The Bell System Technical Journal

July, 1926

The Power of Fundamental Speech Sounds

By C. F. SACIA and C. J. BECK

SYNOPSIS: This paper describes the continuing work on speech power by means of oscillographic studies of vowels, semi-vowels and consonants. A previous paper considered the characteristics of a few individual sounds from the power standpoint, but the principal emphasis was placed upon speech as a whole. In this later analysis, sounds are considered individually on the basis of instantaneous and mean power. A practical application of the results is suggested.

CONTINUING the work done on speech power by means of power oscillograms,[1] we have made additional reductions in the data relative to the vowels, semi-vowels and consonants and have also prepared a smaller amount of data on the power of the semi-vowels and the consonants from the amplitude oscillograms.[2] This is a preliminary study of the subject, at least in so far as the latter two classes of sounds are concerned, for these records of speech sounds were made to show all sounds in their true relative value hence the consonant sounds, being greatly inferior to the vowels were measurable to a correspondingly smaller degree of accuracy. We have gathered such data as the existing records could yield before future plans are completed to make a more comprehensive study of consonants.

Stop consonants are not so well characterized by the power data as are other types. The unvoiced stop consonants have two properties: a puff whose main frequency component is of the order of 50 cycles with a few ripples of high frequency; and a modifying effect upon the beginning or end of the vowel which immediately precedes or succeeds it. Hence, such a consonant is more of a controlling factor and lacks the essential properties of a discrete sound. In giving the data on the puff where it is measurable, we separate the low and high frequency components. In the case of the voiced stop consonants the vocal cord vibrations give the consonant more character of its own.

Mean Power and Peak Power

In the paper on speech power and energy, the "mean power," P_m, was derived (in the case of the vowel sounds) as the mean of the power taken throughout the interval of the vocal cycle. By the assumption of an appropriate arbitrary interval instead, say of the order of one

[1] B. S. T. J. Vol. IV No. 4. "Speech Power and Energy," by C. F. Sacia.
[2] B. S. T. J. Vol. IV No. 4. "Sounds of Speech," by I. B. Crandall.

one-hundredth of a second, the definition applies as well to consonant sounds and in addition has the same practical significance as that of the mean power of a vowel.

Mean power is thus a variable function of time, starting from zero, rising to a maximum and eventually falling to zero again as the sound is being uttered.[3] In studying an aggregate of speech sounds it is impracticable to have the final results in terms of these mean power curves; the most important discriminant of such a curve of any sound is its maximum ordinate, P_m. This value was used in the earlier study and has been given the name "syllabic power" when used in connection with the syllable as a whole. In the present case we shall abbreviate by simply calling it the "mean power of the sound." Similarly, when we are considering the consonant apart from the rest of the syllable we select the maximum value of P_m for that consonant.

Likewise, in considering the instantaneous power of a sound we select the height of the greatest peak occurring therein and for convenience we call it the "peak power."

All the averages hereinafter tabulated are the arithmetic averages of such maximum ordinates and not the integrated averages.

Normal and Conversational Values

We specify "normal" values as those derived from monosyllables spoken disconnectedly without accent but also without being slighted; while "conversational" values are derived from ordinary conversational speech. It does not follow that the arithmetic average of conversational values for a given sound should equal the average of the normal value, for the reason that some sounds are slighted much more frequently than others, as we shall see later.

The Consonants and Semi-Vowels

Of these sounds two independent sets of data are available: instantaneous peak power and mean power. The former is summarized in Table I. To explain the table in detail we take as an example the consonant, "t" as in "tap." There being one observation upon each of two speakers, the greatest observation showed 19 microwatts (peak) from the lips of the one speaker while the other speaker reached a peak of 13 microwatts, and the average of these two is 16. As in the paper on Speech Power and Energy, the corresponding values of power intensity in microwatts per square centimeter at the condenser transmitter are given in the group at the right. Since the relating factor is

[3] See "Speech Power and Energy," Fig. 1, page 628, for comparison of instantaneous and mean powers.

TABLE I

Normal Values of Peak Power in Microwatts for Two Speakers

(A) CONSONANTS

Consonant		Total from Voice			Per Cm² at Trans.		
Symbol	Key	Max.	Min.	Ave.	Max.	Min.	Ave.
b	bat	7	7	7	0.06	0.05	0.06
p	pot	7	6	6	0.06	0.05	0.05
*p	pot	128	0	64	1.04	0.	0.52
d	dot	7	1	4	0.06	0.01	0.04
t	tap	19	13	16	0.15	0.11	0.13
g	get	9	7	8	0.07	0.06	0.06
k	kit	9	4	6	0.07	0.03	0.05
dh	that	10	8	9	0.08	0.06	0.07
th	thin	1	0	1	0.01	0.	0.01
*th	thin	30	0	15	0.24	0	0.12
v	vat	29	21	25	0.23	0.17	0.20
*f	for	53	10	31	0.42	0.08	0.25
f	for	4	2	3	0.04	0.02	0.03
j	jot	26	23	24	0.21	0.19	0.20
ch	chat	61	43	52	0.49	0.35	0.42
zh	azure	53	23	38	0.43	0.19	0.31
sh	shot	133	97	115	1.08	0.79	0.93
z	zip	42	21	31	0.34	0.17	0.25
s	sit	54	8	31	0.43	0.06	0.25

* Low frequency puff.

(B) SEMI-VOWELS

Semi-Vowel		Total from Voice			Per Cm² at Trans.		
Symbol	Key	Max.	Min.	Ave.	Max.	Min.	Ave.
l	let	226	37	131	1.83	0.29	1.06
ng	ring	169	25	97	1.36	0.20	0.78
n	no	74	21	47	0.59	0.17	0.38
m	me	198	23	111	1.60	0.18	0.89

NOTE: For these two speakers, the peak power of the succeeding vowel was as follows:

	Total	Per Cm²
ū (tool)	206	1.7
á (tap)	860	6.8
ē (teem)	241	1.9

about 127, the intensities 0.15, 0.11 and 0.13 are the first three numbers respectively divided by 127.

These values were derived by measuring the amplitudes of the above-mentioned oscillograms of the acoustic pressure. The maximum or peak amplitudes of the consonant and the succeeding vowel were first measured; the square of the ratio between these is the ratio of the corresponding peak powers. Now the approximate peak powers of these vowels for the two speakers were found (see note under Table I) from the power oscillograms used in our study of speech power. Hence from the product we derive the approximate peak power of the consonant (or semi-vowel). Direct measurement of peak power from the latter oscillograms was impracticable because of the low sensitivity of the instantaneous power recorder [4] and the before-mentioned fact that the power of the consonants and semi-vowels is low relative to that of the vowels.

Since frequencies of the order of 50 cycles are of negligible importance in speech, the 50-cycle puff has been separated from the other components in the case of the unvoiced stop consonants. This is justified by the fact that the utterances of such a sound by two speakers may seem exactly alike to the careful listener, whereas a large puff may be present in one case and none in the other.

The values thus far considered represent "normal" values in speech —not accented and yet not slighted.

TABLE II

Conversational Values of Mean Power in Microwatts for 16 Speakers

(A) CONSONANTS

Consonant		Speaker's Power		Number of Measurable Observations	Per Cm² at Trans.	
Symbol	Key	Max.	Av.		Max.	Av.
d	dot	2.9	0.08	4	0.023	0.0006
t	tap	6.0	0.14	14	0.049	0.0012
k	kit	4.8	0.34	20	0.039	0.0027
v	vat	2.4	0.03	1	0.019	0.0002
f	for	3.6	0.08	1	0.029	0.0006
j	jot	3.6	0.47	8	0.029	0.0038
ch	chat	7.9	1.44	19	0.064	0.0116
sh	shot	6.0	1.83	9	0.049	0.0148
z	zip	7.2	0.72	31	0.058	0.0058
s	sit	8.7	0.94	115	0.070	0.0076

(B) SEMI-VOWELS

Semi-Vowel		Speaker's Power		Number of Measurable Observations	Per Cm² at Trans.	
Symbol	Key	Max.	Av.		Max.	Av.
l	let	9.6	0.33	13	0.078	0.0026
ng	ring	3.6	0.35	2	0.029	0.0028
n	no	18.0	2.11	146	0.145	0.0170
m	me	16.8	1.85	31	0.136	0.0149

4 In recording the power, separate vibrators had been used for instantaneous and mean powers.

Our measurements of mean power, on the other hand, were made from power records of conversational speech, with a greater variety of observations and speakers. Stress, therefore, plays an important part here.

In Table II is given a compact summary of the direct measurements made on the power oscillograms. Thus consider "d" as in "dot." 2.9 microwatts was the greatest observed value for any speaker, while the average of all observations (including accented and unaccented utterances) was but 0.08. Only four observations, however, were large enough to be measured. As before, we give the corresponding intensities in microwatts per square centimeter at the transmitter in the next two columns.

To show the occurrence of stress in the utterance of these sounds in ordinary speech, we give in Fig. 1 the stress frequency-distribution curves [5] of several oft-occurring sounds. These curves are derived in the same manner as were the syllabic stress curves in the study of speech power. They exhibit the marked degree in which the consonants differ in stress for ordinary speech. For example, among the consonant sounds, "t" and "sh" represent extreme types. The former is either slighted or strongly accented with but little intermediate gradation while the blunt characteristic of the latter indicates the most nearly uniform distribution of stress into all shades from zero to maximum. Similarly with the three semi-vowels shown, "l" and "m" are extreme types.

The Vowels

Some attention was given to vowel power in the other paper where under the heading of "Relative Power of Vowels" (on page 634) were charted what we have classified as normal values of mean power. These were derived from the mean power curves of disconnected monosyllables. Although they were charted separately for male and female voices, we shall not differentiate between the two in the following. In Tables III and IV are summarized the four sets of data based upon the speech from 16 voices. Here we see the influence of stress by comparing the conversational and normal values. This effect is noteworthy in the case of "o" (ton) "à" (tap) and "i" (tip) which average considerably less power in conversational speech than in normal syllables. Another point of interest is the comparison of peak and mean values. For example, in the normal data, the ratio of peak to mean (i.e. the

[5] The abscissa represents the relative number of observations ($s/\underline{s}$) whose relative power values exceed the magnitude of the ordinate, n, a numeric varying between zero and one.

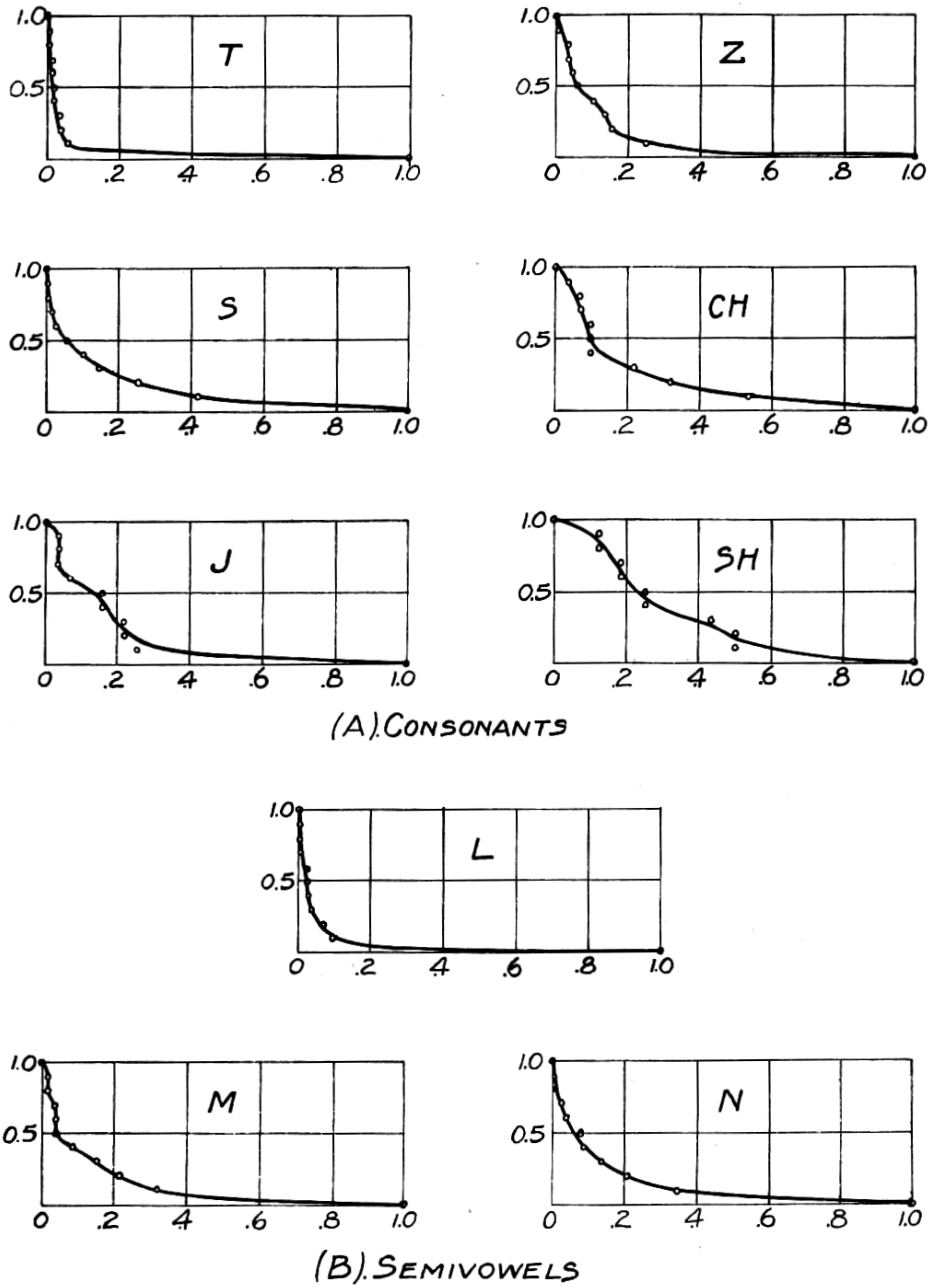

Fig. 1. Power Stress Curves.

TABLE III—VOWELS

Peak Power in Microwatts for 16 Speakers

Vowel		Total from Voice				Per Cm² at Trans.				Number of Measurable Observations Conversational Values
		Normal Values		Conversational Values		Normal Values		Conversational Values		
Symbol	Key	Max.	Av.	Max.	Av.	Max.	Av.	Max.	Av.	
ū	tool	620	290	760	180	5.0	2.3	6.1	1.5	61
u	took	890	470	—	—	7.2	3.8	—	—	0
ō	tone	1310	540	900	330	10.6	4.4	7.2	2.7	62
ò	talk	1240	630	1580	600	10.0	5.1	12.8	4.8	32
o	ton	1240	600	1720	300	10.0	4.8	13.9	2.4	248
a	top	1650	760	1580	660	13.3	6.1	12.8	5.3	127
à	tap	1860	1020	1380	290	15.0	8.2	11.1	2.3	38
e	ten	1720	660	1510	340	13.9	5.4	12.2	2.8	125
ā	tape	1380	580	1720	470	11.1	4.7	13.9	3.8	32
i	tip	1240	520	1330	190	10.0	4.2	10.8	1.5	198
ē	teem	1510	430	960	190	12.2	3.5	7.8	1.5	56
r̄	err	—	—	550	200	—	—	4.4	1.6	33

Note: The dash indicates that observations were not available.

TABLE IV—VOWELS

Mean Power in Microwatts for 16 Speakers

Vowel		Total from Voice				Per Cm² at Trans.				Number of Measurable Observations Conversational Values
		Normal Values		Conversational Values		Normal Values		Conversational Values		
Symbol	Key	Max.	Av.	Max.	Av.	Max.	Av.	Max.	Av.	
ū	tool	60	33	53	13	0.49	0.27	0.43	0.11	64
u	took	108	40	—	—	0.87	0.32	—	—	0
ō	tone	82	38	68	22	0.66	0.31	0.55	0.18	64
ò	talk	91	43	125	47	0.74	0.35	1.01	0.38	32
o	ton	84	33	107	15	0.68	0.27	0.86	0.13	284
a	top	111	48	130	34	0.89	0.39	1.05	0.28	128
à	tap	96	40	40	9	0.78	0.33	0.32	0.07	48
e	ten	79	27	88	17	0.64	0.22	0.71	0.13	141
ā	tape	58	26	62	20	0.47	0.21	0.49	0.16	32
i	tip	53	30	55	9	0.43	0.24	0.44	0.07	250
ē	teem	65	27	78	12	0.52	0.22	0.63	0.10	64
r̄	err	—	—	30	10	—	—	0.24	0.08	40

Note: The dash indicates that observations were not available.

square of the peak factor) is greater for centrally located vowels and is greatest for "à" (tap) as was mentioned in the earlier paper. Referring to the normal values of peak power we find a surprising degree of regularity in the increase of these values from a minimum for "ū" (tool) to a maximum for "à" (tap) and the falling off again to minimum for "ē" (teem). The one slight irregularity is the vowel "o" (ton). (We have omitted "r̄" (err) from this comparison because it has no well defined place on the Vietor triangle which forms the basis for this arrangement of the other vowels).

TABLE V—SPEECH SOUNDS

Speech Sound	Key	Relative Power, Arbitrary Units		C Relative Power Attenuation to give 80% Articulation
		A Mean Power Conversational values for 16 speakers	B Peak Power Normal values for 2 speakers	
ò	talk	1870	688	826
a	top	1380	1430	474
ō	tone	875	630	619
ā	tape	808	632	567
e	ten	664	975	364
o	ton	616	688	474
ū	tool	532	344	349
ē	teem	484	402	421
r̄	err	384	– see note	924
à	tap	366	2170	645
i	tip	346	688	295
n	no	84	78	36
m	me	74	185	38
sh	shot	73	192	216
ch	chat	58	87	64
s	sit	38	51	11
z	zip	29	52	17
j	jot	19	41	98
ng	ring	14	162	134
k	kit	14	10	43
l	let	13	218	157
t	tap	6	26	32
d	dot	3	7	60
f	for	3	6	9
v	vat	1	41	13
u	took	– see note	688	347
zh	azure	–	63	–
dh	that	–	15	–
g	get	–	13	60
b	bat	–	11	30
p	pot	–	11	24
th	thin	–	1	1

NOTE: The dash indicates that observations were not available.

Relative Power of Speech Sounds

A direct comparison of most of the fundamental sounds will now be made. In Table V—A are shown the conversational values (averaged) of the mean power for each sound for 16 speakers. The units are taken arbitrarily in order to show only the relative values. As might have been expected, the vowels rank the highest, the semi-vowels next and the consonants the lowest, although we find a few consonants interspersed among the semi-vowels. In Table V—B is the similar arrangement for the normal values of peak power for the two speakers. Data on a larger number of sounds are available for this group, but the same general order prevails: vowels, semi-vowels and consonants. Minor differences in order (note "v" as in "vat") may be expected to occur because of the influence of stress upon the conversational value. But in both cases the ratio of the maximum to the minimum is of the order of 2000. This similarity is striking in view of the difference in the modes of utterance and the numbers of speakers in the two cases.

Finally, in Table V—C are shown relative values[6] derived on the basis of relative attenuation in power required to bring the articulation (as judged by the average ear) to 80%. Since disconnected monosyllables

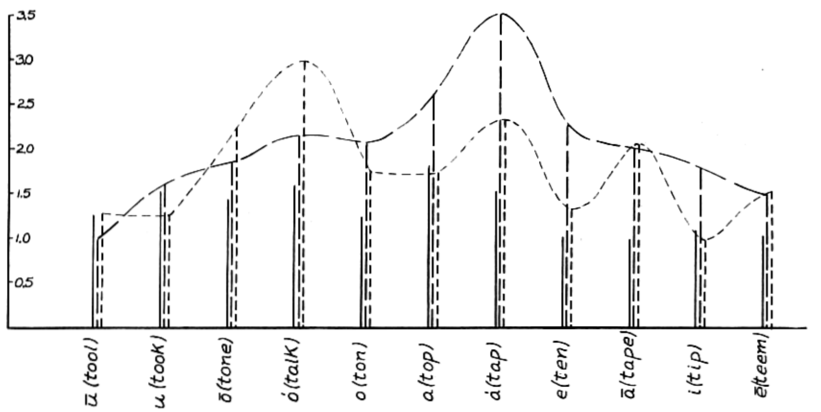

Fig 2. Comparative Chart Relative Normal Values of Vowel Sounds.
— — — — —Peak Power.
————————Mean Power.
- - - - - - - - - -Relative Power Attenuation Required to Give 80% Articulation.

were used in this test the values are normal values in our present category. Although the same general order of the other two tables

[6] Taken from the paper presented by Harvey Fletcher before the Modern Languages Association, December 1923. Values are there called relative "intensity" which term we avoid here because of the acoustic meaning already assigned to intensity: power per square centimeter.

prevails here, there are considerable differences throughout which may well be expected since the ear is used in making the balance. The frequency response characteristic of the ear is the complicating factor in this case. The ratio of maximum to minimum here is of the order of one thousand or about one-half the absolute power ratio found in the two preceding tables.

A more orderly comparison between power and "relative attenuation" exists in the case of the vowels alone as shown in the chart of Fig. 2. Thus the peak power and "relative attenuation" most nearly correspond at the ends of the chart (especially the left) where there is resonance of lower frequency in the vowels. The vowel "o" again shows a peculiarity in that the two trends—as shown by the envelopes—intersect here. Peak power predominates over "relative attenuation" in the three successive vowels "a," "à," "e," which have strong resonance in the region from 600 to 1200 cycles. The vowel "i" gives the only erratic turn in this comparison, differing considerably from the two adjacent vowels.

As for loudness in the ordinary sense, let us note a phenomenon of rather common occurrence in these days of good quality sound reproducing apparatus. One may be listening to well reproduced speech at ordinary volume when suddenly a slightly accented syllable containing "à" (tap) comes through with noticeable overload distortion and its accompanying disagreeable effect upon the ear. Although the listener does not judge this sound to be any louder than numerous accented sounds preceding and following it, still the fact remains that there has been considerable overload due to the peaks of the wave being cut off by the amplifier. Where do we look for the explanation? As noted in the earlier paper this vowel has the highest peak factor, and we have already seen in Table III that it normally contains the greatest peak power. In spite of this therefore, it would seem that the loudness of this sound does not predominate over the loudness of the sounds in the first half of the chart, as does the peak power. This phenomenon can also be demonstrated, for the vowel "ē" (teem) and to a lesser degree even for the vowels which intervene between these two in the tables and chart of the vowel sounds.