibility for average speech. This threshold is determined by the loudest

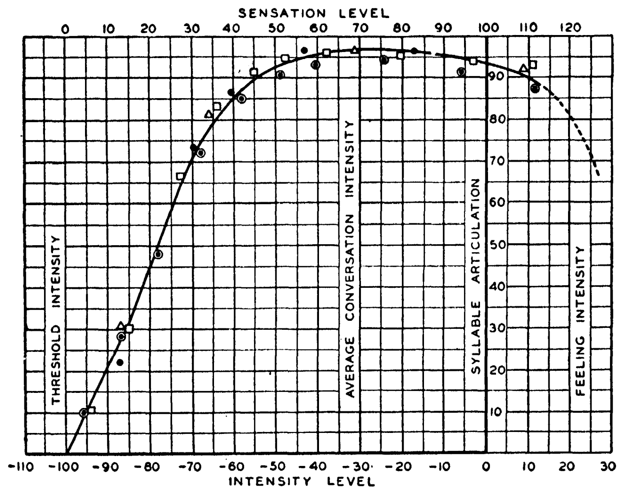


FIG. 127.—ARTICULATION VS. INTENSITY OF RECEIVED SPEECH.

vowel sound, namely, the sound ó (awl). For convenience, a a scale of sensation level is also given at the top of Fig. 127.

It is seen that the intensity level for best interpretation is at − 30 or at a sensation level of 70 db. As the sensation level increases from 0 to 50 db, the articulation rises rapidly from 0 to values greater than 90 per cent. From 50 db to 70 db there is only a slight increase and from 70 db to 120 db there is a slight decrease. It is interesting to note that for a range of intensity levels of more than 70 db, the articulation changes by less than 10 per cent. These results hold only for the condition when no noise is present either in the room or in the telephone system. It is important to notice that in the range where the speech becomes hard to interpret lowering the intensity level 1 db reduces the articulation about $2\frac{1}{2}$ per cent.

When reproduced speech is decreased to an intensity value near the threshold of audibility, the loudest components are in the frequency range from 700 to 1500 cycles (see Fig. 82). For this frequency range the threshold intensity level α_0 is -93. The reason why speech can be reduced to an intensity level of -100 before it becomes inaudible is because the loudest speech sounds have a level which is 6 to 10 db greater than the average level.

The articulation data were analyzed to determine the articulation for each of the fundamental sounds. The curves shown in Figs. 128, 129, 130, 131, 132, and 133 give the results for each of the fundamental sounds. The abscissas are given in terms of the intensity level for average speech rather than for the speech sound itself. The threshold intensity level for each of the speech sounds as determined by the method described in

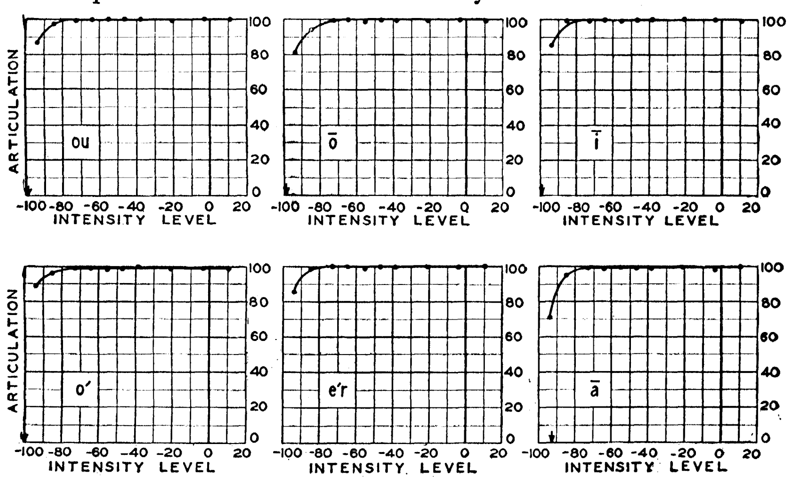


FIG. 128.—ARTICULATION VS. INTENSITY LEVEL.

Part One, Chapter III, is indicated by the small arrow. It was from these articulation data that the figures given in the last column of Table VIII were calculated. The sounds in each of the groups have similar characteristics from a recognition standpoint.

In Table XXXII are shown the results of the articulation tests for the range of intensities usually used in conversation.

As a check against these results obtained with the high-quality telephone circuit, tests were made with the observers stationed

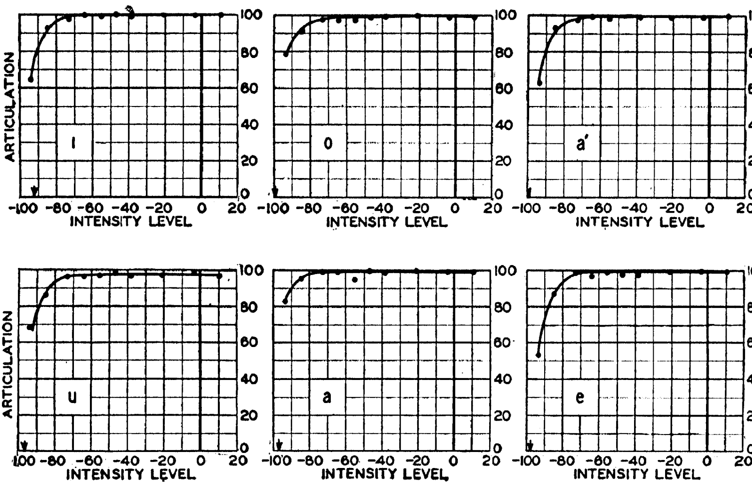


FIG. 129.—ARTICULATION VS. INTENSITY LEVEL.

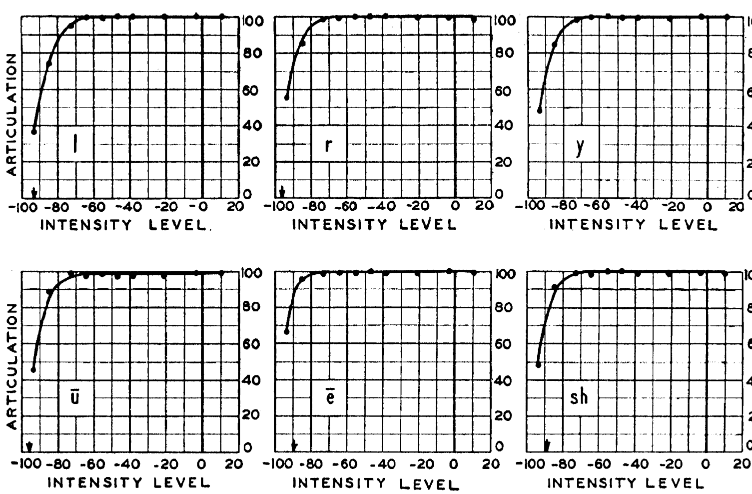


FIG. 130.—ARTICULATION VS. INTENSITY LEVEL.

at 3 feet away from the speaker, thus permitting the speech sounds to be transmitted through the air. These tests checked the results given in Table XXXII within the observational error. Instead of the articulation, 100 minus the articulation

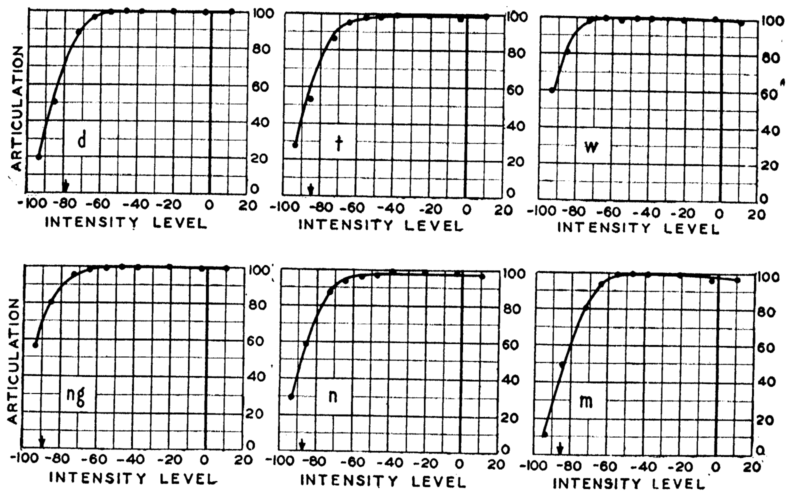


FIG. 131.—ARTICULATION VS. INTENSITY LEVEL.

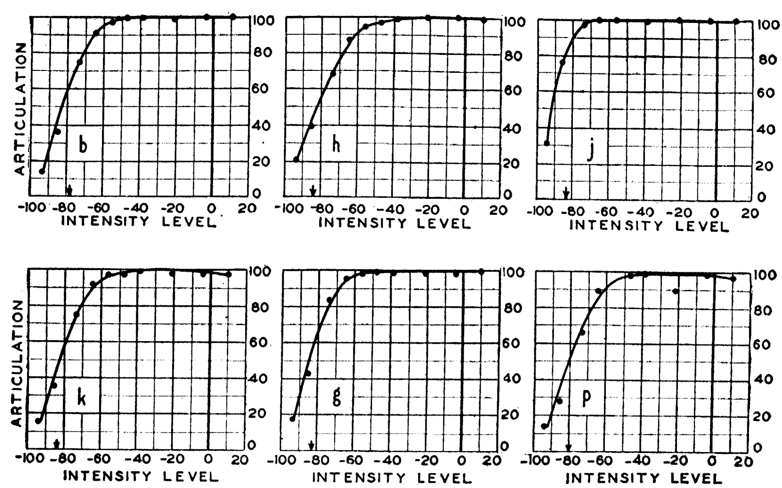


FIG. 132.—ARTICULATION VS. INTENSITY LEVEL.

or the articulation error is given. The speech sounds are arranged according to the magnitude of the articulation error and, consequently, according to the relative difficulty of recognizing them. It will be noticed that the consonants are usually harder to recognize correctly than the vowels. However, the speech sounds e and l, r, ng form notable exceptions to this rule since the former is among the most difficult, while the latter are among the very easiest speech sounds to recognize at normal intensities. At all intensities, the sounds th, f, and v are the most difficult to recognize. The sound z, which is readily recognized at normal intensities, becomes very difficult at weak intensities. The sounds i, ou, er, and ó are missed

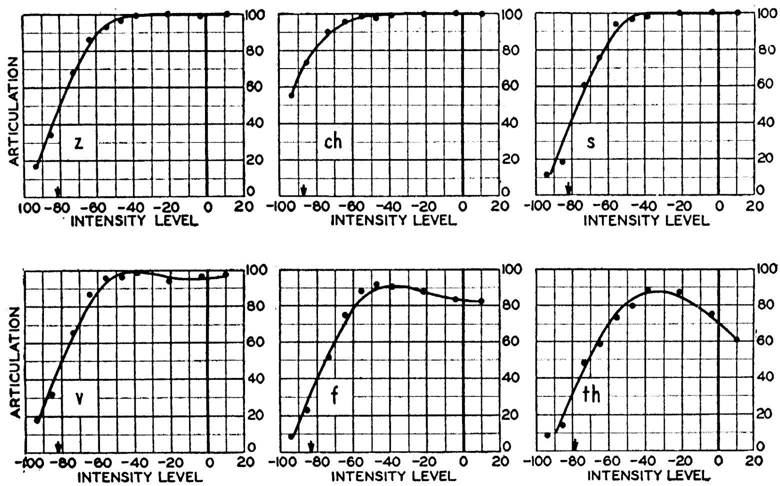


FIG. 133.—ARTICULATION VS. INTENSITY LEVEL.

less than 10 per cent of the time even when very near the threshold value for average speech. It is seen from Table XXXII that for intensities commonly used in conversation the sounds v, f, and th count for more than half of the mistakes in the recognition of the fundamental speech sounds.

There is a characteristic difference between the shape of the curves for the vowels and for the consonants. For the former the curves run along horizontally and then drop off very abruptly. For the latter the drop is more gradual. The threshold points marked on the axis of abscissas do not correspond in general to zero articulation. A consonant sound may sometimes be identified by the modification produced on the following or preceding vowel even though it is below the threshold as determined by an isolated sound. It might seem logical to consider this modification of the vowel as part of the consonant. If it is so considered, then it is evident that as long as the vowel is heard there is always a chance of identifying

TABLE XXXII

ARTICULATION ERROR OR THE PER CENT OF TIMES THE SOUND IS MISINTERPRETED

Speech Sound	Key Word	Average	Speech Sound	Key Word	Average
l	look	0.2	g	gold	1.0
ī	time	0.2	a	top	1.1
ou	town	0.3	b	ball	1.1
ng	sing	0.3	n	no	1.1
r	red	0.3	ē	team	1.2
z	zest	0.3	h	hat	1.2
ér	term	0.4	sh	ship	1.5
y	you	0.4	á	tap	1.5
ō	tone	0.5	ch	cheap	1.8
d	day	0.5	s	say	1.8
i	tip	0.6	k	keep	1.8
t	ten	0.6	ū	tool	2.2
m	man	0.7	u	took	2.5
j	jump	0.8	p	pay	2.5
o	ton	0.8	e	ten	2.8
ó	talk	0.8	v	view	3.9
w	we	0.9	f	fall	12.7
ā	take	1.0	th	then	17.3

the consonant preceding or succeeding it, and consequently the threshold of a consonant so considered will be the same as that for the vowel. It is for this reason that all of the curves seem to go through the same zero articulation point. For example, it is seen that, for the sounds in Figs. 133 and 134, the articulation is still above zero when the characteristic part of the consonant is 10 or 15 db below the threshold for the isolated sound. The vowels should have zero articulation points corresponding to their respective threshold values. The curves do not extend far enough to verify this fact. The only vowel curve which does not seem to satisfy this condition is that for the sound ē. However, Table X shows that the average phonetic power of this sound is about one-third that of the sound ó and consequently its threshold value should be only about 5 db different from that of ó instead of 10 db. The former value agrees with the articulation curve.

CHAPTER V

Effect of Frequency Distortion upon the Recognition of Speech Sounds

When pure tones having equal intensities are produced successively in front of the transmitter a transmission system is said to have frequency distortion if unequal intensities are produced at the receiving end. It is evident, however, that if this lack of uniform response to different frequencies exists for those frequencies only which are either below or above the hearing range, no distortion will be noticeable to the ear.

To determine the importance of the various frequencies for carrying the properties which determine the recognition of sound, the following experimental tests were performed. By means of the high-quality telephone system illustrated in Fig. 126, speech sounds were converted into electrical waves. These electrical waves were sent through electrical filters which had the property of transmitting only certain frequency ranges. One type of filter known as the "low pass" filter transmits only frequencies below a certain limit, a schematic diagram of the circuit arrangement to produce this effect being shown in Fig. 134. The other type of filter known as the "high pass" filter passes only those frequencies above a certain limit, the circuit arrangement for producing this effect being shown in Fig. 135. By means of the arrangement of coils and condensers shown in these two figures, the amplitudes of those frequencies outside of the band which we desire to transmit are reduced to less than $\frac{1}{1000}$ of their normal value while those in the band are only slightly changed.[1] By means of a switch-

[1] For a complete theory of this action of electrical filters, see Chapter XVI of book by K. S. Johnson entitled "Transmission Circuits for Telephone Communication."

ing mechanism various values can be given to the coils and condensers so as to make the limiting frequency at any desired value.

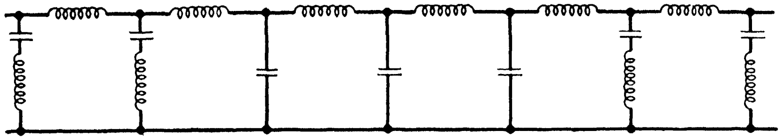


FIG. 134.—LOW PASS FILTER.

Articulation tests were made, using various filter combinations. The solid curves in Fig. 136 show the results of these tests. The ordinates give the syllable articulation and the

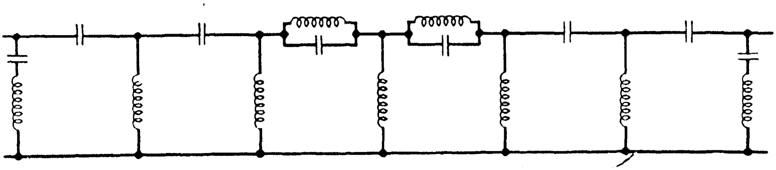


FIG. 135.—HIGH PASS FILTER.

abscissas give the cut-off or limiting frequency of the filter. For example, on the curve labelled "Articulation L" the point (1000,40) means that a system which transmits only frequen-

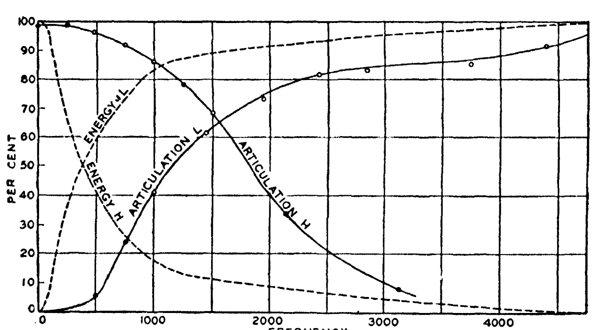


FIG. 136.—EFFECT UPON THE ARTICULATION AND ENERGY OF SPEECH OF ELIMINATING CERTAIN FREQUENCY REGIONS.

cies below 1000 cycles per second has a syllable articulation of 40 per cent. Similarly, on the curve labelled "Articulation

H" the point (1000,86) means that a system which transmits only frequencies above 1000 cycles per second has a syllable articulation of 86 per cent. The dotted curves in this figure show the per cent of the total speech energy which is transmitted through such filter systems. These curves were derived from the data given in Part One, Chapter III. In obtaining these results the intensity of the received speech was adjusted by means of an attenuator in the circuit so that a maximum articulation was obtained for each filter system.

It will be seen that although the fundamental chord tones with their first few harmonics carry a large portion of the speech energy and are important from the standpoint of the naturalness of the reproduced speech, they carry practically none of the properties which determine the correctness with which the speech sounds are understood. A filter system which eliminates all frequencies below 500 cycles per second eliminates 60 per cent of the energy in speech, but only reduces the articulation 2 per cent. A system which eliminates frequencies above 1500 cycles per second eliminates only 10 per cent of the speech energy, but reduces the articulation 35 per cent. A system which eliminates all frequencies above 3000 cycles per second has as low a value for the articulation as one which eliminates all frequencies below 1000 cycles per second. This last statement may appear rather astonishing since it is contrary to the popular notion of the relative importance of various voice frequencies. In this connection it should be pointed out that the articulation is not a measure of the naturalness of the reproduced speech. Although a system transmitting only those frequencies above 1000 cycles will give an articulation of 85 per cent, a degree of recognition that for many purposes might be satisfactory, the speech reproduced by it will sound very peculiar, its naturalness being destroyed. As indicated, the elimination of frequencies below 500 cycles produces only a small effect upon the articulation but it produces a much larger effect upon the naturalness.

The two solid curves intersect on the 1550 cycle abscissa and at 65 per cent articulation, which shows that using only frequencies above or frequencies below 1550 cycles an articulation of 65 per cent will be obtained. The two dotted curves necessarily intersect at 50 per cent.

The data were analyzed to find the effect upon each of the fundamental sounds. The results of this analysis are shown in Figs. 137, 138, 139, 140, 141, and 142. These curves

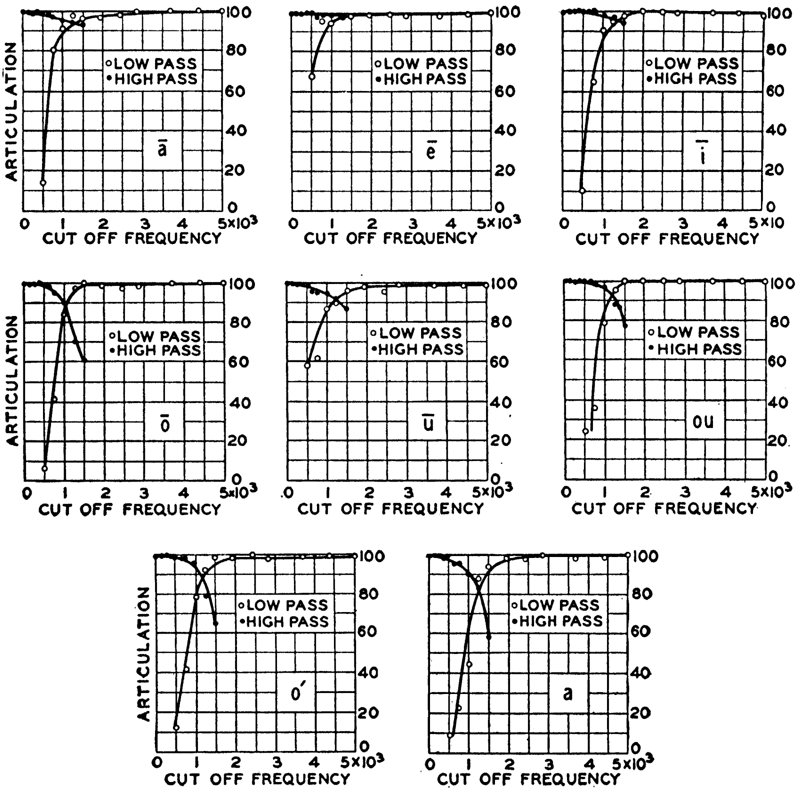


Fig. 137.—Long Vowels.

indicate that some of the sounds are fairly well localized in a limited frequency range while others seem to have characteristics extending throughout the entire range. For example, the sound ē could be recognized correctly 98 per cent of the time when either the range of frequencies above 1700 cycles or the range of those below 1700 cycles was used. On the other hand, the sound "s" was only slightly affected by eliminating frequencies below 1500 cycles but its characteristics were practically destroyed by eliminating frequencies above 4000 cycles. The short vowels, u, o, and e, are seen to have important char-

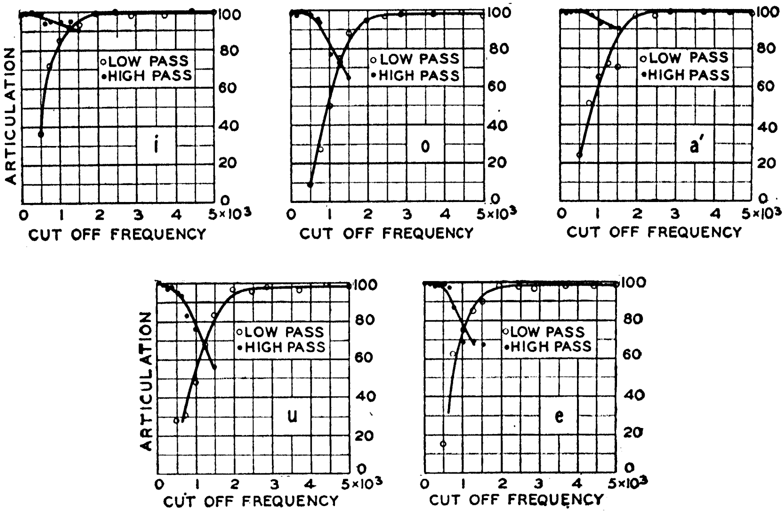


Fig. 138.—Short Vowels.

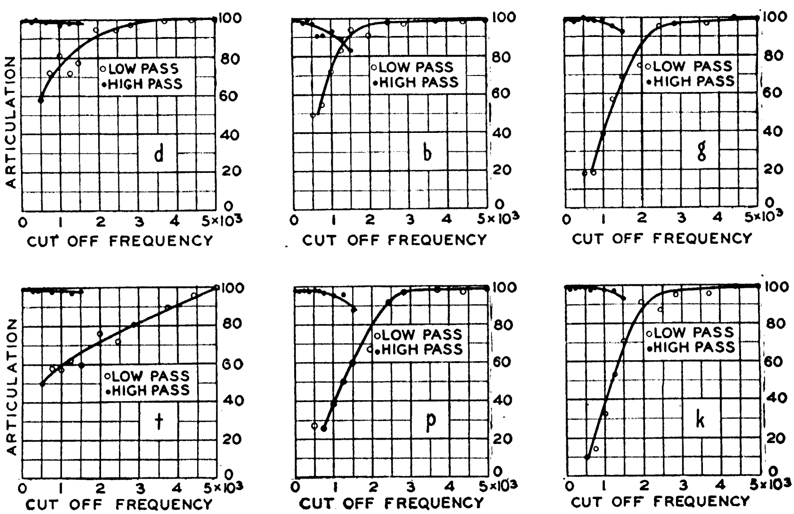


Fig. 139.—Stops.

acteristics carried by frequencies below 1000. More than a 20 per cent error is made in recognizing these three sounds when the frequency components below 1000 are eliminated. On the

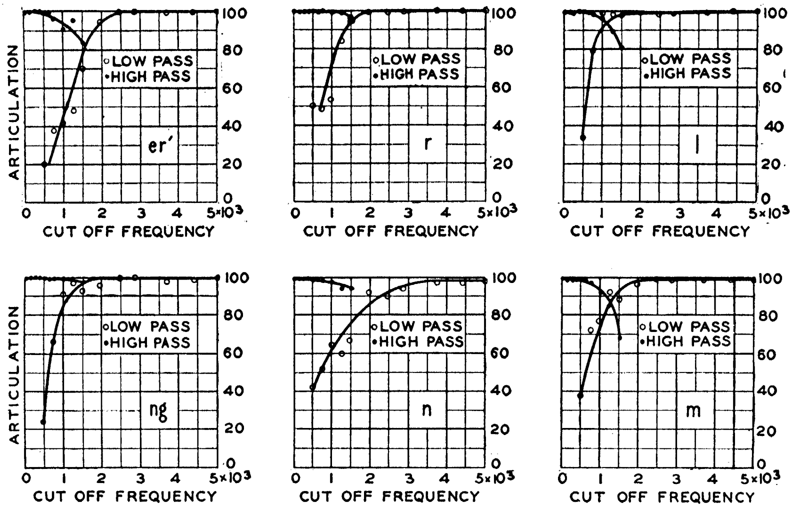


FIG. 140.—SEMI-VOWELS.

other hand the elimination of frequencies above 2000 cycles for these sounds produces only slight effects. The long vowels and the diphthong sounds seem to have sufficient distinguishing characteristics in either half of the frequency range to be iden-

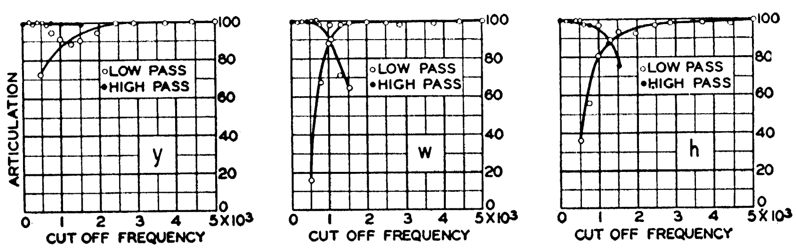


FIG. 141.—TRANSITIONALS.

tified. The intersection point of the curve for these sounds is always above 90 per cent, showing that by using a frequency range on either side of the intersection point, the sounds can be readily identified. The fricative sounds are seriously affected by the elimination of the high frequencies. The elimination of frequencies above 3000 reduces the articulation of the sound "s" to 40 per cent, the sound "th" to 66 per cent, the sound "z" to 80 per cent, the sound "t" to 81 per cent, and the sound "f" to 85 per cent. All other sounds are reduced less

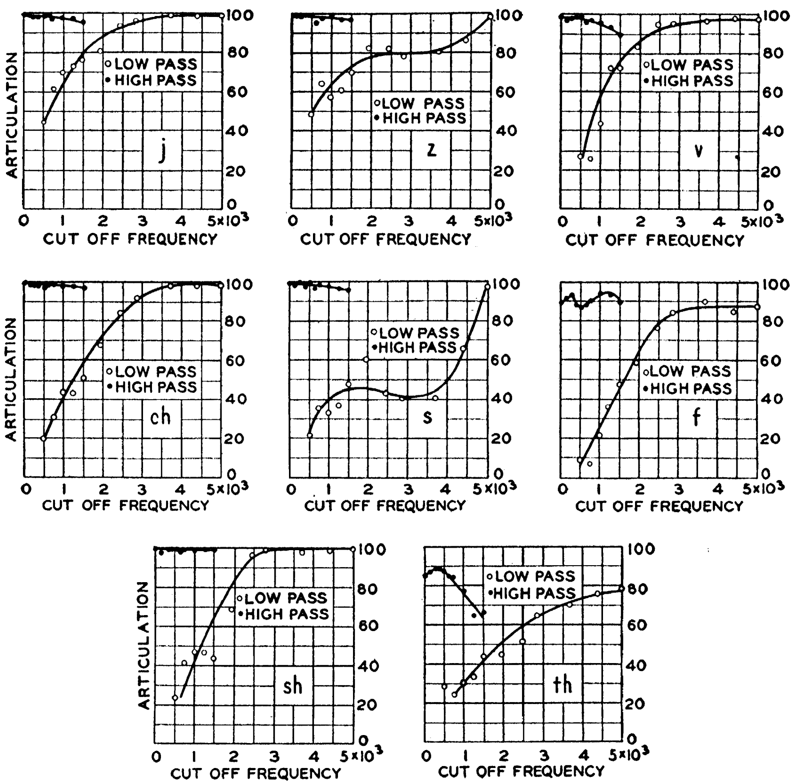


FIG. 142.—FRICATIVES.

than 10 per cent by the elimination of this frequency range. The pure vowels, the diphthongs, and the semi-vowels are affected only a negligible amount by the elimination of this region. The curves indicate that for the unvoiced stop consonants the frequencies in the region of 1000 and 3000 cycles are the important ones for carrying the recognition properties.

The sound "t" has a noticeable characteristic, namely, that the elimination of all sounds below 1500 cycles produces no noticeable effect upon its recognition. It is also the first one of this group to be affected by the elimination of the high frequencies. The transitional sounds w and h seem to have important characteristics in the frequency region between 700 and 2000 cycles. The sound y has characteristics similar to the sounds ī and ē. When all frequencies below 1500 cycles are eliminated it still has an articulation of 99 per cent.

It must be remembered that in spite of the fact that high articulation values are obtained for these sounds under the distorting conditions mentioned, the quality of the sound is materially altered, that is, the naturalness is considerably reduced. However, there are some characteristics of the sound which seem to be sufficient to identify it in spite of its greatly altered quality. For example, ē sounds very much like ū when frequencies above 1000 cycles are eliminated, but from the fact that the per cent articulation at this point is over 90 per cent it is evident that some features are still preserved in the low-frequency region for the sound ē that distinguish it from the sound ū.

Tests made with women calling vs. men calling over these filtering systems indicated that the fricative sounds formed by women's voices require very much higher frequency ranges to properly transmit them than those produced by men. The elimination of frequencies above 4400 cycles reduces the articulation for women's voices for the sound "s" to 68, for the sound "th" to 56, and for the sound "f" to 88, while for men's voices these values are 95, 75, and 97, respectively.

In most systems proposed for use certain frequency regions are not entirely eliminated as in the filter systems but only suppressed various amounts. In order to increase the efficiency of the system from a loudness standpoint, resonance is frequently introduced. This is accomplished by adjusting the mass reactions and elastic constraints in the mechanical parts or the inductances and capacities in the electrical parts so that they annul each other at certain frequencies and thus permit large amplitudes for small driving forces. This results in magnifying the amplitudes in certain frequency regions above those in other regions.

The effect of such resonance on the ability of the system to properly transmit speech was investigated by means of the circuit shown in Fig. 126. An electrical network consisting of an inductance coil and a condenser connected in parallel was bridged across the circuit. By adjusting the inductance and capacity of these to the proper values, the desired resonant characteristic was given to the system. For illustration, the

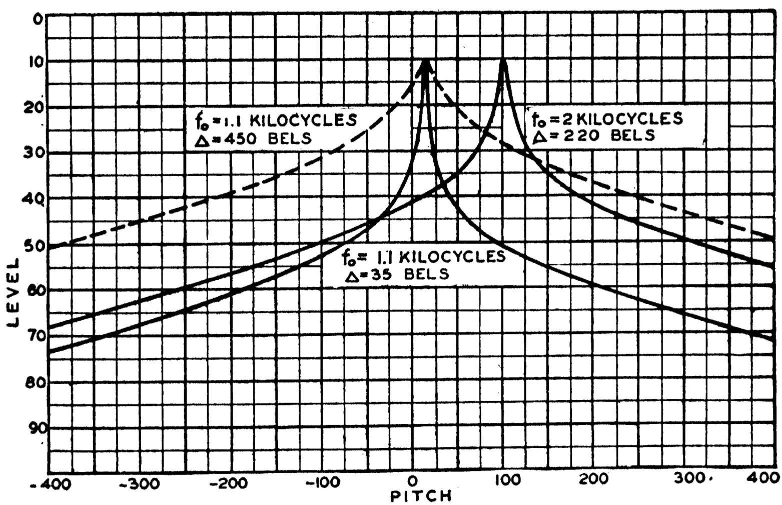


FIG. 143.—FREQUENCY CHARACTERISTICS OF RESONANT SYSTEMS.

characteristics of three of the systems used in the articulation tests are given in Fig. 143. System No. 1 has a resonant frequency at 1100 cycles and a damping constant of 450 bels per second; system No. 2 a resonant frequency at 1100 cycles and a damping constant of 35 bels per second; system No. 3 a resonant frequency at 2000 cycles and a damping constant of 220 bels per second. The ordinates represent the intensity level at which pure tones having the pitch represented by the abscissas will emerge from the receiver when they are created with a zero intensity level at the transmitter. Since they represent the necessary amplification to make the reproduced tones equal in intensity to the original tones, it is convenient to speak of these ordinates as representing the number of db below unity reproduction. Curves are drawn for the case when the attenuators in the amplifiers are set so that the resonant frequencies in each case are at the same level, namely, 10 db below unity reproduction. The frequency characteristics of singly-resonant systems such as these are represented by curves which are symmetrical about the resonant frequency when the coordinates such as shown are used, namely, pitch for the abscissas and db below unity reproduction for the ordinates. (See Appendix F.)

System No. 2 was the most sharply resonant of any of the systems used in the test, yet it would be called very highly damped compared to a system like a tuning fork or a piano string. It was seen in Part Three, Chapter VI, that all of the tuning forks used for clinical purposes have damping constants smaller than 3 db per second. System No. 2, then, damped out its natural frequency of vibration one hundred times faster than the most highly damped tuning fork. Even so, it produces a serious impairment to the speech. However, this impairment is mainly due to the unequal response for different frequencies rather than to "hangover effects" due to transients.

A series of articulation tests were made with systems having resonant frequencies at 1100 cycles but with various amounts of damping. These tests indicated that the articulation decreased from 96 per cent to 90 per cent, as the damping of the system was changed from infinity, that is, corresponding to the high quality circuit, to a damping of 130 bels per second. The system whose characteristic is shown in Fig. 143 by the dotted line had an articulation of 92 per cent. It is rather surprising to find that such large departures from uniform response can exist in a transmission system without producing more serious impairment to the speech which is transmitted over it. As the damping of the transmission system becomes smaller than 130 bels per second, the articulation obtained decreases at a more rapid rate, being about 80 per cent when the damping is reduced to about 35 bels per second.

Tests which were made with a series of transmission systems in which the damping was kept approximately constant, while the resonant frequency varied, indicated that when the resonant frequency varied only between 900 cycles per second and 2000 cycles per second there was only a small change in the articulation which was obtained with the system. When the resonant frequency was outside of this range, the articulation decreased. It should be emphasized here that for producing systems of highest quality such resonances within the voice range should be entirely avoided.

Articulation tests were made with these resonant systems throughout a wide range of intensities. The articulation varied with intensity in a manner very similar to that shown in Fig. 127 for a high-quality system. It was found that at the high intensities the resonant system produced a greater relative impairment than that produced by the high-quality system. This is due to the fact that the frequencies near the resonant frequency are so loud as to become very annoying before the other frequencies are sufficiently loud to indicate the characteristics of the speech sounds. This is particularly noticeable when a set designed for aiding the deafened has pronounced resonant peaks. A user complains of having his ear "banged" by certain vowel sounds before the loudness is sufficient to hear the consonants.

CHAPTER VI

Effect of Other Types of Distortion upon the Recognition of Speech Sounds

Speech waves are distorted in a great many different ways besides those mentioned in the last two chapters. One way which is frequently encountered in amateur radio receiving sets is to "overload" the vacuum tubes. When the input speech energy is greater than that which the set is designed to handle, the output speech waves are distorted. Measurements upon telephone systems containing vacuum tube amplifiers show that considerable distortion of this type can be tolerated before any appreciable loss in articulation is produced. The results of one such series of measurements are shown by the curves in Fig. 144. The lower curves show the relation between the input and output levels for the speech as judged by listening tests. As seen from these curves, at an input level of about 20 decibels the output vacuum tube reaches its capacity. For higher levels the output speech energy is no longer proportional to the input. It is seen, however, that the articulation decreases only from 79 to 77 as the input level increases 15 decibels above its overload point as indicated by the curve. For higher levels, the articulation drops off rapidly. One can readily notice the effect of overloading by the presence of a peculiar high hissing sound even before the articulation is noticeably affected. These tests were made with a telephone system containing resonant elements which accounts for the articulation being below 80 per cent in the range of levels where no overloading takes place. Under such conditions there are two opposing factors operating, one tending to increase and one tending to decrease the articulation. On account of the nonlinearity the weak consonant sounds are made stronger in comparison with the vowel sounds which would tend to increase the articulation. On the other hand the introduction of component frequencies not in original sounds tends to decrease the articulation. It is probably due to these opposing factors that the articulation changes so little when large distortions of this type are produced. Under certain conditions the articulation of a circuit will very definitely increase when a non-linear element is introduced itno the system. This condition is obtained when the first factor predominates over the second one. A few circuits fulfilling this condition have been tested in the laboratory.

The effect of overloading upon the transmission of music is very much more marked than upon the transmission of speech. Tests made with music indicated that when the input was increased more than 5 decibels above the overload point, that is, for input levels higher than − 15 on the scale shown in Fig. 144, the reproduced music was noticeably affected. Tests were made for both vocal and instrumental music by a number of observers listening to the quality of the reproduced music. This type of distortion is very common in radio receiving sets, being due to either poor operation or poor design of the set.

Another type of distortion which is interesting is that produced by a phonograph when the speed of the turntable during reproduction is different from that used when recording. The component frequencies in such reproduced sounds can be obtained from those of the recorded sounds by multiplying each by a common factor. For example, if the speed is twice that of the normal, then the frequency of each component in the reproduced sound is doubled. There are two effects of such distortion which are readily noticed; namely, the pitch of the speech sounds is raised and the syllables are spoken very much more rapidly. The effect upon the articulation is not so obvious, but since the characteristic frequencies of the various speech sounds are shifted, it becomes more difficult to properly understand them.

To investigate this effect, the standard articulation lists were recorded on disc records. This was done in cooperation with The Victor Talking Machine Company. By means of an electromagnetic reproducer the sounds were converted into electrical energy and sent into the high quality telephone system described above, and the syllables were then observed in the usual manner. Speeds of rotation from about ½ to 1½ the normal were tried. The results of these tests are shown in Fig. 145. Changes of speed less than 10 per cent produce very little effect. For greater changes the articulation falls off rapidly. Decreasing the speed has a greater effect than

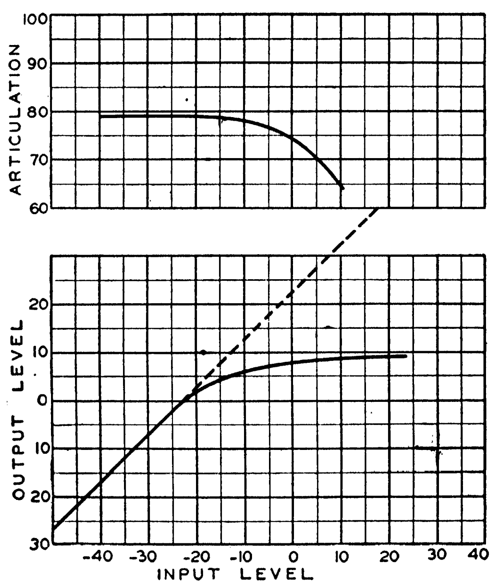


FIG. 144.—EFFECT OF OVERLOADING UPON THE ARTICULATION OF SPEECH.

increasing the speed. In all the types of distortion discussed above the harmonic relationship between the fundamental and overtones is maintained.

In another type of distortion which is peculiar to carrier telephone systems this relationship is not maintained. When the frequency of the carrier introduced at the receiving end differs from the frequency of the carrier at the sending end, the frequencies in the speech spectrum are shifted by a definite amount. Each component frequency is either increased or decreased by the same number of cycles. For example, if this shift is 50 cycles, a vowel pronounced at a pitch corresponding

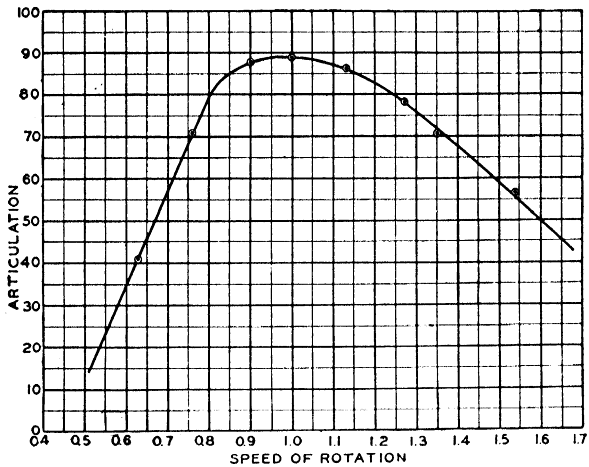


FIG. 145.—EFFECT UPON ARTICULATION OF MULTIPLYING COMPONENT FREQUENCIES BY A COMMON FACTOR.

to 100 cycles would be reproduced by such a system with the components at 150, 250, 350, 450, etc. As will be seen, the harmonic relationship is destroyed.

Articulation tests with such a system gave the results which are shown in Fig. 146. It is seen that the shifts greater than 10 cycles produce a noticeable effect, but it is remarkable that shifts as high as 300 or 400 cycles are possible without completely destroying the intelligibility. As will be seen from the curves of Figs. 125 and 146, an average person would interpret short sentences correctly over 90 per cent of the time, even when all the component frequencies are shifted upward 400 cycles. The data indicate that shifting the speech frequency downward produces a more serious deterioration than by producing the same shift upward. This type of distortion has a much more serious effect on music than on speech. Since musical tones are rich in harmonics, it is evident that a shift which changes the harmonic relationship will produce a serious quality damage to musical tones.

Another common type of distortion is that produced in rooms which are reverberant. In such rooms there are two causes which are operating to decrease the articulation of the

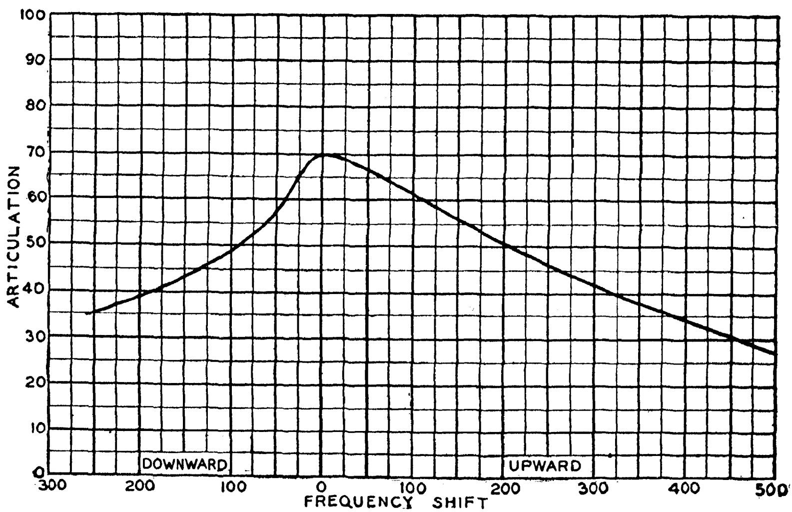


FIG. 146.—EFFECT OF FREQUENCY SHIFT UPON THE ARTICULATION OF SPEECH.

received speech. The first one is due to the persistence of the sound after the source has been silenced. This phenomenon is frequently referred to as the "hangover" effect. It not only distorts the speech sounds by making the endings and beginnings different, but it tends to mask the succeeding speech sounds. Particularly, the vowel sounds tend to mask the succeeding consonants. The second effect is similar to the frequency distortions discussed in Chapter V. Due to the selective absorption properties of the room some frequencies are reproduced much more efficiently than others. Both of these effects are dependent upon the position of the observers in the room.

It is a common practice to describe the acoustic properties of a room in terms of its reverberation time. The reverberation time is the number of seconds after the source is shut off before the sound has decreased its intensity 6 bels. For example, a room having a reverberation time of one second has a damping of 6 bels per second. In the past, very little attention has been given to the source of sound in making reverberation time measurements. Frequently the reverberation time is given without stating whether it is for speech, pure tones or for some other sound. This condition has probably arisen because of the difficulty of making such measurements. The observational error of the common reverberation test is so large that it is hard to distinguish between different classes of sounds. To more completely describe the acoustic properties of the room, a curve should be given which

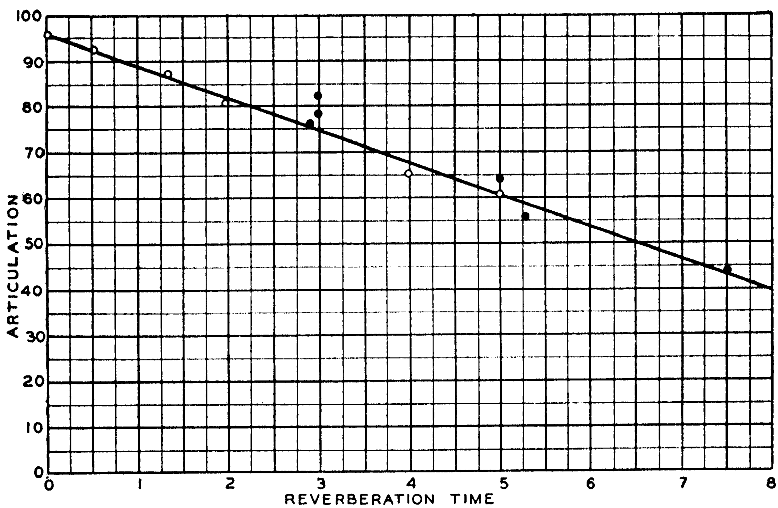


FIG. 147.—ARTICULATION VS. REVERBERATION TIME.

shows the reverberation time or damping for each frequency. Very frequently an auditorium is poor because of the lack of damping for the low frequencies only. Acoustic treatment which produces an absorption for the high frequencies in such a room would make it worse rather than improve it.

Knudsen has made some articulation measurements in rooms of various reverberation time, but gives only one value for this time for each room; presumably that which corresponds to speech. One set of tests was carried out in a room whose reverberation time was changed by bringing into it various amounts of hair felt. The blank dots shown in Fig. 147 give the results of these tests. Knudsen also made tests in auditoriums having different acoustic properties both before and after acoustic treatment. These results are shown by the solid dots in Fig. 147. The tests were made with the standard articulation lists by using three observers who were stationed at different parts of the room. These results indicate that it is a good approximation to say that the articulation decreases about 7 per cent for each additional second in reverberation time. However, the amount of articulation reduction will undoubtedly depend upon whether the high frequencies or the low frequencies are damped out most quickly. There is no one-to-one relationship between articulation and reverberation time but the above results indicate the general effect in large rooms.

CHAPTER VII

Effect of Noise and Deafness upon the Recognition of Speech Sounds

The effect of noise upon the ability to hear is very similar to the effect of partial deafness. As stated in the chapter on "Noise," one of the best ways of describing a noise is to give its deafening effect. For this reason the effect of noise and the effect of deafness upon the ability to recognize speech are considered together in this chapter.

When a noise is present at the ear, the threshold for hearing other sounds is shifted. The noise audiogram gives the amount of this shift. It is the deafening effect. The effect of the noise upon the ability to recognize sounds has four aspects. First, the average threshold shift produces an effect equivalent to the lowering of the intensity of the speech sounds. Second, the unequal threshold shift for different frequency ranges produces a distortion effect equivalent to that produced by a transmission system having unequal responses for different frequency ranges. Third, for the higher intensities there is an intermodulation between the noise and speech which takes place during their transmission through the middle ear. Fourth, the presence of the noise tends to distract the attention from the perception of the sounds.

Although the presence of a noise may be annoying, if one concentrates his attention on the speech sounds he can identify them when they have sufficient intensity to be clearly above the noise. This is true only when the intensity of the speech is not greater than 70 or 80 decibels above the threshold. When the noise is so great that intensities larger than this must be used, the ability to recognize them will always be somewhat less than in a quiet place due to the intermodulation effect.

The curve representing the relation between intensity and articulation of speech sounds received in the presence of a noise whose audiogram is flat, can be obtained from the curve of Fig. 127 by shifting those points on the curve below a sensation level of 70 decibels an amount equal to that shown by the noise audiogram. A system of curves obtained in this way is shown in Fig. 148. The abscissas give the sensation level and the intensity level of the speech sounds and the ordinates the

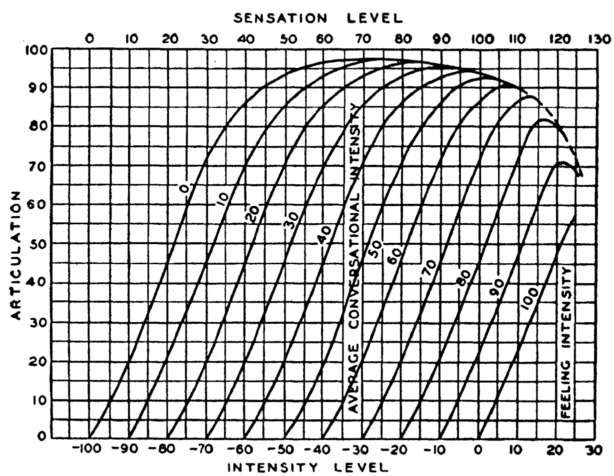


FIG. 148.—ARTICULATION VS. INTENSITY OF RECEIVED SPEECH IN THE PRESENCE OF NOISE.

articulation which is obtained under the different noise conditions. The number on each curve gives the threshold shift produced by the noise as indicated by a noise audiogram. Experiments have shown that at the higher intensities no values of articulation will be obtained which are greater than those obtained in a quiet place. Consequently, the curves at these intensities have been estimated.

It is difficult to produce a noise which has a perfectly flat audiogram. In Figs. 149 and 150 are shown some articulation curves taken in the presence of noise. The noise audiogram corresponding to each case is shown in the figures. When the interfering noise is composed essentially of the high frequen-

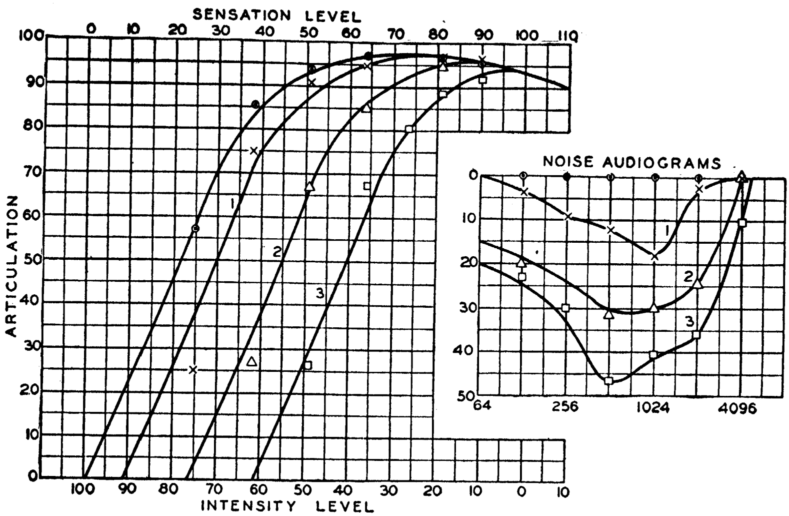


FIG. 149.—EFFECT OF NOISE UPON ARTICULATION.

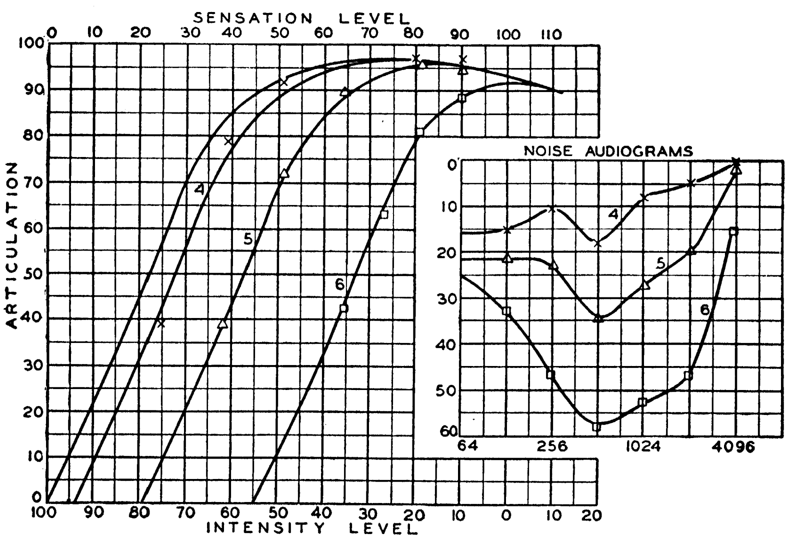


FIG. 150.—EFFECT OF NOISE UPON ARTICULATION.

cies, thus producing a noise audiogram whose threshold has shifts mainly in this region, the articulation curves cannot be obtained by such a simple procedure as given above. For example, in Fig. 151 are shown the curves representing the articulation obtained in the presence of pure tones. The noise for curve No. 7 is a 2000-cycle pure tone at a sensation level of 78 and curve No. 8 is for a 1000-cycle tone having the same sensation level. It is seen that the curves No. 7 and No. 8 cannot be obtained from the "no-noise" curve by

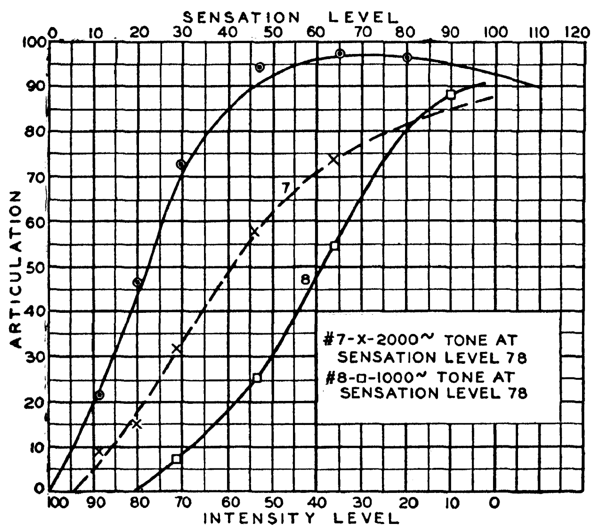


FIG. 151.—EFFECT OF NOISE UPON ARTICULATION.

making horizontal shifts as was the case for the first six curves given.

Experiments were made using pure tones as the interfering sounds. When the frequency is below 500 cycles the resulting intensity level-articulation curves are of the first class mentioned above; that is, the principal effect upon the curve is to produce a horizontal shift. When the frequency is above 500 cycles, the shift is always considerably more for the high than for the low intensity levels as illustrated in Fig. 151. As a general rule since the low-pitch tones produce a greater masking for the frequencies which are important for recognizing speech, especially when they are very intense, they also cause a greater reduction in articulation than the high tones of equal sensation level. Also noises of any complex character produce, in general, a greater reduction in articulation than that produced by the pure tones of the same sensation level since the threshold shift which they produce covers a much wider frequency range.

As an illustration of how these data might be used to answer practical questions, suppose it is desired to find the interference to conversation in the presence of typewriter noise corresponding to the audiogram of Fig. 60. Throughout the important speech range the threshold shift is about 45 decibels. In ordinary conversation in a room the sensation level is about 70 decibels. According to the curves of Fig. 149 the articulation corresponding to this condition is 60 per cent. Under such articulation conditions, the persons conversing would probably raise their voices about 10 decibels in intensity level, thus increasing the articulation to 80 per cent. In a similar way, one can find the average interpretation under any specified noise condition whose corresponding audiogram can be considered as approximately flat.

What has been said about noise can also be said about deafness. When the audiogram representing the degree of hearing can be represented by an approximately flat curve, the main effect upon the intensity articulation curve is to produce a horizontal shift. The articulation to be expected can be found in the manner described in the last paragraph. For example, if the noise audiogram mentioned above represented the degree of deafness for a person rather than a noise condition, then that person would obtain an articulation of approximately 60 per cent when people were conversing with normal intensity and placed about three feet apart. The intensity level of speech from the speaker in a large auditorium such as a theatre or a church is usually between 50 and 60 decibels.

Under such circumstances the deafened person mentioned above would have serious difficulty understanding anything at all. It must be remembered that these figures apply to the average person and that there will be considerable variations from this average, depending upon the intellect of the deafened person. With the speech levels emerging from a telephone receiver in practise, unless a person's hearing is reduced more than 30 decibels below normal, he should have little difficulty in understanding most conversations over the telephone. It is necessary to maintain this high speech level over the telephone because of the room noise which is present at most subscribers' stations and also on account of noises which are picked up by the line.

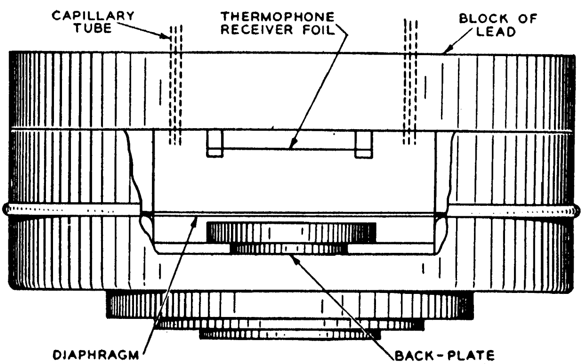
For certain types of deafness there is added to this shift effect a distortion effect. For such cases the articulation is always lower than that calculated by the above method. Persons having nerve deafness, which is indicated by the audiogram dropping rapidly at the high frequencies, have a greater difficulty in trying to understand speech than those whose audiograms are approximately flat. Some persons having this type of deafness are unable to interpret speech regardless of the level at which it is introduced into the ear.

In this connection it is interesting to note that those sounds which are most difficult to hear and interpret such as f, th, s, etc., are among the easiest sounds to interpret by noting the position of the lips. On the other hand the vowel sounds which are easier to hear are difficult to interpret from the lip positions. For this reason, hearing and lip-reading materially aid each other under those conditions where it is difficult to make the proper interpretations. Deafened persons realize this and can recognize very much better those speech sounds which they hear when at the same time they are able to clearly see the lip positions.

APPENDICES

APPENDIX A

A DRAWING showing the arrangement for calibrating a condenser transmitter by means of a thermophone is given in Fig. 1. Hydrogen gas is sent into one of the capillary tubes shown and out of the other until the enclosed chamber is filled with this gas. Capillary tubes are used for this purpose so that from an acoustic standpoint the chamber may be considered completely closed. The measurements are made with hydrogen gas so that the wavelength will be as large as possible compared to the size of the chamber. This is necessary, for in



APPENDIX A. FIG. 1.—ARRANGEMENT FOR CALIBRATING ELECTROSTATIC TRANSMITTER.

the development of the formula given below it was assumed that variations of pressure at different positions within the chamber are all in phase. When the wavelength is comparable to the size of the chamber, standing wave patterns are set up. Under such conditions this formula does not hold.

A direct current I_0 is sent through the gold foil strip, heating it to a temperature θ_0. The final temperature which it assumes depends upon I_0 and also upon the heat capacity of the foil, the properties of the gas, and the size of the chamber. It can be calculated from the change in resistance of the foil. An alternating current $I_1 \cos \omega t$ is then superimposed upon the direct current. It causes fluctuations in the temperature of the gold strip and also in the gas immediately surrounding it. This in turn causes fluctuations in pressure which depend upon the rate at which the heat is conducted and radiated away from the gas layer next to the strip. The following formula was deduced by E. C. Wente of Bell Telephone Laboratories, and its validity has been demonstrated experimentally. The pressure variation p is given by

$$p = \frac{.478 R I_0 I_1 \cos(\omega t - \Phi) M}{\left[\left(GQ - UF - \frac{Aa\omega}{\alpha}\right)^2 + \left(FQ + UG - \frac{Aa\omega}{\alpha}\right)^2\right]^{1/2}}$$

R = resistance of thermophone strip;
I_0 = direct current;
I_1 = amplitude of alternating current;
ω = 2π times the frequency of variation;
Φ = phase between pressure and current;
M = correction factor which is nearly unity and is given by

$$M = \left(1 - \frac{S}{V_0\alpha} + \frac{S^2}{2V_0^2\alpha^2}\right)^{1/2}\left(1 - \frac{2}{\alpha' l} + \frac{2}{(\alpha' l)^2}\right)^{1/2}$$

S = area of the walls of the chamber:
V_0 = volume of the chamber;
α is related to the heat diffusibility and is given by

$$\alpha = \sqrt{\frac{\rho_0 \omega C_p}{2K}}$$

ρ_0 = density of the gas;
C_p = specific heat of the gas at constant pressure;
K = the heat conductivity of the gas;
α' is the same as α except that the constants involved refer to the thermophone strip.

$$G = \frac{V_0 \alpha T_a}{2a p_0}\left[1 - \frac{k - 1}{k}\frac{\theta_0}{T_a}\left(1 - \frac{2a}{\alpha V_0}\right)\right]$$

$$Q = 2a\alpha K + 8aC\theta_0^3$$
$$U = a\gamma\omega + 2a\alpha K$$
$$F = \frac{V_0\alpha T_a}{2ap_0}\left(1 - \frac{k-1}{k}\frac{\theta_0}{T_a}\right)$$

k = ratio of the specific heats of the gas at constant volume and at constant pressure $= \frac{c_p}{c_v}$.
A = the reciprocal of the mechanical equivalent of heat;
a = area of one side of the thermophone strip;
θ_0 = temperature of the thermophone strip due to I_0;
γ = heat capacity per unit area of the strip;
T_a = average temperature of gas within the enclosure;
p_0 = average pressure within the enclosure;
C = Planc's radiation constant.

The phase θ is determined by $\tan\theta = \dfrac{FQ + GU - \dfrac{Aa\omega}{a}}{GQ - FU - \dfrac{Aa\omega}{a}}$.

For frequencies above 50 cycles the effect of radiation can be neglected and the above formula is simplified thus:

$$p = \frac{.478 I_0 M}{\sqrt[N]{1 + (B + D)^2}} IR$$

where p and I are either maximum or effective values and

$$N = \frac{\gamma\omega V_0\alpha}{2p_0}\left(T_a - \frac{k-1}{k}\theta_0\right)$$

$$B = 1 + \frac{4K\alpha}{\gamma\omega}$$

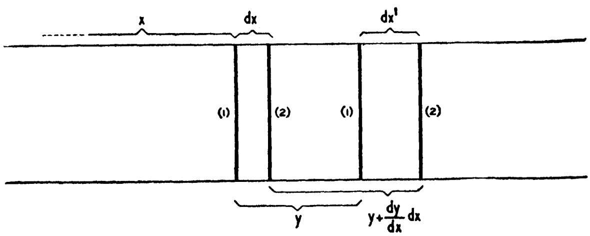
$$D = \frac{2a}{V_0\alpha}\,\frac{\frac{k-1}{k}\theta_0}{T_a - \frac{k-1}{k}\theta_0}.$$

Neglecting the radiation effect causes a 2 per cent error at 32 cycles. The correction factor M varies from .7 at 32 cycles to .98 at 1000 cycles.

APPENDIX B

Derivations of Equations Giving Relations of Pressure, Density, Velocity, Displacement, and Power in a Plane Wave

The equilibrium position of layer (1) is at x and layer (2) at $x + dx$ as shown in Fig. 1. At a time t these layers are at $x + y$ and $x + y + \frac{dy}{dx}dx$.



Appendix B. Figure 1.

Since the mass included between them is the same in both instances

$$\rho_0 dx = (\rho_0 + \rho)dx^1 = (\rho_0 + \rho)\left(1 + \frac{dy}{dx}\right)dx \qquad (1)$$

or neglecting ρ compared to ρ_0

$$\frac{dy}{dx} = -\frac{\rho}{\rho_0}. \qquad (2)$$

The forces acting on the layer of air at a time t are

$$P + p - \left(P + p + \frac{dp}{dx}dx\right) = -\frac{dp}{dx}dx.$$

This is equal to the mass times the acceleration of the layer or

$$-\frac{dp}{dx}dx = \rho_0 dx \frac{d^2y}{dt^2} \tag{3}$$

From the adiabatic law of gases

$$\frac{P+p}{P} = \left(\frac{\rho_0+\rho}{\rho_0}\right)^{\gamma} \quad \text{or} \quad p = \frac{\gamma P}{\rho_0}\rho. \tag{4}$$

Substituting from (2) and (4)

$$\frac{d^2y}{dt^2} = a^2\frac{d^2y}{dx^2} \tag{5}$$

where

$$a = \sqrt{\frac{\gamma P}{\rho_0}}. \tag{6}$$

Equation (5) is the desired equation of motion. A solution of this equation is

$$y = A\cos\left(\omega t + \frac{\omega}{a}x\right) \tag{7}$$

where A and ω are arbitrary constants. If equation (7) is written in the familiar form of a wave equation, namely

$$y = A\cos 2\pi\left(\frac{t}{T} + \frac{x}{\lambda}\right) \tag{8}$$

it is seen that a is the velocity of propagation of the wave since it is equal to the frequency $\frac{\omega}{2\pi}$ times the wave length λ. Also A is the amplitude of vibration of each air particle as the wave passes. A summation of any number of terms of the above form will also satisfy the equation. These equations enable one to write down the relations between the amplitude and velocity of the vibrating particles of air, the excess density, and the excess pressure caused during the passage of the wave. If the same letters are used to represent r.m.s. values instead of instantaneous values, then the amplitude y of the vibrating particles is given by

$$y = \tfrac{1}{2}\sqrt{2}A. \tag{9}$$

The velocity v of an air particle is

$$v = \left(\frac{dy}{dt}\right)_{\max} = y\omega \tag{10}$$

with a phase of 90° compared to the displacement. The excess density ρ is by equations (2) and (7)

$$\rho = \rho_0\frac{\omega y}{a} = \rho_0\frac{v}{a} \tag{11}$$

and is opposite in phase to the velocity v. The excess pressure p is given by equation (4) or

$$p = a^2\rho = \rho_0 av = \rho_0 a\omega y \tag{12}$$

with a phase of 180° from the velocity v or 90° behind y. The quantity $(\rho_0 a)$ is called the radiation resistance and may be designated by r, then

$$p = rv. \tag{13}$$

This equation is similar in form to the equation expressing ohms law for electrical circuits; p corresponding to potential difference, r to electrical resistance, and v to electrical current. By analogy, then, it might be inferred that the intensity J of the sound or power flowing through a square centimeter would be given by

$$J = pv = \frac{p^2}{r} = rv^2. \tag{14}$$

That these equations do hold is easily shown as follows. The kinetic energy in a tube of the sound waves which is a centimeters long, is given by

$$\left.\begin{aligned} \text{K.E.} &= \int_0^a \tfrac{1}{2}\rho_0 dx\cdot\left(\frac{dy}{dt}\right)^2 \\ &= \tfrac{1}{2}\rho_0 A^2\omega^2\left[\tfrac{1}{2}a + \tfrac{1}{2}\int_0^a \cos 2\left(\omega t + \frac{\omega}{a}x\right)\cdot dx\right] \end{aligned}\right\} . \tag{15}$$

The last term is fluctuating with time but is always negligibly small compared to the first term. Therefore,

$$\text{K.E.} = \tfrac{1}{4}\rho_0 A^2\omega^2 a = \tfrac{1}{2}rv^2. \tag{16}$$

Since there is a constant interchange between potential and kinetic energy, they must be on the average equal. The total energy in a tube of unit cross-section which is *a* centimeters long is then

$$J = rv^2$$

which is the intensity of the sound, since this energy moves along with the wave and passes through a unit cross-section every second. The fractional change in pressure $\frac{p}{P}$ is related to the fractional change in density by the formula

$$\frac{p}{P} = \gamma\frac{\rho}{\rho_0} \tag{17}$$

as will be seen from equations (12) and (6).

For example, consider these relationships for the average speech intensity. For air, the radiation resistance is 41.5 at 20° C. The average speech intensity at 10 centimeters from the mouth is approximately .01 microwatt or .1 erg per second. Consequently, the excess pressure p is 2 bars as compared to 1,000,000 bars for the undisturbed state. The velocity v is .05 cm/sec. The displacement y, if it is assumed that most of the energy is at 100 cycles, is $\frac{1}{1000}$ millimeter. At other frequencies, it is inversely proportional to the frequency. The excess density ρ is 2×10^{-9} grams/cc.

APPENDIX C

LET the pressure variation of the air in front of the drum of the ear be designated by p. Since the pressure of the air in the middle ear balances the undisturbed outside air pressure this change in pressure multiplied by the effective area of the ear drum is the only effective force that produces displacements. Let the displacement of the fluid of the cochlea near the oval window be designated by X. If Hooke's law held for all the elastic members taking part in the transmission of sound to the inner ear, then

$$X = kp \tag{1}$$

where k is a constant.

It would be expected from the anatomy of the ear that Hooke's law would start to break down even for small displacements. So in general the relation between the force p and the displacement X can be represented by

$$X = f(p) = a_0 + a_1 p + a_2(p)^2 + a_3(p)^3 + \dots \tag{2}$$

where the coefficients α_0, α_1, $\alpha_2 \dots$ belong to the expansion of the function into a power series. Now if p is a sinusoidal variation, then

$$p = p_0 \cos \omega t, \tag{3}$$

where $\frac{\omega}{2\pi}$ is the frequency of vibration. Substituting this value in (2), terms containing the cosine raised to integral powers are obtained. These can be expanded into multiple angle functions. For example, for the first four powers

$$\cos^2 \omega t = \tfrac{1}{2} \cos 2\omega t + \tfrac{1}{2}. \tag{4}$$

$$\cos^3 \omega t = \tfrac{1}{4} \cos 3\omega t + \tfrac{3}{4} \cos \omega t. \tag{5}$$

$$\cos^4 \omega t = \tfrac{1}{8} \cos 4\omega t + \tfrac{1}{2} \cos 2\omega t + \tfrac{3}{8}. \tag{6}$$

It is evident then that the displacement X will be represented by a formula

$$X = b_0 + b_1 \cos \omega t + b_2 \cos 2\omega t + b_3 \cos 3\omega t + \ldots$$

In other words when a periodic force of only one frequency is impressed upon the ear drum, this same frequency and in addition all its harmonic frequencies are impressed upon the fluid of the inner ear.

If two pure tones are impressed upon the ear then p is given by

$$p = p_1 \cos \omega_1 t + p_2 \cos \omega_2 t.$$

If this value is sutstituted in equation (2), terms of the form $\cos^n \omega_1 t$ and $\cos^m \omega_2 t$ and $\cos^n \omega_1 t \cos^m \omega_2 t$ are obtained. The first two forms give rise to all the harmonics and the third form gives rise to the summation and difference tones. For example, the first four terms are

$$\begin{aligned}
a_0 &= a_0. \\
a_1 p &= a_1(p_1 \cos \omega_1 t + p_2 \cos \omega t). \\
a_2(p)^2 &= a_2[\tfrac{1}{2}p_1^2 \cos 2\,\omega_1 t + \tfrac{1}{2}p_2^2 \cos 2\,\omega_2 t \\
&\quad + p_1 p_2(\cos(\omega_1 - \omega_2)t + \cos(\omega_1 + \omega_2)t) \\
&\quad + \tfrac{1}{2}(p_1^2 + p_2^2)]. \\
a_3(p)^3 &= a_3[(\tfrac{3}{4}p_1^3 + \tfrac{3}{2}p_1 p_2^2) \cos \omega_1 t + \tfrac{1}{4}p_1^3 \cos 3\,\omega_1 t \\
&\quad + (\tfrac{3}{4}p_2^3 + \tfrac{3}{2}p_1^2 p_2) \cos \omega_2 t + \tfrac{1}{4}p_2^3 \cos 3\,\omega_2 t \\
&\quad + \tfrac{3}{4}p_1^2 p_2 \cos(\omega_2 t + 2\,\omega_1 t) + \tfrac{3}{4}p_1^2 p_2 \cos(\omega_2 t - 2\,\omega_1 t) \\
&\quad + \tfrac{3}{4}p_1 p_2^2 \cos(\omega_1 t + 2\,\omega_2 t) \\
&\quad + \tfrac{3}{4}p_1 p_2^2 \cos(\omega_1 t - 2\,\omega_2 t)].
\end{aligned}$$

Therefore, unless there is a linear relation between a force acting on the ear drum and the displacement at the oval window, that is, unless all the coefficients in equation (2) are zero except a_1, the harmonics and the summation and difference tones will be impressed upon the fluid in the cochlea of the inner ear.

APPENDIX D

The relation between hearing loss in sensation units and the maximum distance for hearing and interpreting speech is established as follows. It is well known that the intensity of sound decreases as the inverse square of the distance from the source. This is true only where there are no solid objects in the vicinity which cause reflections. Stated mathematically,

$$\frac{I}{I_0} = \frac{d_0^2}{d^2} \tag{1}$$

where I is the intensity at a distance of d and I_0 the intensity at a distance of d_0 from the sound source. From the definition of hearing loss given

$$\text{H.L.} = 10 \log \frac{I}{I_0}. \tag{2}$$

Combining equations (1) and (2)

$$\text{H.L.} = 20 \log \frac{d_0}{d}. \tag{3}$$

It has been found that the speech sounds entering a normal ear from a caller whose lips are $1\frac{1}{2}$ inches or $\frac{1}{8}$ foot from the opening of the ear canal may be attenuated 80 decibels before 50 per cent of the called numbers are recognized incorrectly. This was determined experimentally by means of the high-quality telephone system described in Part Four, Chapter IV. The dial of the attenuator was first set so that the speech emerging from the telephone receiver had an intensity equal to that produced when talking directly into the ear at $\frac{1}{8}$ foot distance. The attenuator was then turned until the person of normal hearing could only recognize half of the numbers called. The difference in the settings was found to be equal to 80 decibels. In a similar way it was found that the corresponding differences when an average whisper, or *pp* voice, and a *ff* voice was used were approximately 50 decibels, 65 decibels, and 95 decibels, respectively.

The normal distances for interpreting numbers can be obtained from these figures. From equation (3) it follows that

d_0 = 39.5 feet for the average whisper;

d_0 = 222 feet for the *pp* voice;

d_0 = 1250 feet for the *mf* voice, and

d_0 = 7040 feet for the *ff* voice.

Using these values for d_0 and equation (3) the corresponding values of H.L. (hearing loss) and distance d given in Table XXII, were calculated.

Let the maximum distances for hearing the average whisper, *pp* voice, *mf* voice, and *ff* voice be d', d'', d''', and d'''', for the patient, and d_0', d_0'', d_0''', and d_0'''', for a person of normal hearing. Then from equation (3),

$$20 \log \frac{d_0'}{d'} = 20 \log \frac{d_0''}{d''} = 20 \log \frac{d_0'''}{d'''} = 20 \log \frac{d_0''''}{d''''}$$

or

$$\frac{d_0'}{d'} = \frac{d_0''}{d''} = \frac{d_0'''}{d'''} = \frac{d_0''''}{d''''} \qquad (4)$$

This equation states that regardless of the intensity of voice used, the ratio of the hearing distance for the normal ear to that for the ear of the patient is the same. Let x', x'', x''', and x'''' be the increased hearing distances corresponding to a hearing improvement q for the four types of voices. Then

$$q = 20 \log \frac{d_0'}{d'} - 20 \log \frac{d_0'}{d' + x'} = 20 \log \left(1 + \frac{x'}{d'}\right)$$

Similarly

$$q = 20 \log \left(1 + \frac{x''}{d''}\right) = 20 \log \left(1 + \frac{x'''}{d'''}\right) = 20 \log \left(1 + \frac{x''''}{d''''}\right)$$

or,

$$\frac{x'}{d'} = \frac{x''}{d''} = \frac{x'''}{d'''} = \frac{x''''}{d''''} \qquad (5)$$

Combining equations (4) and (5)

$$\frac{x''''}{x'} = \frac{d_0''''}{d_0'} = \frac{7040}{39.5} = 178$$

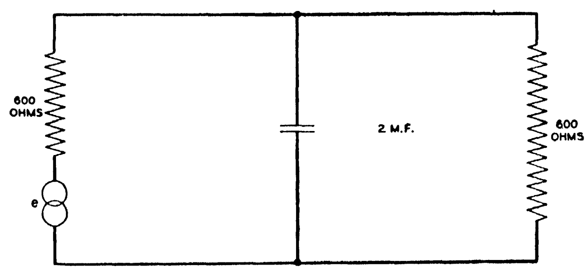
$$\frac{x'''}{x'} = \frac{d_0'''}{d_0'} = \frac{1250}{39.5} = 31.7$$

$$\frac{x''}{x'} = \frac{d_0''}{d_0'} = \frac{222}{39.5} = 5.6.$$

This leads to the very important conclusion that if a patient shows an improvement of 1 foot when using the average whisper he will also show an improvement of 5.6 feet, 31.7 feet, and 178 feet for the other three intensities of calling. This is true regardless of the amount of the improvement; if it is small, the additional distance will be added to a large distance; if large, to a small distance.

APPENDIX E

If a condenser is connected across the two wires of a long transmission line which has a characteristic impedance of 600 ohms, what will be the loudness loss of the reproduced speech? It is easily seen that the ratio r of the current flowing into the second impedance before and after the condenser is strapped across the line is



Appendix E. Figure 1.

$$r = \sqrt{1 + (.0012\pi f)^2}$$

where f is the frequency. The value of y then becomes

$$y = [1 + (.0012\pi f)^2]^{-1/3}.$$

When these values of y were plotted against values of $x = \int_0^f G(f)df$ which are obtained from Fig. 119, the resulting areas were found to be .49 and .39 for the high quality and resonant circuits, respectively. The corresponding values in decibels are 9.3 and 12.3, which are the effective losses required.

APPENDIX F

In singly-resonant systems the velocity (v) (current for electrical systems) produced by a force (electromotive force for electrical systems) is known to be given by

$$v = \frac{F}{R + j\left(m\omega - \frac{S}{\omega}\right)} \quad (1)$$

where R is the resistance, m the mass (inductance for electrical systems), and S the elastic constant (the reciprocal of the capacity for electrical systems). The resonant frequency in kilocycles $f_0 = \frac{\omega_0}{2000\pi}$ is given by the condition that the last term of the denominator is zero or

$$f_0 = \frac{1}{2000\pi}\sqrt{\frac{S}{m}}. \quad (2)$$

The damping constant Δ in bels per second is known to be related to m and R by the equation

$$\Delta = .434\frac{R}{m}. \quad (3)$$

Let α be the number of bels down from the velocity amplitude corresponding to the resonant frequency; then obviously α is given by

$$\alpha = 2 \log \left|\frac{v_0}{v}\right| \quad (4)$$

where the parallel vertical bars indicate absolute values must be taken. Then substituting and reducing

$$\alpha = \log \left[1 + 7.45 \times 10^6\left(\frac{f_0}{\Delta}\right)^2\left(\frac{f}{f_0} - \frac{f_0}{f}\right)^2\right] \quad (5)$$

Or if the value of the pitch P in octaves be introduced from the relation $P = \log_2 f$, this reduces to

$$\alpha = \log\left[1 + 7.45 \times 10^6\left(\frac{2^{P_0}}{\Delta}\right)^2(2^{P-P_0} - 2^{P_0-P})^2\right] \tag{6}$$

which shows that α has the same value for those tones which have the same difference in pitch above or below the resonant pitch. In other words, curves which represent P and α are symmetrical about P_0.

INDEX OF NAMES

Abraham, 134
Adrian, E. D. 126
Arnold, H. D., 134, 196

Beck, C. J., 70
Bevier, L., 20
Boltzmann, F. L., 132
Boring, E. G., 118
Bowlker, 193
Bunch, C. C., 215

Crandall, I. B., 26, 54, 77, 134

Dewey, G., 81
Dean, L. W., 215

Fowler, E. P., 184, 199
Free, E. E., 106
Fry, T. C., 191

Graham, F. H., 181
Gray, A. A., 118, 124

Harrison, H. C., 23
Hart, V. W., 127
Hartley, R. V. L., 191
Helmholtz, H. von, 14, 19, 46, 116, 118
Herman, L., 20, 47
Hewlett, C. W., 139

Johnson, K. S., 279
Jones, I. H., 118, 215

Keith, 117
Kingsbury, B. A., 227
Knudsen, V. O., 118, 145, 215, 296
Koenig, R., 16
Koenig, W., 51
Kranz, F. W., 134, 215

Lane, C. E., 96, 139, 166, 176, 183, 195

Macfarlan, D., 209
Mackenty, J. E., 13
MacKenzie, D., 77, 227
Marx, 123
Maxfield, J. P., 23
Mayer, A. A., 167, 180
Merritt, E., 17
Miller, D. C., 17, 58
Moore, C. R., 90

Nichols, E. L., 17

Paget, Sir Richard, 3, 11, 58
Porter, E. L., 127
Rayleigh, Lord, 133
Riesz, R. R., 150
Ripman, W., 7
Rousselot, P., 16

Sabine, W. C., 15, 227
Sacia, C. F., 26, 65, 70, 75
Scott, L., 16
Scripture, E. W., 21, 47, 125
Seashore, C. E., 215
Shambaugh, G. E., 116
Sivian, L. J., 68
Stewart, G. W., 189
Stewart, J. Q., 10, 269
Stumpf, C., 58

ter Kuile, E., 116
Toepler, A., 132

Wead, 133
Webster, A. G., 133
Wegel, R. L., 90, 139, 166, 176, 183, 199
Weiss, O., 15
Wente, E. C., 27, 134, 306
Wheatstone, C., 47
Wien, 133
Wilkinson, G., 118, 124
Willis, 47
Wittmaach, 123
Wrightson, Sir Thomas, 117, 129

Yoshii, 123

SUBJECT INDEX

A

Acoumeter
  relation of measurements with, to sensation units, 206–208
  test for hearing acuity, 206–208
Acoustic spectra
  effect of pitch on, 51, 54, 55
  of musical sounds, 90–95
  of vowel sounds, 50–55
  of watch tick, 208
Acuity
  distances for hearing average speech, 187
  in presence of noise, 185–187
  methods of testing, 198–221
  of normal ear, 132–142
  relation between various tests for, 202–221
Alphabet
  International Phonetic Association, 5
  National Education Association, 5
  Simplified Spelling Board, 5
  Standard (Bell Telephone Laboratories), 5, 6
American Federation of Organizations for the Hard of Hearing, 212
Analyzer, Electrical Harmonic, 51–55, 88–90
Anatomy
  of hearing, see Organs of hearing
  of speech, see Organs of speech
Anvil (ossicle), 112, 113, 118, 121
Articulation
  consonant, 255
  effect of deafness on, 297–302
  effect of frequency distortion on, 279–289
  effect of intensity on, 270–278
  effect of noise on, 297–301
Articulation
  effect of resonance distortion on, 287–289, 318, 319
  effect of reverberation on, 294–296
  effect of speed of rotation of phonograph on, 291–293
  effect of vacuum-tube overload on, 290–292
  letter, 255
  relation between syllable and letter, 266, 268
  relation to intelligibility, 266
  relation to probability, 268, 269
  sound, 255
  standard test for, 256–258
  syllable, 255
  vowel, 255
Artificial larynx, 12, 13
Artificial production of speech sounds
  acoustic, 11, 12
  electrical, 10, 11
Audibility
  methods of measuring, 132–140
  threshold of, 132–142
  threshold of feeling, 132–143
Audiometer
  buzzer, see Buzzer audiometer
  description of, 103–105
  phonograph, see Phonograph audiometer
  pitch-range, see Pitch-range audiometer, 104, 216–218
  use in measuring hearing, 211–221
  use in measuring noise, 103–105
Audiogram
  noise, definition, 104
  typical deafness, 200, 217
  typical noise, 106, 186, 299
Audition, see Hearing

Auditory Masking, 167–187
see Masking
Auditory Meatus, 111, 113
Auditory Nerve Action, 125–131
Auditory Sensation Area
definition, 144
diagram, 141
Average Speech Power, 64–71

B

Balance method of measuring noise, 105
Basilar membrane, 114–117, 121, 122
positions on, for sensing pitches, 164–166
response to musical tones, 252–254
vibration form, 180–185
Beats
binaural, 188–196
objective, 194
subjective, 194–196
Bel, definition, 68, 69
Binaural
beats, 188–196
location of complex sounds, 192–196
loudness, comparison with monaural loudness, 237, 238
masking, 172, 173
Bone conduction, 174
function in producing objective binaural beats, 194
Buzzer audiometer (Nos. 3-A and 5-A)
description, 103, 214, 215
use in measuring hearing, 214, 215, 217–220
use in measuring noise, 103

C

Canal, ear, 111, 112
Capsule, manometric (Koenig), use in recording speech waves, 16, 17
Cavities
nose, 4
throat, 4
Characteristic frequencies
of music, 87–98
of speech, 51–63
Characteristics of speech
energy distribution, 76–80
Characteristics of speech
frequency analyses, 51–63
pitch, 51–55, 60, 62
power, 64–75
typical speech waves, 29–49
Children, school, methods of measuring hearing of, 211–214
Classification of speech sounds, 6
Cochlea, 112–117
Coin click
relation to sensation units, 206–208
test for hearing acuity, 206–208
Components, frequency, see Frequency
Condenser transmitter
calibration, 305–307
description, 27
Conduction
bone, see Bone conduction
nerve, 125–131
Consonants
articulation of, see Articulation
duration of, 57–62
fricatives, 9, 10
power, 71, 74
sensation level, 73
stops, 9, 10
unvoiced, 9, 10
voiced, 9, 10
Conversion scale
hearing loss in sensation units to maximum hearing distances for speech, 202–206, 314–316
hearing loss in sensation units to per-cent loss, 201
hearing loss in sensation units to watch tick, acoumeter, tuning fork and coin click test results, 202–221
sensation level to loudness, 229–232
Cords, vocal
action, 5
description, 5
diagram, 4

D

Deafness
audiograms of typical cases, 200, 217
effect on articulation, 297–302
hearing-loss scale, 199–201
methods of measuring, 198–221

Deafness
relation between various tests for, 202–221
Decibel, definition, 68, 69
Density, relation to pressure, velocity, displacement and power in a plane wave, derivation of equations for, 308–311
Difference tones, 175–180
effect on pitch, 250–254
production by ear overload, 312, 313
Differential intensity sensitivity, 145–151, 158, 159
Differential pitch sensitivity, 151, 152, 158–164
Diphthongs, formation, 8, 9
Displacement, relation to density, pressure, velocity and power in a plane wave, derivation of equations for, 308–311
Distances
for hearing average speech, 187
relation between sensation units loss and maximum hearing distance for speech, 202–206, 219, 314–316
Distortion
effect of frequency distortion on articulation, 279–289
effect of frequency distortion on pitch, 245–254
effect of resonance distortion on articulation, 287–289, 318, 319
Drum, ear, 113, 118
Duration of speech sounds, 56–62

E

Ear
canal, 111, 112
characteristics, see Hearing
drum, 113, 118
functions, 118–125
inner, description, 112–117
middle, description, 111, 112
outer, description, 111, 112
overload of, 312, 313
Electrical harmonic analyzer
description, 88–90
speech records from, 51–55
Electrostatic transmitter
calibration, 305–307
description, 27
Energy, frequency distribution in speech, 76–80
Equal loudness, curves for, 230
Eustachian tube, 112, 113
Evolution of speech, 3, 4

F

Feeling, threshold of, 132–143
Film, motion-picture, use in recording speech waves, 24–26
Filters, electrical
circuits of typical, 280
use of, see Frequency distortion and Frequency components
Formation of speech sounds, 5–10
Frequency
analyses of distribution in speech, 51–63
analyses of line noise, 101
component multiplication, effect on articulation, 291–293
components, effect on pitch and quality of musical sounds, 245–254
distortion, effect on loudness, 233–237
distortion, effect on articulation, 279–289
energy distribution in speech, 76–80
level, 153–157
limits for hearing, 143, 144
minimum perceptible, differences of, 151, 152, 158–164
ranges in music, 87–98
ranges in speech, 51–63, 76–80
Frequency of occurrence
of speech sounds, 81–84
of syllables, 81–83
of words, 81–83
Functions
of hearing organs, 118–125
of speech organs, 4, 7–10

H

Hammer (ossicle), 112, 113, 118, 121
Harmonics
effect on pitch and quality of musical sounds, 245–254

Harmonics
- production by ear overload, 312, 313
- subjective, 175–180

Harmonic analyzer, electrical, 51–55, 88–90

Harmonic theory of vowel production, 47–51

Hearing
- binaural beats, 186–196
- description of organs, 111–117
- distances for hearing average speech, 187
- function of organs, 118–125
- in presence of noise, 185–187
- interference, 167–187
- limits of, 132–143
- location of sound images, 188–196
- loss, see Hearing loss
- loss scale, 199–201
- loudness, 225–245
- masking, 167–187
- methods of testing acuity, 198–221
- nerve conduction, 125–131
- positions on basilar membrane for sensing pitches, 164–166
- relation between various tests for, 202–221
- scale, 199–201
- sensitivity for intensity differences, 145–151, 158–164
- sensitivity for pitch differences, 151, 152, 158–164
- sensitivity, 132–142
- subjective tones, 175–180, 250–254, 312, 313
- theories of, 118–131
- vibration forms of basilar membrane, 180–185

Hearing loss
- percent, definition of, 199
- relation between various tests for, 201–221, 314–316
- scale, 199–201

Helicotrema, 116, 117, 118, 121

Helmholtz resonator, use in recording speech waves, 14

I

Images, sound, effect of intensity and phase on location of, 188–196

Inharmonic theory of vowel production, 47–51

Initial speech intensity, definition, 70

Instantaneous speech power, 64–65

Intelligibility (see also Articulation)
- definition of, 256
- methods of measuring, 264–266
- relation to articulation, 266

Intensity
- effect on location of sound images, 188–192
- initial speech, 70
- level, 153–157
- level for audibility, 132–142
- level for feeling, 132–143
- minimum perceptible differences, 145–151, 158–164
- of music, 95–98
- of speech, see Speech power
- of speech sounds, effect on articulation, 270–278
- relation to loudness, 230, 232
- sensitivity of normal ear, see Sensitivity

Interference
- effect of noise on speech and music, 99, 101, 102, 185–187
- masking, 167–187 (see also Masking)

International Phonetic Association, alphabet of, 5

Interpretation, see Articulation and Intelligibility

L

Language, see Speech sounds

Larynx, 4, 5
- artificial, 12, 13

Letter articulation, 255
- relation to syllable articulation, 266, 268
- see Articulation

Level
- frequency, 153–157
- intensity, 153–157
- phonic, 153–157

Limits of audition
lower pitch limit, 141, 143, 144
methods of measuring, 132–142
threshold of audition, 132–142
threshold of feeling, 141–143
upper pitch limit, 141, 143, 144
Line noise, 101
Lips, use in sound production, 4, 7, 8
Location, binaural, of complex sounds, 192–196
of sound images, 188–196
Loudness
comparison of binaural and monaural, 237, 238
effect on, of eliminating various frequencies, 233–237
equal, curves for, 230
losses, computation of, 239–241, 317
of complex sounds, 231, 232
of pure tones, 226–231
relation to intensity, 230, 232
relation to sensation level, 230, 232
standard, 225, 226
theory of, 225–244
Lungs, use in sound production, 4

M

Manometric capsule (Koenig), use in recording speech waves, 16, 17
Masking
binaural, 172, 173
effects of complex sounds, 185, 186
effect of one pure tone upon another, 167–180
monaural, 167–172
subjective tones, 175–180
Mean speech power, 64–66
Meatus, auditory, 111–113
Mechanism of hearing
description of organs, 111–117
function of organs, 118–125
nerve conduction, 125–131
positions on basilar membrane for sensing pitches, 164–166
theories of hearing, 118–131
vibration form of basilar membrane, 180–185
Mechanism of speaking, see Speech sounds and Organs of speech
Membrane
basilar, see Basilar membrane
Reissner's, 114
tectorial, 115, 116
tympanic, 112, 113, 118
Minimum audible pressure, 132–142
Minimum perceptible differences in frequency, 151–152, 158–164
Minimum perceptible differences in intensity, 145–151, 158–164
Monaural
location of complex sounds, 193
loudness, comparison with binaural loudness, 237, 238
masking, 167–172
Motion-picture film use in recording speech waves, 24–26
Musical instruments, see Musical sounds
Music, effect of noise on, 99, 101, 102
Musical sounds
acoustic spectra of, 90–94
effect of frequency components on pitch and quality of, 245–254
effect of harmonics on, 245–254
frequency ranges of, 95–98
intensity of, 95–98
physical properties of, 87–98

N

National Education Association alphabet of, 5
Nerve conduction, 125–131
Noise
audiograms, 104, 106, 186, 299
definition, 99
effect on articulation, 297–301
effect on speech threshold, 73, 185–187
interference effect of, 99, 101, 102, 185–187
masking effects, 185–187
method of measuring, 102–105
results of surveys, 106, 107
room noise, 101–106
telephone line noise, 101
typical wave form of street noise, 100

O

Objective tones, 194
effect on pitch, 250–254
Organ of Corti, 115
Organs of hearing
description, 111–117
functions, 118–125
positions on basilar membrane for sensing pitches, 164–166
vibration form of basilar membrane, 180–185
Organs of speech
description, 4, 5
diagram, 4
use, 4, 7–10
Origin of speech, 3, 4
Oscillograms of typical speech waves, 29–49
Oscillograph, use in recording speech waves, 26–28
Ossicles, 112, 113, 118, 121
Oval window, 112, 113, 118, 121
Overload
of ear, production of overtones by, 312, 313
of vacuum tube, effect on articulation, 290–292
Overtones
effect on pitch and quality of musical sounds, 245–254
production by ear overload, 312–313

P

Pantomime, influence upon word formation, 3
Passages, vocal, use in sound production, 4
Peak speech power, 64–66, 70, 71, 74, 75
Perception
of music, see Musical sounds
of speech, see Speech sounds
tonal, minimum time for, 152, 153
Phase, effect on location of sound images, 188–196
Phaser, description, 189
Phonautograph (Koenig and Scott), use in recording speech waves, 16, 17
Phonetic speech power, 64–66, 70, 71, 74 75
Phonic level, definition of, 153–157
Phonodeik (D. C. Miller), use in sound recording, 17–19
Phonograph
effect of speed of rotation on articulation, 291–293
electrical recording for, 23
use for recording speech waves, 19–22
Phonograph audiometer (No. 4–A)
description, 105, 211–214
use in measuring hearing, 211–214
use in measuring noise, 105
Phonoscope (Weiss), use in recording speech waves, 15, 16
Photo-electric cell, use in film recording of speech, 26
Pinna, 112
Pitch
criterion for determining, 245
definition of, 245
differential sensitivity for, 151, 152, 158–164
effect of frequency components on pitch of musical sounds, 245–254
effect of difference tones on, 250–254
effect of harmonics on, 245–254
effect on acoustic spectra of vowels, 51, 54, 55
limits for hearing, 141, 143, 144
of male and female speech, 53, 54, 60, 62
range in musical instruments, 97
recognition of, 245–250
scale, definition of, 153–157
Pitch-range audiometer (Nos. 1–A and 2–A)
description, 104, 216–218
use in measuring hearing, 216–221
use in measuring noise, 104
Power
in music, 95–98
in speech, see Speech power
relation to density, pressure, velocity and displacement in a plane wave, derivation of equations for, 308–311

Pressure
for feeling threshold, 132–143
minimum audible, 132–142
minimum perceptible differences of, 145–151, 158–164
of music, 95–98
of speech sounds, see Speech power
relation to density, velocity, displacement and power in a plane wave, derivation of equations for, 308–311
Probability, relation to articulation, 268, 269

Q

Quality
effect of harmonics on, 245–254
of musical sounds, effect of frequency components on, 245–250

R

Range of frequencies
in music, 95–98
in speech, 51–63, 76–80
see also Frequency
Recognition
of pitch, 245–250
of speech, see Articulation
Recording speech waves
condenser transmitter, 27
Helmholtz resonator, 14
manometric capsule, 16, 17
motion picture, 24–26
oscillograph (high quality), 26–28
phondeik, 17–19
phonograph, 19–24
phonautograph, 16
phonoscope, 15, 16
photographing condensations and rarefactions, 15
Records of speech waves
frequency analyses, 51–55
typical oscillograms, 29–49
Reissner's membrane, 114
Resonance
characteristic resonant frequencies of speech, 51–63
distortion, effect on articulation, 287–289, 318, 319
Resonance
theory of hearing, 118–125
Resonator, Helmholtz, use in recording speech waves, 14
Reverberation
effect on articulation, 294–296
time, definition, 295
Room noise, see Noise
Round window, 112, 113, 118, 121, 122

S

Scala
media, 114
tympani, 112, 114, 121, 122
vestibuli, 112, 114, 121
School children, methods of measuring hearing of, 211–214
Sensation area, auditory, 141, 144
Sensation level
definition, 68–70
of speech sounds, 73
relation to intensity, 230, 232
relation to loudness, 230, 232
Sensation unit, definition, 68, 69
Sensitivity
differential intensity sensitivity, 145–151, 158–164
differential pitch sensitivity, 151, 152, 158–164
methods of testing, 198–221
of normal ear, 132–142
see also Hearing
Sign language, 3
Similarity of languages, 4
Simplified Spelling Board, alphabet of, 5
Sound articulation, see Articulation
Sound images, effect of intensity and phase on location of, 188–196
Sounds
of music, see Musical sounds
of speech, see Speech sounds
Speech
characteristics, see Characteristics of speech
effect of noise on, 73, 99, 101, 102
initial speech intensity, 70
see also Speech sounds
Spectra, acoustic, see Acoustic spectra

Speech organs, 4, 5, 7–10
Speech power
average, 64–71
frequency distribution, 76–80
instantaneous, 64, 65
mean, 64–66
peak, 64–66, 70, 71, 74, 75
phonetic, 64–66, 70, 71, 74, 75
syllabic, 64–66, 74
Speech sounds
articulation of, see Articulation
artificial production of, 10–12
classification, 6
duration of, 56–62
formation, 5–10
frequency analyses, 51–63, 76–80
frequency of occurrence, 81–84
origin and evolution, 3, 4
pitch of, 51–55
power, 64–74
recognition of, see Articulation
sensation level, 73
methods of measuring recognition of, 255–268
theories of vowel production, 46–51
typical speech waves, 29–49
voiced and unvoiced, 5
Speech waves
frequency analyses of, 51–55
methods of recording, 14–28
typical oscillograms of, 29–49
Standards, see Units
Steady state (harmonic) theory of vowel production, 47–51
Stirrup (ossicle), 112, 113, 118, 121
Structure
of hearing organs, see Organs of Hearing
of speech organs, see Organs of Speech
Subjective tones, 175–180
beats, 194–196
effect on pitch, 250–254
production by ear overload, 312, 313
Summation tones, 175–180
effect on pitch, 250–254
production by ear overload, 312, 313
Surveys of noise, 106, 107
Syllable articulation, 255
relation to letter articulation, 266, 268
Syllable. See also Articulation
Syllabic speech power, 64–66, 74
Syllables, frequency of occurrence, 81–83

T

Talking machine, see Phonograph
Tectorial membrane, 115, 116
Telephone line noise, 101
Theories of hearing, 118–131
location of sound images, 188–196
loudness theory, 225–244
positions on basilar membrane for sensing pitches, 164–166
vibration form of basilar membrane, 180–185
Theories of speech production, see Speech sounds
Theories of vowel production
harmonic (steady state), 46–51
inharmonic (transient), 47–51
Thermophone
constants, 305–307
use in measuring threshold intensities, 134–137
Threshold of audibility, 132–142
effect of noise on, 73, 185–187
Threshold of feeling, 132–143
Tonal perception, minimum time for, 152–153
Tones
difference, 175–180, 250–254, 312, 313
objective, 194, 250–254
subjective, 175–180, 250–254, 312, 313
summation, 175–180, 250–254, 312, 313
Tongue, use in sound production, 4, 7–10
Transient (inharmonic) theory of vowel production, 47–51
Transitionals, 8
Transmission unit, 68, 69
Transmitter
condenser, 27, 305–307
electrostatic, 27, 305–307
Tube, eustachian, 112, 113
Tuning fork
relation to sensation unit, 208–211
test for hearing acuity, 208–211
Tympanic membrane, 112, 113, 118

U

Units
  of hearing loss, 199–201
  of loudness, 225, 226
  sensation, 68, 69
  transmission, 68, 69
Unvoiced sounds, production, 5

V

Vacuum tube, overload of, effect on articulation, 290–292
Velocity, relation to density, pressure, displacement and power in a plane wave, derivation of equations for, 308–311
Victor Talking Machine, 292
Vocal organs, see Organs of speech
Voiced sounds, production, 5
Vowels
  acoustic spectra of, 50–55
  articulation of, see Articulation
  duration of, 56
  power, 71, 74
Vowels
  pure, 7
  semi, 9
  sensation level, 73
  theories of production, 46–51
Vowel triangle
  description, 7
  diagram, 6

W

Watch tick
  acoustic spectrum, 208
  relation to sensation units, 206–208, 217–219
  test for hearing acuity, 206–208
Waves, speech, see Speech waves
Weber–Feschner law, 145
Window
  oval, 112, 113, 118, 121
  round, 112, 113, 118, 121, 122
Words
  formation of, see Speech sounds
  frequency of occurrence, 81–83