The Bell System Technical Journal

July, 1931

Some Physical Characteristics of Speech and Music *

By **HARVEY FLETCHER**

Kinematic and statistical descriptions of the physical aspects of speech and music are given in this paper. As the speech or music proceeds, the kinematic description consists in giving the principal melodic stream, namely, the pitch variation and also the intensity and the quality variations. For speech and song, the quality changes are principally described by giving, besides the main melodic stream, two secondary melodic streams corresponding, respectively, to the resonant pitches of the throat and mouth cavities. To this must also be added the positions of the stops and the high pitched components of the fricative consonant sounds as functions of the time. The statistical description consists in giving the average, the peak, and the probable variations of the power involved as the various kinds of speech and music proceed. These general ideas are illustrated by numerous experimental data taken by various instrumental devices which have been evolved in the Laboratories during the past fifteen years.

A speech or musical sound is transmitted from the mouth of a speaker or from a musical instrument through the air to the ear of the listener by means of a pressure wave, a succession of condensations and rarefactions of the air. Such a wave spreads in all directions away from the source of sound and soon encounters solid objects which cause reflections. These reflected waves combine with the original one and thus modify the pressure changes taking place at any point. In this paper we shall be concerned chiefly with the pressure changes which take place before reflections occur.

Speech is composed of fundamental sounds called vowels and consonants. As a conversation proceeds there is a constant shifting from one of these sounds to another, only one of them being sounded at one time. Most of these sounds may be continued as a steady tone and hence may be designated as continuants. The others require that the sound stream be interrupted and are therefore called stops. The first class includes the long and short vowels, the diphthongs, the semi-vowels, and the fricative consonants, the sounds ā, ĭ, ou, l and s being typical, respectively, of each of these groups. The pure stops are p, t, ch, and k. In producing the corresponding voiced stops, b, d, j and g, the voiced stream is not entirely interrupted, although the tones from the vocal cord are very much subdued. A conversation,

---

* Presented as invited paper in Symposium on Acoustics, American Phys. Soc., Dec. 30–31, 1930, Cleveland, Ohio. Published in Rev. of Modern Physics, April, 1931.

then, consists of a succession of continuants and stops and a physical interpretation of speech consists, therefore, of a description of these continuants and a discussion of the manner of joining the continuants together either directly or by means of stops.

Melodic Streams of Speech

As an example of how this analysis of speech may be made consider the sentence, "Joe took father's shoe bench out," an oscillogram of which is shown in Fig. 1.[1] This silly sentence was chosen because it is used in our laboratory for making tests on the efficiency of telephone transmitters. This sentence together with its mate "She was waiting at my lawn" contains all of the fundamental sounds in the English

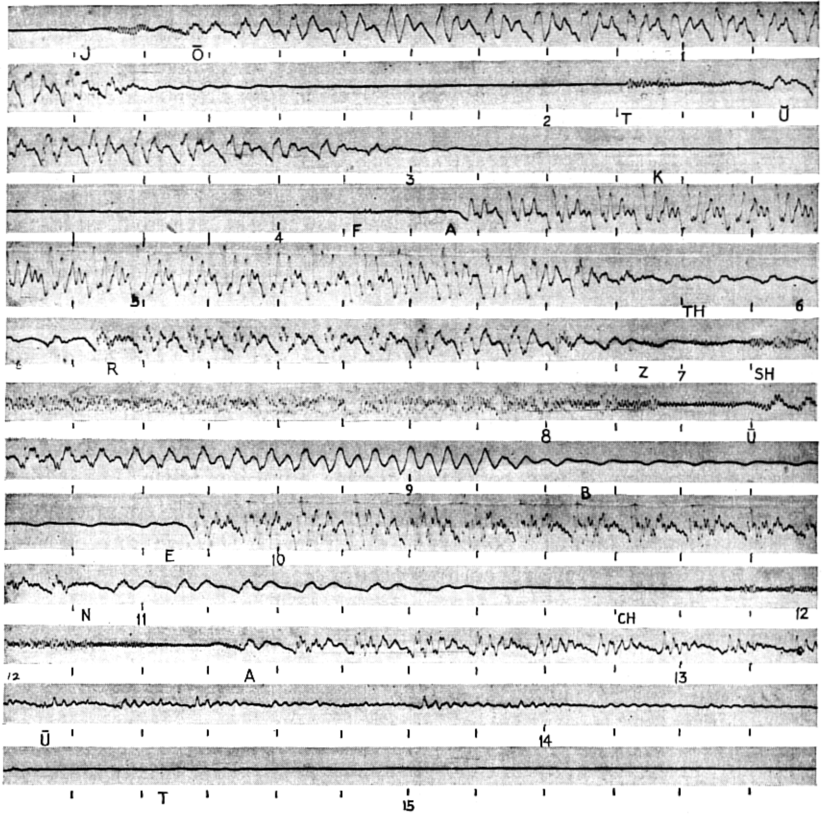

Fig. 1—Oscillogram: "Joe took Father's shoe bench out"—spoken.

---

[1] This oscillogram and the others following it were taken with the new high quality and high speed oscillograph which has recently been developed in our laboratory. It has an approximately uniform response for amplitude and phase from 20 to 10,000 cycles per second.

language that contribute toward the loudness of speech. In Fig. 1 the ordinates are proportional to the pressure change in bars and the abscissas are time intervals of .01 second. The eighteen fundamental sounds in this sentence are joined together without the stream of sound being interrupted except for the stops t, k and ch. The stop consonant b is voiced so that although the vocal cord sound is interrupted by the closing of the lips, it continues to sound in a subdued way until the stop is removed and the e sound begins. Pauses, that is, silent intervals, are made between sentences and sometimes between words. It will be noticed that a brief pause was inserted at the intervals .17 to .21 and .32 to .335 and .34 to .41 and 1.16 to 1.18 seconds. There is no such pause between "shoe" and "bench."

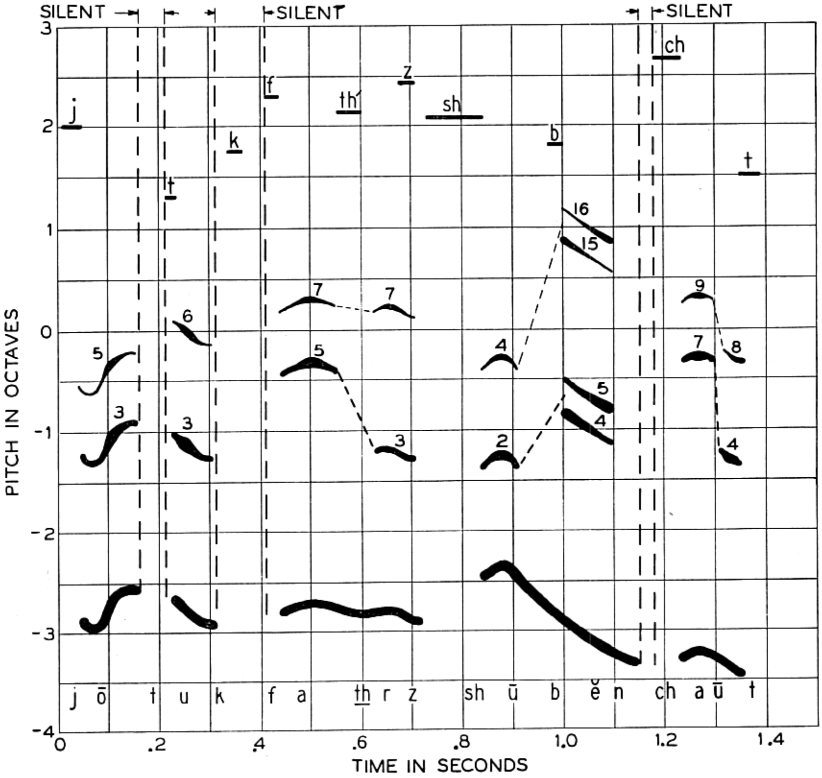

Fig. 2—Melodic curves: "Joe took Father's shoe bench out"—spoken.

Speech, then, consists of a series of comparatively steady states of vibration joined together in time, either by silences or transitions from one steady state to another. Each one of these steady states is characterized by a pitch and a tone quality, and the sequence is essentially a melody. The melody of the sentence whose wave form is shown in Fig. 1 may be illustrated graphically as indicated in Fig. 2. In this figure the ordinates represent the pitch in octaves below or above a tone having a frequency of one kilocycle per second; or if the frequency f is measured in kilocycles, then the pitch P is given by the equation

$$P = \log_2 f. \tag{1}$$

The abscissas represent the time in seconds. The lower curve gives the changes in the pitch of the fundamental and represents the melody as ordinarily understood in music. The middle two curves represent the pitch positions of the strongest harmonics. The location of these positions is determined by the resonant properties of the throat and mouth cavities. These curves may be considered as secondary melodic streams. The combination of these two secondary melodic streams is interpreted by the senses as a sequence of spoken vowels rather than as a series of pitch changes. The small number above each part of the curve gives the number of the harmonic which is augmented by the resonance of the mouth or throat. For the sound e in bench the 4th harmonic was the strongest at the beginning of the sound, but the 5th came in strongest near its end. I have tried to indicate the relative intensities of the harmonics as the sound proceeds by the relative thicknesses of the lines. An examination of the oscillogram shows that the intensity of the harmonic always increases as its pitch becomes nearer the characteristic pitch for the vowel being spoken.

As indicated by the short lines at the top of the chart, there exists at certain intervals high pitched components which are characteristic of the fricative sounds. The unvoiced sounds t, k, f, z and sh, exist only when the three melodic streams are stopped. The high pitched components of the voiced sounds, j, *th* and b, are superimposed upon the three melodic streams.

Besides these four important streams of speech (Fig. 2), there are a great many others with intensities which are in general much lower, but when combined with the main streams they determine the kind of voice, that is, whether it is smooth and musical or rough and harsh. The main melodic stream for a woman's voice is between the pitches -1 and -2 octaves while for a man's voice it is between -3 and -2 octaves. The secondary melodic streams produced while speaking the same sentence are approximately the same for man and woman and of pitches shown in Fig. 2.

In Fig. 3 is shown an oscillograph of the sentence "How are you?".

This sentence contains no stops. The sound stream is not interrupted; it is just a continuous variation from one vowel to another. In Fig. 4 the main melodic stream is given.

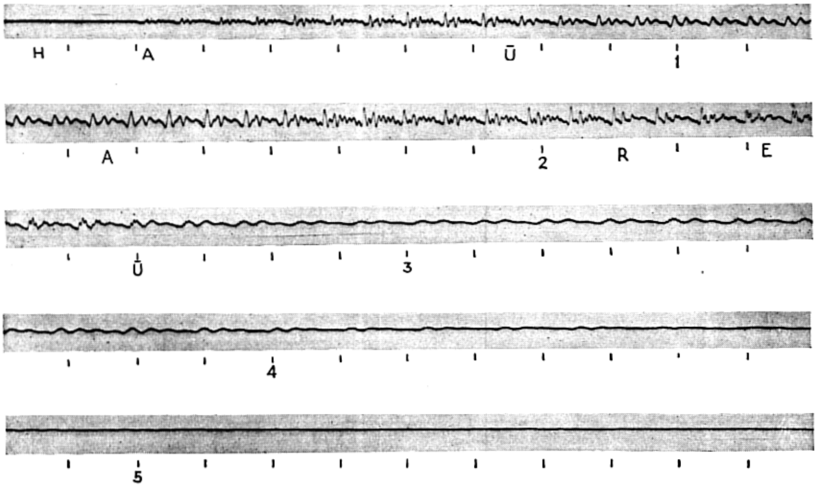

Fig. 3—Oscillogram: "How are you?"

In Fig. 5 an oscillograph of the sentence "Joe took father's shoe bench out" is shown when the vowels of this sentence are intoned on the simple melody do-re-me-fa-me-re-do, and in Fig. 6 the melodic

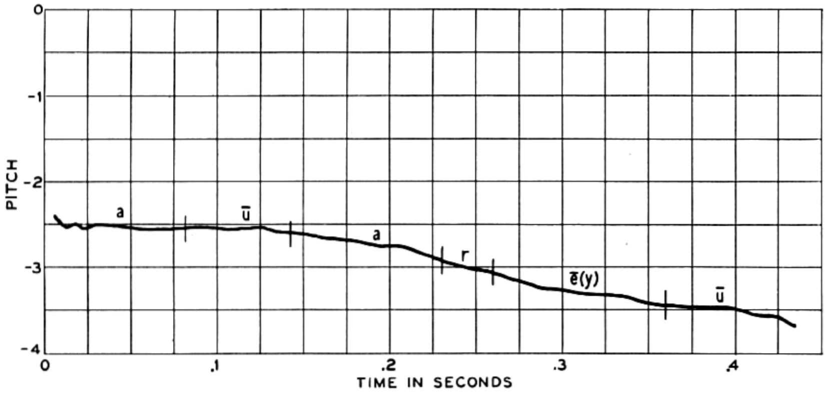

Fig. 4—Melodic curve: "How are you?"

streams are given. In this case only the characteristic resonant pitch positions for the two secondary melodic streams are given. The chief difference between this figure and that for the spoken sentence is in the main melodic stream. For purposes of comparison the curves of the spoken and sung sentence are enlarged and shown together in Fig. 7. In the case of the sung sentence the pitch changes are in definite intervals on the musical scale while for the spoken sentence

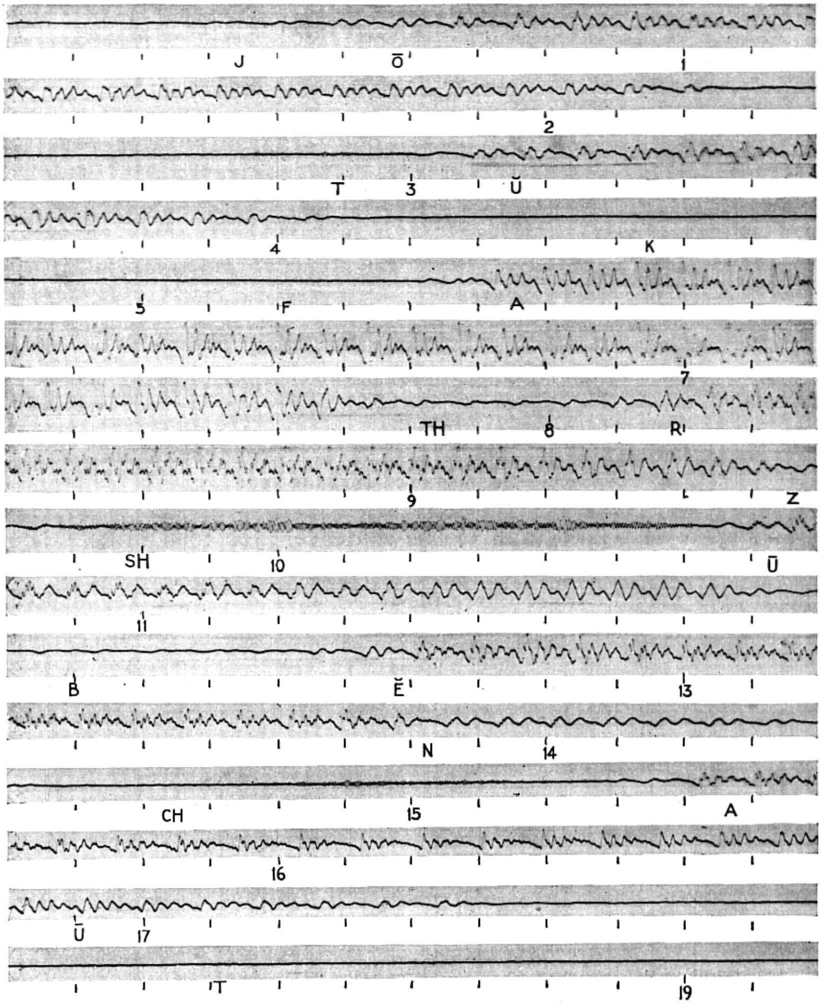

Fig. 5—Oscillogram: "Joe took Father's shoe bench out"—sung.

the pitch varies irregularly, depending upon the emphasis given. The pitch of the fricative and stop consonants is ignored in the musical score, and since these consonants form no part of the music they are generally slid over, making it difficult for a listener to understand the

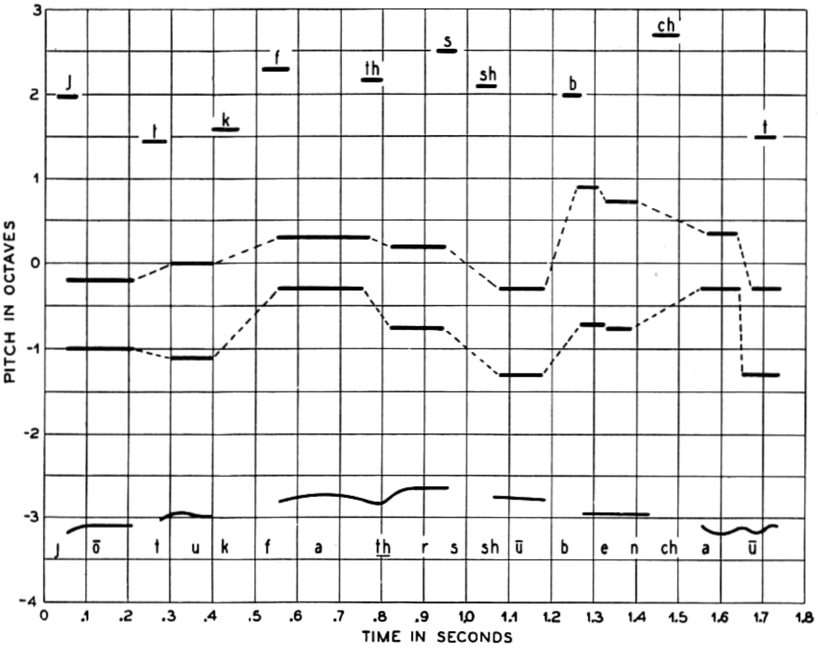

Fig. 6—Melodic curves: "Joe took Father's shoe bench out"—sung.

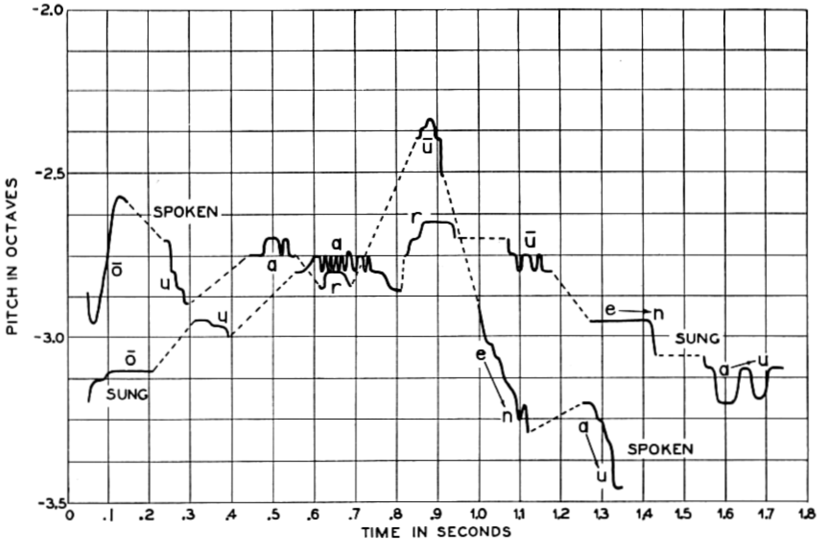

Fig. 7—Melodic curves: "Joe took Father's shoe bench out"—spoken and sung.

meaning of the words. Some of my friends in the musical profession object to this statement of the situation but I think you will agree that a singer's principal aim is to produce beautiful vowel quality and to manipulate the melodic stream so as to produce emotional effects. To do this, it is necessary in singing to lengthen the vowels and to shorten and give less emphasis to the stop and fricative consonants. It is for this reason that it is more difficult to understand song than speech.

CHARACTERISTIC PITCH OR FREQUENCY LEVELS FOR THE VOWELS

Now let us examine part of the speech wave of Fig. 1 in more detail. Consider the vowel in the word "shoe."

The fundamental cycle was repeated 170 times per second. It is evident that the second harmonic is very much magnified until it is nearly as intense as the fundamental. In Fig. 8 is shown another oscillogram of ū intoned at 120 cycles per second. In this case the 3rd harmonic is magnified. An analysis of a number of ū sounds shows that components falling between 300 and 400 cycles per second are always reinforced. This reinforcement is probably due to the resonance characteristic of the mouth cavity.

Fig. 8—Oscillogram of vowel ū.

Similar characteristic low pitch regions exist for the vowels in the words, put, tone, talk, ton and father. A characteristic *high* pitch region also exists for these sounds but the intensity of the components falling in it are much less. For the vowels in the words tap, ten, pert, tape, tip and team there are two characteristic regions of reinforcement which are of approximately the same intensity and which are independent of the fundamental pitch. This is illustrated in Fig. 9, which gives a spectrum analysis of the vowel "ē" pronounced at the four pitches indicated. The characteristic regions are at 375 cycles per second and 2400 cycles per second corresponding to pitches − 1.4 octaves below and + 1.3 octaves above the reference pitch.

Experimental work[2] has indicated that for American speech the characteristic pitch regions for the vowels and semi-vowels are those shown in Fig. 10. For the first six vowels the components corre-

---

[2] "Speech and Hearing," Harvey Fletcher, pp. 58, 59.

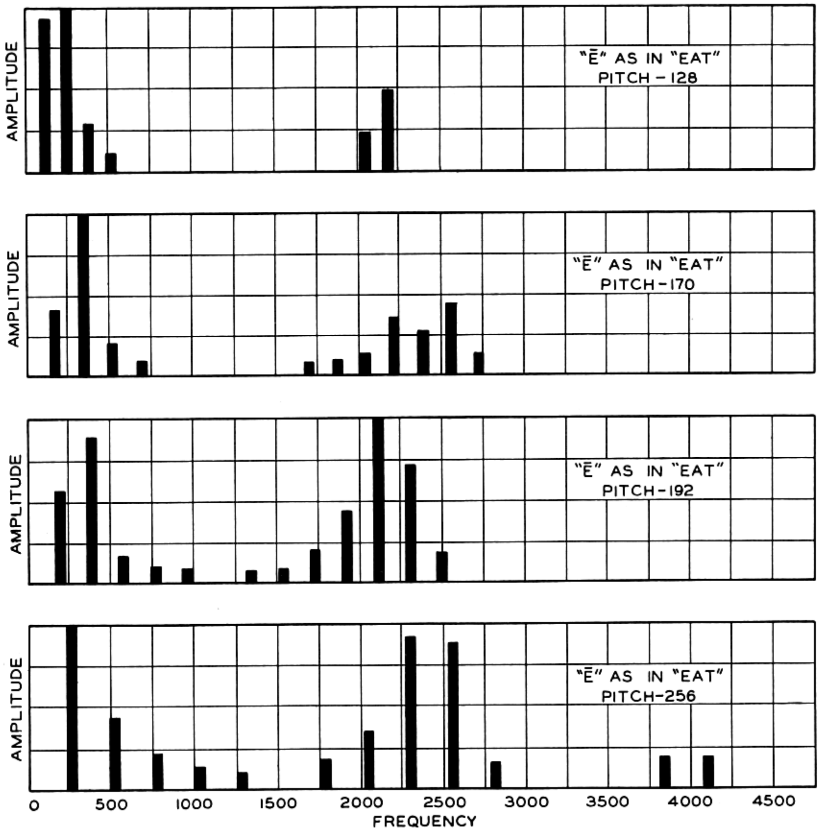

Fig. 9—Spectra of "Ē" intoned at different pitches.

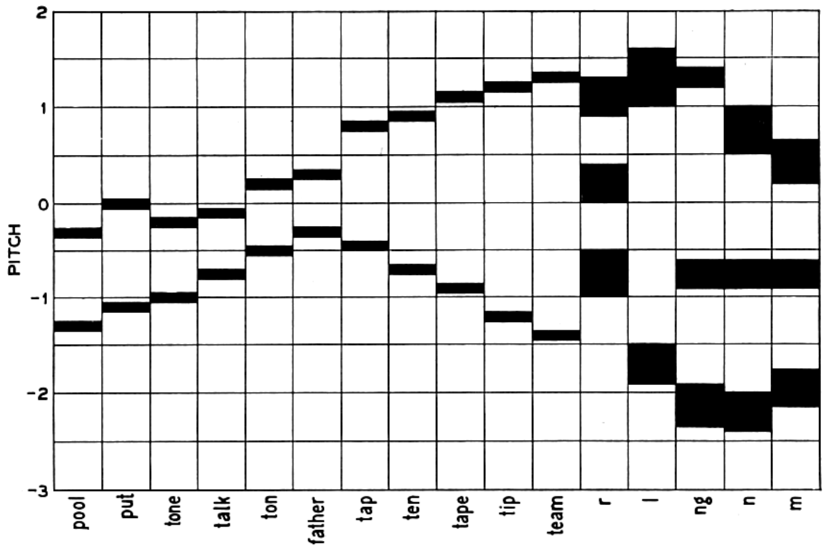

Fig. 10—Characteristic resonance positions for the spoken vowels.

sponding to the characteristic region of high pitch are much less intense than those of low pitch. For the other vowels the intensities of both regions are about alike.

OSCILLOGRAMS OF THE UNVOICED CONTINUANTS

Now let us examine more closely the wave forms for the fricative sounds, s, sh, f, th. They are shown in Fig. 11. These show only part of the oscillogram produced when each of these sounds was continued for about one second. It is seen that these sounds contain components having high pitches mostly above + 1. It is seen that they do not have the wave form repeated as uniformly as was the case with the vowel sounds. They seem to be composed of a series of explosions. For example, the oscillogram for "sh" looks very much like one obtained from the sound of a sky rocket.

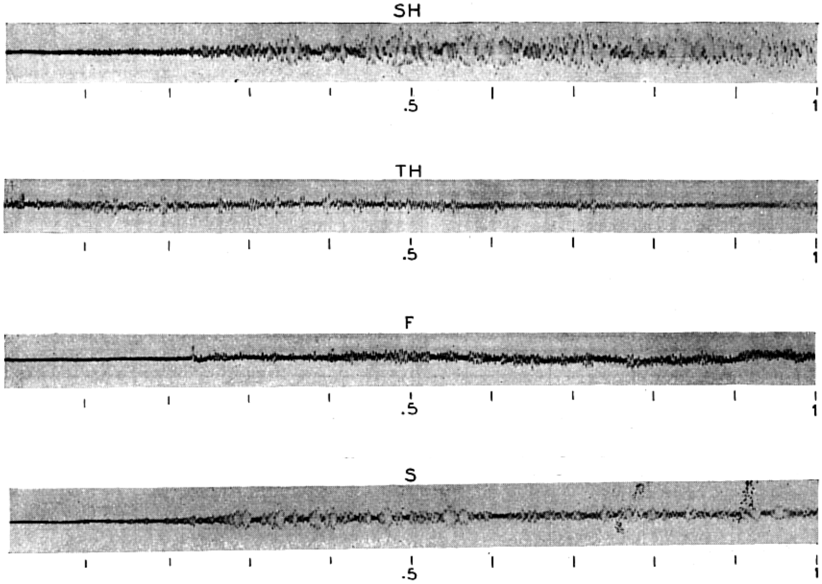

Fig. 11—Oscillograms of fricative consonants.

The f and th sounds are magnified six times in amplitude compared to the sh and s sounds. Although much fainter they still show this explosive character. There are 40, 45, 37 and 55 waves per each .01 second interval, respectively, for these four sounds corresponding to 4000, 4500, 3700, 5500 cycles per second.

ACOUSTICAL POWER OF SPEECH WAVES

Keeping this picture before us, as to the physical composition of speech, and its kinematic nature, let us now consider some statistical averages. If ten different persons spoke the sentence discussed above, there would be a considerable range of differences in the frequencies and intensities used to transmit it through the air. To get a typical cross-section of American speech, it would require at least 100 such sentences pronounced by at least 5 men and 5 women. This would involve the analysis of 18,000 fundamental sounds besides the transitions between them. Also, as was seen from the oscillograms given above, the wave form changes even where it is ideally supposed to be constant so that three or four sample waves from each steady state condition should be analyzed to find the components in each sound. Thus, we have the problem of recording and analyzing about 70,000 such waves. To analyze such a wave by the usual academic methods, namely, to plot the wave to a definite scale and then analyze it into its components by means of a Henrici or similar analyzer, would require at least two or three hours. So such a job for analyzing only the steady-state part of speech would require about 210,000 hours, or 100 years working seven hours a day for 300 days per year. In other words, such a method of attacking the problem is altogether too slow. To find the average intensities and frequencies involved in conversational speech, much more powerful methods for obtaining statistical averages were adopted.

There is a to and fro movement of the air particles simultaneously with the alteration of the air pressure. When the source is so far away that the disturbance can be considered as a plane wave, then the following relations exist between the pressure p, the displacement y, the velocity v, and the acceleration a of a layer of air particles, and the frequency of vibration $\frac{\omega}{2\Pi}$, namely,

$$y\omega^2 = v\omega = a, \tag{2}$$

$$p = rv, \tag{3}$$

where r is the radiation resistance of the air and is given by the product of the air density by the velocity propagation of the wave. The intensity J of the sound at any point is the power passing through a square centimeter of the wave front and is given by

$$J = \frac{p^2}{r}\,. \tag{4}$$

If J is expressed in microwatts and p in bars, this reduces to

$$J = \frac{p^2}{415}\,. \tag{5}$$

The intensity level I is defined by

$$I = \log_{10} J \tag{6}$$

and is expressed in bels. These relations hold for any complex sound as well as for a pure tone if p is interpreted as the root mean square value of the pressure change.

It is seen then that all of these quantities can be determined by making experimental measurements of the pressure change. For accomplishing this the following methods were used.

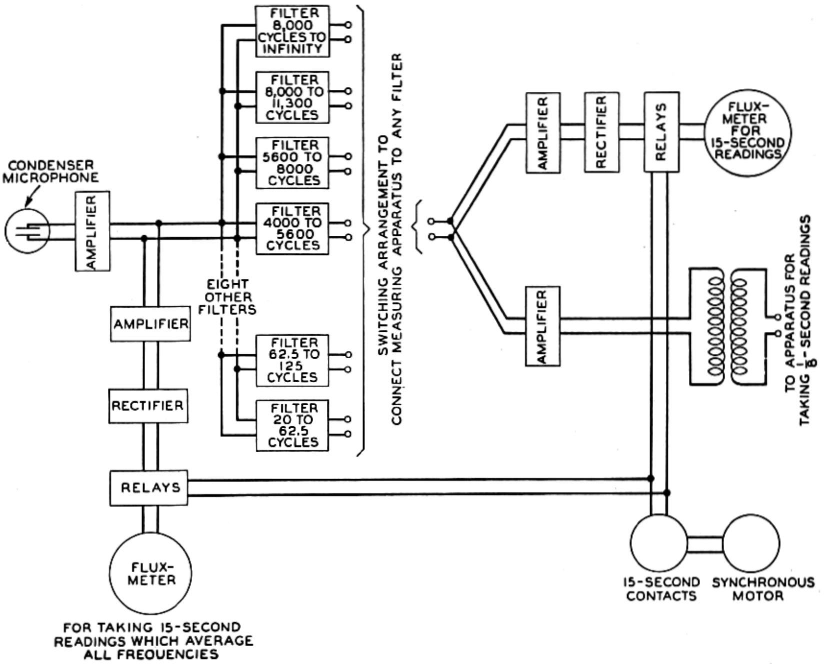

Fig. 12—Schematic of electrical circuit for measuring the average power-frequency distribution of sounds.

The speech to be analyzed is picked up by a Wente condenser microphone and sent into a vacuum tube circuit. This circuit is arranged so that any one of 14 band pass filters can be inserted. After passing through the filter the electrical speech wave is then sent through a rectifier and finally into a meter. A schematic[3] of

---

[3] See paper entitled "A New Analyzer of Speech and Music" by H. K. Dunn (*Bell Laboratories Record*, November, 1930) and also paper entitled "Absolute Amplitudes and Spectra of Certain Musical Instruments and Orchestras" by Sivian, Dunn & White, *Jour. Acous. Soc. of America*, Jan., 1931.

the circuit is shown in Fig. 12. Two kinds of meters are used. The first is a flux meter as shown in Fig. 12 for integrating the speech energy over any desired interval. When the rectifier is designed to give a value which is proportional to the average voltage, then the deflection of the needle of the flux meter will be proportional to the average pressure times the time. In other words, this device will read the average pressure during any desired time interval. In this way it is possible to find the average pressure in any one of the 14 bands. If the rectifier is adjusted so that the reading is proportional to the square of the impressed voltage then the reading will correspond to the average power. Knowing the calibration[4] of the transmitter

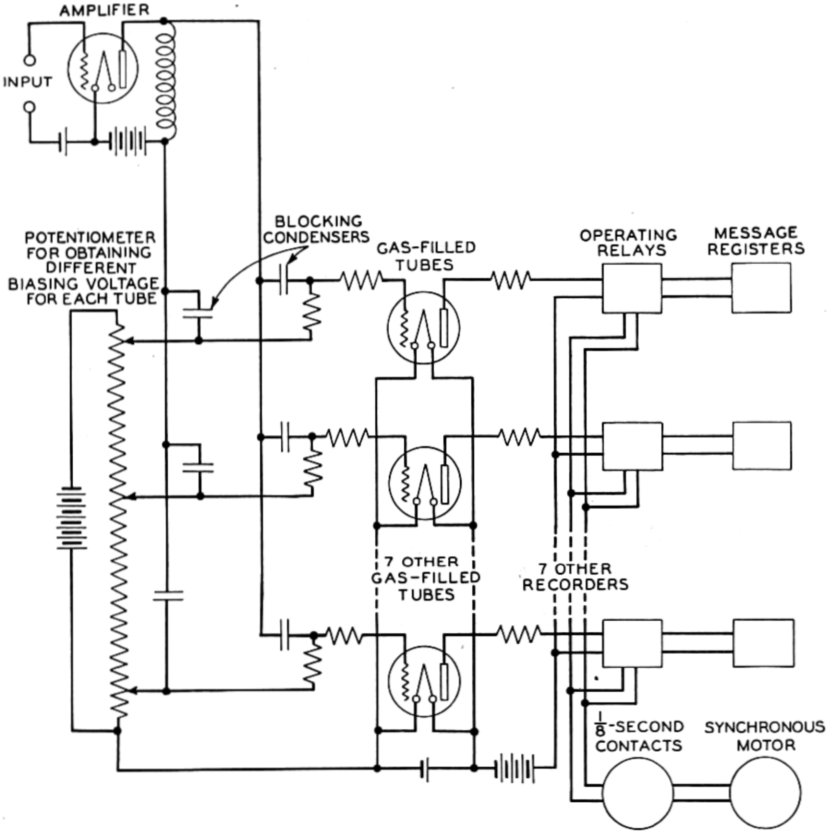

Fig. 13—Schematic of electrical circuit for measuring the peak power-frequency distribution of sounds

---

[4] "Speech and Hearing," page 305, and also paper entitled "Absolute Calibration of Condenser Transmitters" by L. J. Sivian, *Bell System Tech. Jour.*, Jan., 1931.

and also its distance from the mouth of the speaker, it is possible to calculate approximately the average speech power.

The other type of meter shown in Fig. 13 consists of a series of parallel circuits, each containing an argon filled three-electrode tube connected in such a way that in adjacent circuits the tube breaks down and allows the passage of current for voltage levels which are 6 db (decibels) apart. Ten such circuits then cover a range of 54 db. In each of these circuits a relay and counter are connected so that for each tube discharge the counter operates. In this way the number of times the tube breaks down is automatically registered. The speech wave coming from the rectifier is sent into this meter where the peak values are measured; that is, the number of times the pressure exceeds a value fixed by each of these circuits will be registered automatically by the corresponding counter. The apparatus is arranged so that every other 8th second interval is measured, the intervening interval

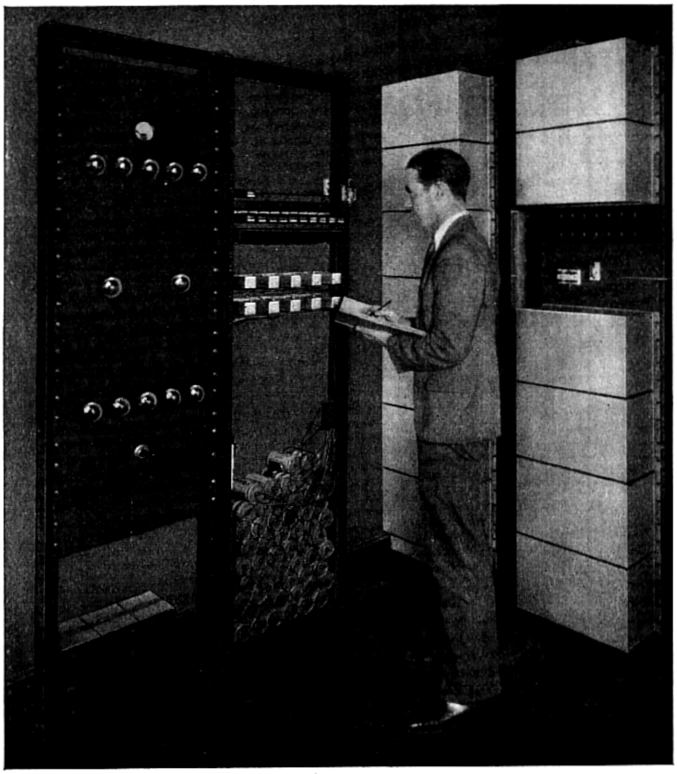

Fig. 14—Photograph of the level analyzer.

being required for resetting the apparatus. In Fig. 14 an observer is shown reading the message registers after a test has been taken. The breakdown tubes are seen at the left and the filters at the right mounted on relay racks.

It is thus seen that with this apparatus 1000 observations may be recorded on a four minute conversation, the final results being read directly from the series of counters.

By the use of this and similar apparatus the following results have been obtained. The average conversational speech power is 10 microwatts or 100 ergs per second. About 1/3 of the time no sound is flowing due to the pauses and the stops to form consonants so that the average conversational speech power is about 50 per cent higher than this value if the silent intervals are excluded. Some of the speakers will use a greater and some a lesser speech power than this average. In Table I are shown the results with a large number of speakers. It will be seen that about 7 per cent of the speakers will use in conversation average powers less than 1/16 the average while about 4 per cent will use powers which are from 4 to 8 times as much as the average. This value of 10 microwatts per second is of course for average conversational intensity. When one shouts as loudly as possible, this average speech power is raised about 100 fold and when one whispers about as softly as possible and still produces intelligible speech, it is reduced to about 1/10,000.

TABLE I

RELATIVE SPEECH POWERS USED BY INDIVIDUALS IN CONVERSATION

Region of Average Speech Power	below 1/16	1/16 to 1/8	1/8 to 1/4	1/4 to 1/2	1/2 to 1	1 to 2	2 to 4	4 to 8	above 8
Per Cent of Speakers	7	9	14	18	22	17	9	4	0

For describing in greater detail the powers involved in speech, we will define the terms Mean Speech Power, Phonetic Speech Power and Peak Speech Power. They are defined as follows:

The Mean Speech Power is the average speech power within any one one-hundredth of a second period.

The Phonetic Speech Power is the maximum value of the mean speech power of a fundamental vowel or consonant.

The Peak Speech Power is the maximum value of instantaneous power over the interval considered.

It was seen from the oscillographs that the vowels have much greater phonetic powers than the consonants. Studies of these phonetic powers for average conversation have indicated that for a typical speaker they are as shown in Table II. The most powerful sound is the vowel in the word "awl" which carries about 900 times as much power as the weakest sound which is th as in thigh. This most powerful vowel when intoned without emphasis is about 50 microwatts. The relative position in this table depends upon the emphasis given. An emphasized syllable has about three times as much syllabic power as an average one and as will be seen from the table this is about the range of powers among the different vowels.

TABLE II

o′	680	ū	310	ch	42	k	13
a	600	i	260	n	36	v	12
o	510	ē	220	j	23	*th*	11
a′	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

An analysis of a few oscillograms such as we first considered for determining the peak powers was made and showed that the peak powers are from 10–20 times the phonetic power. It is thus seen that when the vowel in the word "awl" is emphasized, the peak power is from 50 to 200 times the average speech power. To find how frequently these peak powers occur, the apparatus described above using the glow discharge tube circuits was used. The results obtained are shown in Table III.

TABLE III

Per Cent of 1/8 Second Intervals	Number of db the Peak Power in the Interval is Above the Average Level
2	above 20
3	18 to 20
6	16 to 18
8	14 to 16
10	12 to 14
11	10 to 12
11	8 to 10
10	6 to 8
8	4 to 6
6	2 to 4
4	0 to 2
21	Below the average

These values confirm earlier results obtained by oscillographs and give a much more detailed picture of the variation of the peak values as the speech proceeds. About 2 per cent of the time the peak power in 1/8th second intervals exceeds the average power level by 20 db; that is, it is more than 100 times greater. It is seen that a system designed to transmit conversational speech of the best quality should be capable of handling at least 1000 microwatts instead of 10 microwatts. It is also seen that the most frequently occurring peak is at about 10 times the average speech power. For 21 per cent of the time the peaks are below the average level. A large number of the 1/8th second intervals in this class are silent.

To find how the speech powers are distributed throughout the pitch range similar measurements were made introducing successively each one of the 14 band filters as indicated in Fig. 12. These bands were arranged so as to cover about 1/2 octave pitch range except at the two lower octaves where they cover a complete octave. From the measurements on the average speech power in each band the curves in Fig. 15 were constructed. They give the results for average conversational speech for both men's and women's voices. The ordinates are such that the fraction of the total power F which is carried by any pitch interval between P_1 and P_2 is given by

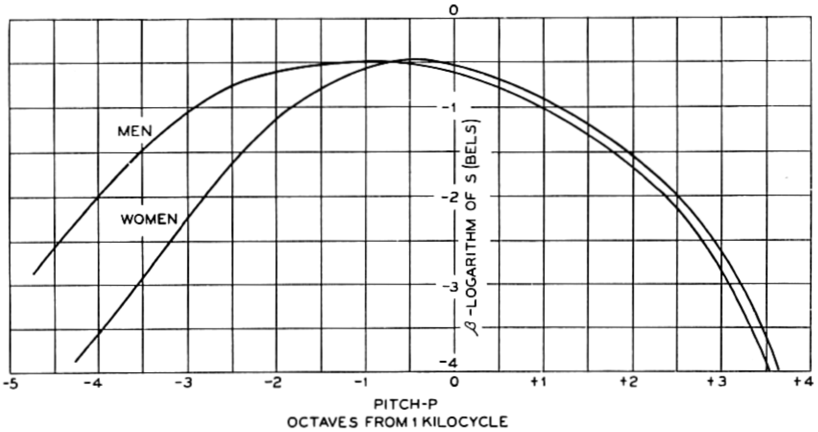

Fig. 15—Distribution function for conversational speech.

$$\text{Fractional energy} = \int_{P_1}^{P_2} S dP.$$

$$F = \int_{P_1}^{P_2} 10^{\beta} \cdot dP. \tag{7}$$

In other words β is the intensity level per octave expressed in bels. For example, the octave containing the most energy in men's voices is -1.75 to $-.75$ and it contains about $10^{-.5}$ or 31 per cent. The octave below -3 contains about 4 per cent and the octave above $+1$ about 5 per cent. For women's voices these figures are 31 per cent for the most intense region, which is the octave from $-.85$ to $+.15$, and .2 per cent and 7 per cent, respectively, for the other two octaves.

Audible Pitch Limits

The audible pitch limits for conversational speech received at various intensities are determined in the following way. It is seen from Table III that the peak power exceeds the average power by 17 db 10 per cent of the time. The loudness of speech near the threshold is probably determined by these louder components. For convenience the term "effective intensity level" will be used when speaking of these components only. With this nomenclature the effective intensity level is 17 db above the average intensity level. Using these figures and assuming that three-fourths of the speech power is radiated through the hemisphere in front of the speaker, then one can calculate that the effective intensity at one meter's distance will be 6×10^{-3} microwatts per square centimeter or at an effective intensity level of 22 db below one microwatt.

To determine the sensation level the pitches and intensities of the components in the vowels must be considered. A study of the frequency spectra of these vowels indicates that the loudest component contains from 1/2 to 1/5 of the total power of the vowel. From this it is concluded that the components determining the threshold are from 3 to 7 db below the effective level of the speech. The threshold of hearing for pure tones in the pitch region between -1 and $+1$ octaves is from -85 to -95 db with an average value of -91 db. Consequently, it is concluded that at the threshold the effective intensity level for the speech is approximately -86 db and the average level approximately -103 db. Since the effective level of the speech at one meter's distance was shown to be -22 db, it is seen that the sensation level at one meter's distance is 64 db. If the speech wave is uninterrupted by reflections then this level decreases 6 db when the distance between the speaker and the listener is doubled. This level will be raised or lowered in accordance with the intensity of the speaking, the variation for different speakers being in accordance with the data in Table I.

For example, using these relations one finds that the most probable average speech power used by a person in conversation is 5 microwatts. The most probable sensation level of such speech at 1 meter's distance is 61 db, at 10 meters' distance it would be only 41 db and could be brought back to level of conversational speech at one meter's distance only by the speaker shouting as loudly as possible.

If we use the peak voltmeter as shown in Fig. 13 and make measurements upon the peaks in 1/8th second intervals in each of the half octave bands the results will be as represented by the curves of Fig. 16.

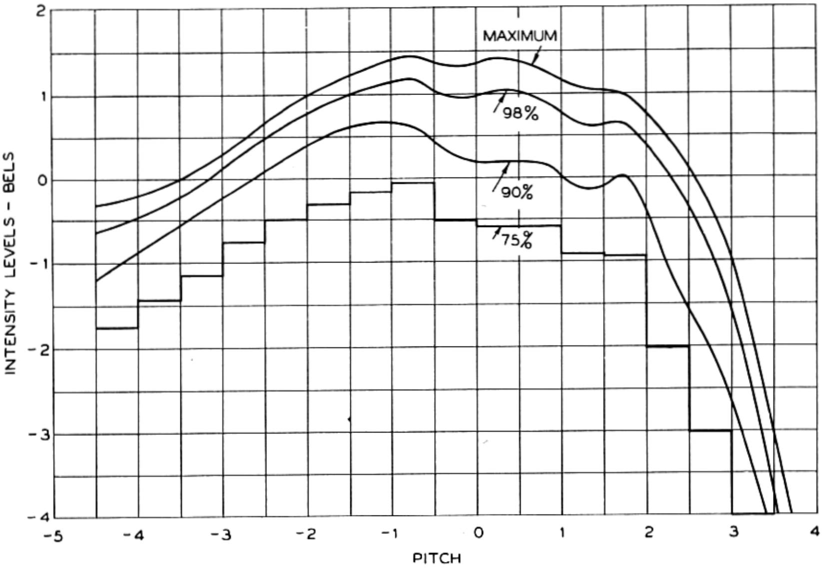

Fig. 16—Peak levels for conversational speech (3 male voices), using ½ octave average pitch intervals.

The top curves give the maximum level of the peak compared to the average intensity. The other two give levels such that the peak levels are below them 98 per cent, 90 per cent or 75 per cent of the time. It will be seen that the most intense peaks occur in the pitch range of − 1 to + 1 octaves. In this pitch range the intensity levels of the maximum peaks for the different components are approximately the same, being 13 or 14 db above the average speech level.

It is interesting to note that in the higher pitch range the curves in this figure are more widely separated than in the lower pitch range. This illustrates an important characteristic of speech, namely, that although components in the pitch range from zero to 2 octaves occur which are just as intense as those in the lower range, they occur less frequently. In other words, the spread in the intensities of the components which are successively occurring as the speech proceeds is very much greater in the higher pitch regions.

As shown above, the threshold is determined for conversational speech when the average speech level is at a − 103 db. For the same reason that only 10 per cent of the peaks having the highest levels determined the threshold for the speech as a whole, the curves labelled 90 per cent of this figure can be used as a basis for determining the sensation level in each of the bands. When the ear of the listener is 10 centimeters from the mouth of the speaker the sensation level will be 84 db and the average intensity level will be − 19 db. If α_0 is the average threshold level for tones in each of the half octave bands, then, if we subtract $\alpha_0 - 19$ from each ordinate of the curve in Fig. 16, we will obtain the sensation level of each half octave band. A curve constructed in this way will be called an audibility curve and is given in Fig. 17. This curve is for the case when the lips of an average male speaker are 10 centimeters from the ear of an average listener. It will be seen that the half octave bands above 3.25 octaves and below − 4.25 octaves are just audible. If the distance between speaker and listener is increased to one meter, which is the most commonly used distance, then the audibility curve would be one which is lowered 20 db from that one shown in Fig. 15 and the audible limits would be + 3 and − 3.5 octaves, corresponding to frequencies of

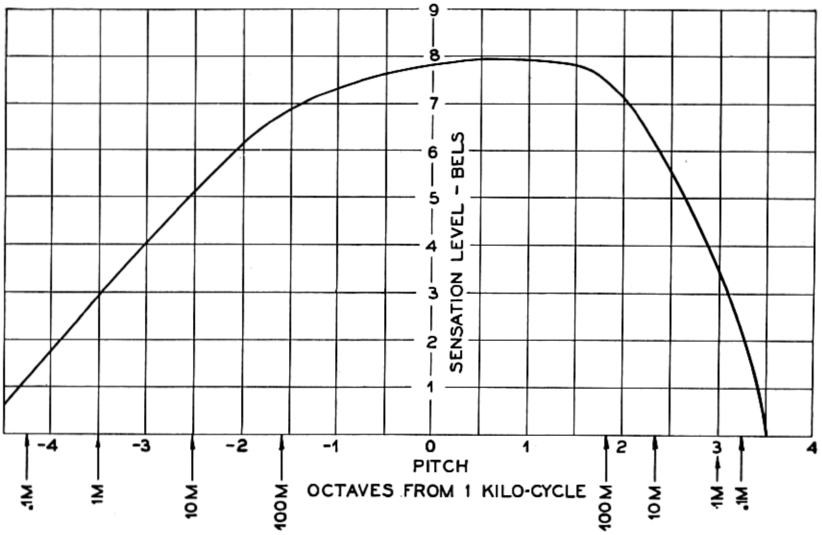

Fig. 17—Speech audibility curve (male voices).

8000 c.p.s. and 90 c.p.s. Similarly, if the distance is increased to 100 meters, the limits will be found to be + 1.85 and − 1.55 octaves. These relations are true only when no other sounds are present. Similar limits are easily determined when the listener is in the presence of any other sound whose noise audiogram is known. In that case, the ordinates in the audibility curve are reduced by an amount equal to the corresponding ordinate in the noise audiogram.

These values are such that any half octave by itself within the pitch limits will transmit audible sounds. This does not necessarily imply that, when the undistorted speech is acting upon the ear, such a half octave will transmit sounds whose presence can be detected. To test this point several observers listened to speech reproduced by a high quality loud speaker system which would reproduce all frequencies from 40 to 15,000 uniformly and into which filters could be introduced. These filters limited at desired cut-off positions the upper and lower frequencies which were reproduced.

A large group of observers then listened to this reproduced speech and they were asked to judge which was filtered and which was unfiltered. The results of such tests are shown in Fig. 18. The ordinates give the per cent of correct observations and the abscissæ the cut-off frequency of the filter. Taking a 60 per cent correct judgment as a criterion for determining the detectable pitch limits, then it will be seen that the lower limit is − 3.5 octaves and the upper limit 3.25 octaves for male speech which agrees with the results taken from the audibility curve established directly from power measurements upon speech and the threshold of hearing as described above. For female speech the limits are − 2.9 and + 3.4 octaves. Summarizing, then, it is seen that the most powerful components carrying conversational speech, which are of any practical importance, are about 4000 or 5000 microwatts while the principal components in

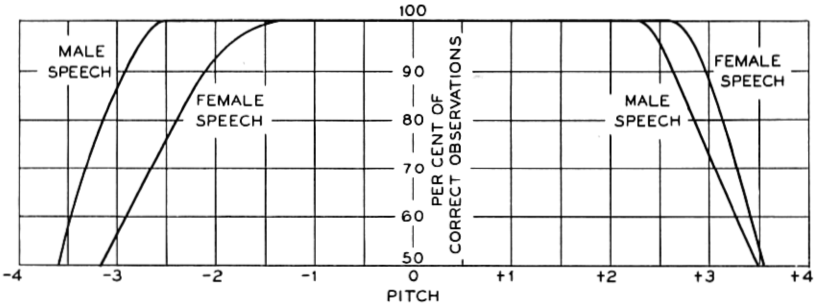

Fig. 18—Audible pitch limits for conversational speech.

the weakest sound carry only about 1/20th of a microwatt. Even for an extremely loud shout or for the most intense singing the maximum power will not exceed more than about 100 times these values; that is, they will not exceed 1 watt. The pitch range necessary for faithfully transmitting men's and women's speech is from − 3.5 to + 3.3 octaves or from 90 to 10,000 cycles per second.

Acoustical Power Produced by Musical Instruments

Now we will look briefly at some of the same results obtained for music by the use of some of these same measuring tools. In Fig. 19 are shown typical waves produced by the clarinet. A complete oscillogram of the waves produced when the instrument played its full range of three octaves on the chromatic scale was taken. The simple waves shown in the figure are those corresponding to the major triad in each of these octaves. The entire record was about 250 feet long. Such musical tones have a much more uniform wave form than those from the voice.

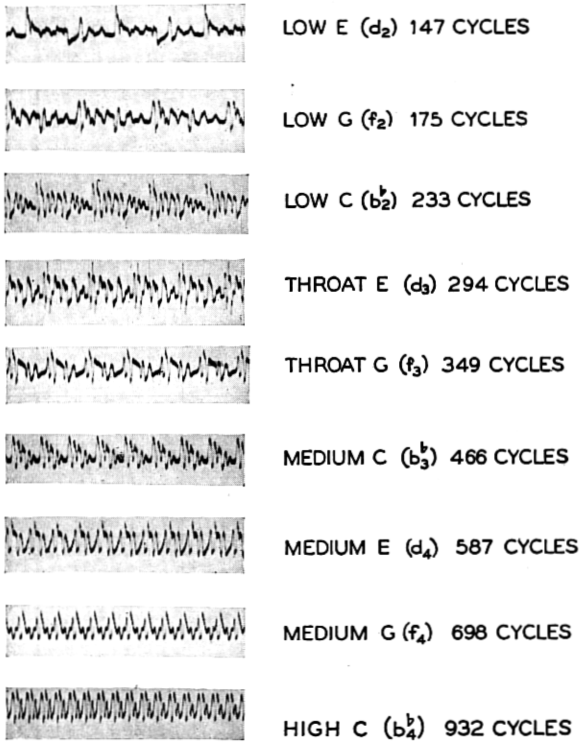

Fig. 19—Major triads of B-flat clarinet.

The measurement of the peak power from typical musical instruments used in an orchestra gave the following results.[6]

TABLE IV

Peak Power of Musical Instruments (Fortissimo Playing)

Instrument	Peak Power in Watts
Heavy Orchestra	70
Large Bass Drum	25
Pipe Organ	13
Snare Drum	12
Cymbals	10
Trombone	6
Piano	0.4
Trumpet	0.3
Bass Saxophone	0.3
Bass Tuba	0.2
Bass Viol	0.16
Piccolo	0.08
Flute	0.06
Clarinet	0.05
French Horn	0.05
Triangle	0.05

The most powerful single instrument is the bass drum which gives powers which exceed 25 watts in successive 1/8th second intervals about 6 per cent of the time it is being played. A 75-piece orchestra

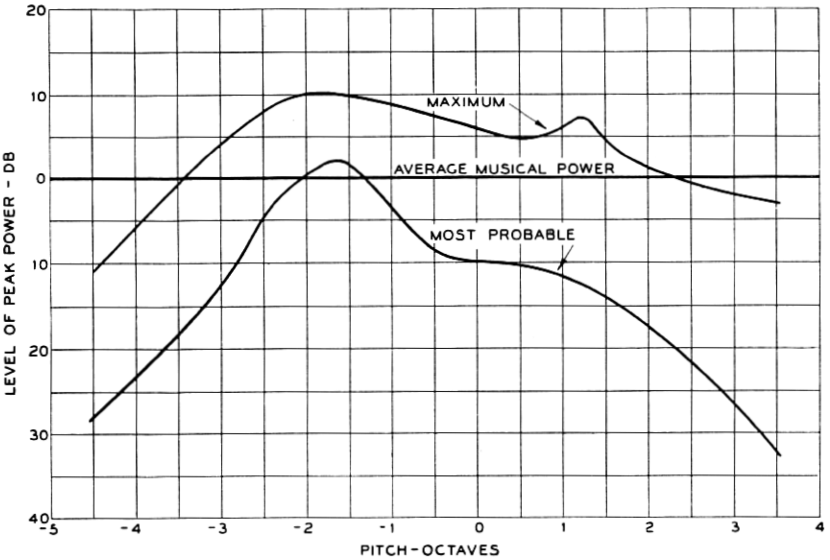

Fig. 20—Maximum and most probable peak levels for a 75-piece orchestra.

---

[6] These results and those in Fig. 19 were taken from a paper by Sivian, Dunn and White entitled "Absolute Amplitudes and Spectra of Certain Musical Instruments and Orchestras," *Jour. Acous. Soc. of America*, Jan., 1931.

playing with full volume will produce peak acoustic powers as great as 70 watts.

When such an orchestra played the four different selections, the maximum peak powers varied from 8 to 66 watts, but the average powers were .08, .07, .07 and .13 watts, respectively. Hence the variation of the average power from selection to selection was much less than that of the peak power. Both the peak powers and also the average powers for the orchestra are about 10,000 times the corresponding powers for conversational speech. In Fig. 20 the curves show how the peak power was distributed among the different pitch bands for this 75-piece orchestra. The curves give the average values for the four selections. The zero line corresponds to a power of approximately 1/10th of a watt. The levels correspond to that which was obtained in the half octave band acting alone. Although the maximum peak was 70 watts for the unfiltered music when the heaviest piece was being played, the most probable peak value in any half octave band is less than 1/10 of a watt except for the octave between -2 and -1 octaves, where it is slightly higher than this value. The distance between the two curves increases as you go to either side of this octave which is approximately that between middle "C" and the "C" above it. This indicates that the components in this region are more nearly alike in intensity and occur more frequently than in the other regions. The top curve indicates that from the standpoint of maximum peak values the half octaves from $-2\frac{1}{2}$ to $+1\frac{1}{2}$ octaves are all about equally important. As the pitch of a component goes below $2\frac{1}{2}$ octaves, its intensity decreases rapidly as indicated in the figure. Very intense peaks occur occasionally with frequencies as high as 10,000 or 12,000 cycles.

To find the lowest level used in orchestral music a violin player was asked to play as softly as is ever customary while playing before the public. Its average power was found to be about 4 microwatts. It is thus seen that the peak power from a large orchestra is about 20,000,000 times the average power produced by soft violin playing.

AUDIBLE PITCH LIMITS FOR MUSICAL SOUNDS

Measurement of the detectable pitch limits was determined in a way similar to that described for conversational speech. The results [7] for typical musical instruments are shown in Fig. 21. For comparison the results for speech and some common noises are also included. It will be seen that the lower limit for music is determined by the bass

---

[7] A more comprehensive report of this work will soon be given in a paper by W. B. Snow.

tuba, the bass viol, and the kettle drum, and its value is about 40 c.p.s. The upper limit is determined by the snare drum, the violin, and the cymbals, and is shown to be about 15,000 c.p.s. Summarizing, then, for music the range of pitches covered by the components is from − 4.7 to + 3.9 octaves, corresponding to the frequency range from 40 to 15,000 cycles per second. The intensity ranges from about 70 watts to 4 microwatts, corresponding to an intensity level range of 73 db going from the average level of the softest violin playing to the peaks in the heaviest playing of a full 75-piece orchestra.

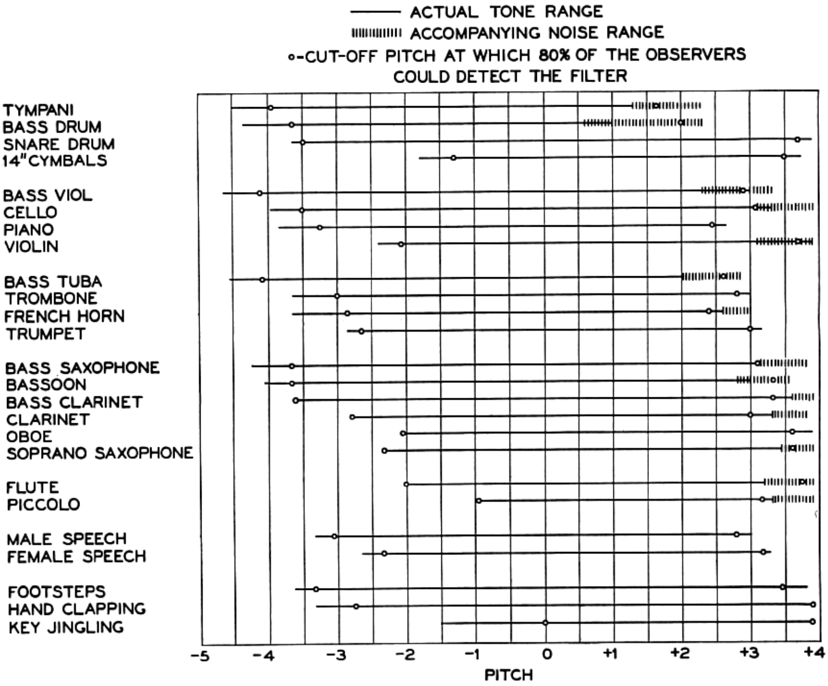

Fig. 21—Audible pitch range for speech, music and noise.