A Statistical Basis for Objective Measurement of Speech Levels

By PAUL T. BRADY

(Manuscript received March 15, 1965)

First-order probability distributions of speech amplitudes are studied to establish a theoretical basis for obtaining a measure of speech level. The logarithm of the long-term waveform of the speech envelope is found to be approximately uniformly distributed above a threshold. The average peak level (apl) is obtained by taking the time average of the log of the envelope waveform and deriving from it the peak of the log-uniform distribution which would have produced the same average. A theoretical analysis of various properties of the apl indicates that, within certain bounds, the apl satisfies a postulated set of requirements of an "ideal" speech level measure. A critical requirement is that the measure remain independent of the value of a threshold employed by a speech detector in the measuring device. It appears that variation in the threshold can typically change the apl by about one db.

The Digital Speech Level Meter is described as an instrumentation of the technique used to obtain the apl. Measurements made with this meter are easily obtained and very repeatable, and are in general agreement with theoretical predictions.

I. INTRODUCTION

1.1 *Object of Study*

The goal of this study is determining a speech level measure ideally having the following properties:

(1.) It is objective, and is not based on the judgment of an observer.

(2.) It is based on measurements made only while speech is present and is not influenced by long silent intervals.

(3.) It is expressible as a single number.

(4.) It varies on a db-for-db basis with attenuation or amplification of the voice signal.

(5.) It is not a function of an arbitrary convention used to take the measurement, such as the value of a threshold in a meter.

(6.) It is not influenced by singular loud transients on the voice circuit.

(7.) It is easily and reliably obtained.

The specification of the *level* of a signal implies a description of certain physical properties of the amplitude of the waveform. The *loudness* of a signal is a measure of the volume of a sound as perceived by a listener. Although it may be possible to correlate level measurements with loudness, no attempt will be made to do so in this study.

1.2 *Outline of Report*

In seeking a measurement satisfying the above requirements, an analysis is made of the statistics of speech levels as they appear above a threshold. This analysis, appearing in Section II, shows the logarithms of these levels are nearly uniformly distributed.

Section III indicates that the peak amplitude occurring in a speech sample satisfies most of the requirements listed above. It may, however, be due to some isolated event (such as coughing or a circuit transient on the voice circuit) which is not characteristic of the general speech process.*

A different measure, the *average peak level* (apl), is therefore proposed in lieu of the sample peak. The apl is a parameter of the postulated uniform level distribution of the speech sample. It is shown that if speech actually is "log-uniformly" distributed, as seems to be the case for some speech samples, the apl is equivalent to the peak. For other speech samples, it will still satisfy some of the requirements stipulated above, and will approximately satisfy the others. Since it is a measure taken over the entire sample, the apl has an advantage over the peak in that it is relatively uninfluenced by singular loud events.

Section IV shows the apl to be a better objective measure of speech levels than the volume unit (VU) presently measured, since the latter exhibits significant observer bias and variability.

Section V describes the Digital Speech Level Meter, an instrument which demonstrates a technique used to obtain the apl. Some of the measurements made with the meter are included in Section VI. These measurements are easily obtained and highly repeatable. It is emphasized that the instrument described here is only an experimental model and may be subject to many revisions before it is suitable for general use.

---

* The second highest peak, the average of the first and second highest peaks, the third highest peak, etc., also satisfy most of the requirements, but are also influenced by extraneous events. In addition, as more peaks are involved in the measurement, the mathematics for a theoretical analysis becomes intractable. Complex functions of several peaks will therefore not be considered in this paper.

II. DISTRIBUTION OF SPEECH LEVELS

2.1 *Density and Cumulative Functions*

In this report, upper case X will be used to denote a random variable and x will denote a particular value which X can assume. Only continuous functions will be studied. The *density function* will be denoted $p(x)$, and the *cumulative function*, $P(x)$. The cumulative function will be Prob. $(X \geqq x)$, the complement of the definition normally used in mathematics, but which is commonly used in speech literature.[1,2,3,4]

2.2 *Establishing a Threshold*

In order to measure speech levels only during the time when speech is "actually present," we must establish an objective indicator of intervals over which the speech waveform is to be observed. Ideally, this indicator should mark off intervals *which would retain their pattern regardless of the level at which the speech sample is played.* If such a pattern could be established then some simple statistic, such as the rms voltage (V_{rms}) measured and averaged only over the prescribed intervals, could satisfy all of the requirements in Section 1.1.

Because of the wide dynamic range of speech, it is virtually impossible to establish the required level-invariant speech patterns if noise is present. A previous study[5] dealt with this problem in some detail, however, and it was shown that on a special simulated toll circuit, a threshold of -40 dbm re OTL* is sufficiently sensitive for detecting most of the speech while avoiding noise operation. Such a threshold detection incorporates no hangover and therefore differs from a conventional speech detector. Expressed mathematically, let X be a random variable such that

$$x = 10 \log \frac{(1000)\,(v^2)}{600} \tag{1}$$

where v is the voltage representing the speech waveform and x is the equivalent level in dbm. Then the speech considered for analysis in this study will be such that

$$\text{Prob}(X \geqq -40 \text{ dbm}) = 1 \text{ (or 100 per cent).} \tag{2}$$

* Zero dbm equals 0.775-volts rms across a 600-ohm resistor and will thereby cause one milliwatt to be dissipated. Although dbm implies a power measurement, it is often used to specify a voltage without regard to power or resistance, as is done here. Zero dbm is about 2.22 db below one volt (zero dbv).

The zero transmission level (OTL) point is a point to which all level points in a telephone toll system can be referred. It is analogous to citing altitude by referring to height above sea level.

Having thus adopted an arbitrary threshold criterion, the task of this study will be to specify the level of a speech sample with a measure which will be relatively insensitive to the threshold value.

2.3 *Source Material*

All of the measurements in this study were made with 8 recorded conversations involving 4 pairs of men and 4 pairs of women. Each conversation was about 7 minutes long except for one which lasted only 3.3 minutes. The recordings were made at the OTL point of a simulated toll circuit. In addition, a "continuous speech" tape was produced by manually editing out conversational pauses, thus condensing each person's speech from 7 minutes to about 1 minute. (A more detailed description of the conversations may be found in Ref. 5.)

2.4 *Instantaneous Level Distribution*

The instantaneous level of speech is interpreted here as the absolute magnitude of the speech waveform at a particular instant of time, expressed in dbm. Shown in Fig. 1 is the computer-obtained cumulative function for the levels occurring in a 67 second sample of continuous speech from subject AD. The speech was 4-kc low-pass filtered and was

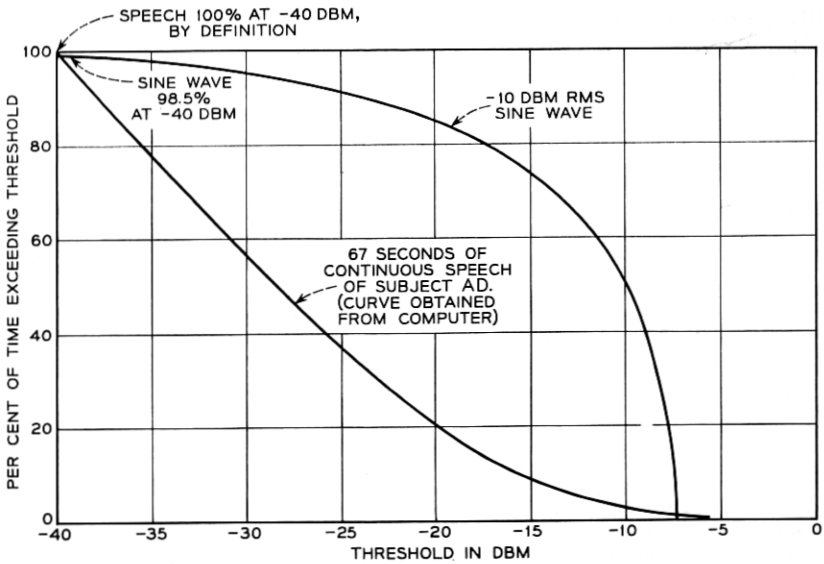

Fig. 1 — Instantaneous cumulative functions of speech and of a sine wave.

sampled at 10 kc for analog to digital conversion. Continuous speech was used because it was more economical for computer work; the original speech contained too many pauses.

The speech curve of Fig. 1 starts at 100 per cent for -40 dbm, by convention stated in (2), and then decreases almost linearly (for a major part of its range) toward a cutoff point near -10 dbm. The approximate linearity of the speech curve is of crucial importance in this study, since the level measuring technique to be described later depends on this property.

To illustrate the contrast between the speech distribution and a sinusoidal distribution, the cumulative function for the instantaneous levels of a full-wave rectified -10 dbm rms sine wave is also included in Fig. 1.

A few of the speech level distributions appearing in the literature are plotted in Fig. 2. The conversion of the original thresholds to the dbm scale is accomplished simply by transferring the shape of the literature data onto the author's graph, ignoring the absolute values of the literature thresholds. This conversion is valid since only the shapes, and not the absolute values, of the different curves will be compared. The curves of Fig. 2 are taken from Sivian,[1] Dunn and White,[2] Davenport,*[3] and Shearme and Richards.[4]

2.5 *The Log-Uniform Distribution as an Empirical Formula*

Figs. 1 and 2 indicate that all of the speech data are very similar, and that the cumulative functions can be approximately drawn as straight lines over much of their range. If the cumulative function were truly linear, then the density function would be uniform over its whole range with value $1/[\text{peak} - (-40)]$, and would be zero outside of this range. This distribution will be called the *log-uniform distribution* since the logarithm of the amplitude is uniformly distributed. In Section III certain properties of the log uniform distribution will be investigated. It is worthwhile first, however, to examine the distribution of the speech wave envelope.

2.6 *The Envelope Distribution*

Let speech be played into a full-wave rectifier, whose output in turn is applied to an RC filter having approximately equal rise and decay times. (Such a circuit is shown in Fig. 11.) The waveform at the filter output will be considered as the speech envelope. It was chosen for use

* Davenport presents *density* functions of measured speech levels, and establishes an empirical formula for the distribution. Plotted in Fig. 2 of the present report is the cumulative function of the empirical formula.

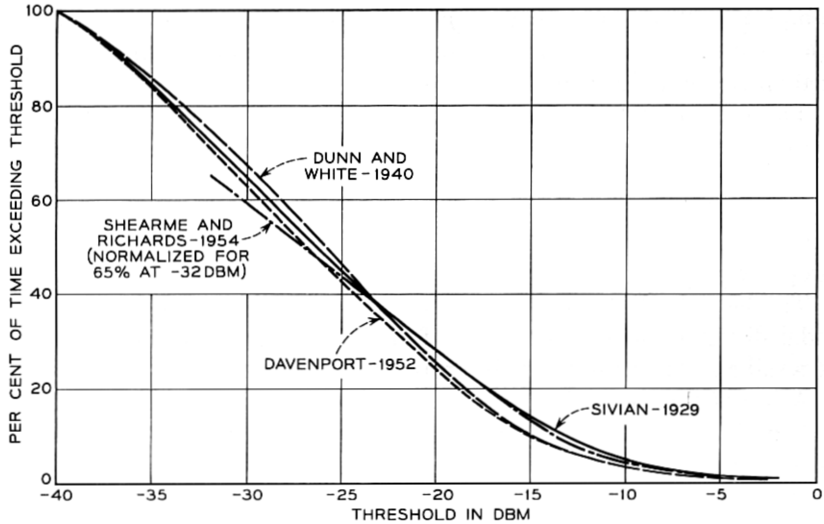

Fig. 2 — Instantaneous speech distributions from four earlier studies. The absolute levels of each curve were adjusted to the DBM scale so that the levels would be roughly equivalent to those in the present study.

in this study since it varies at a lower rate than does the original waveform and thus leads to simpler instrumentation in sampling circuits. It is shown in the next section that the envelope level distribution is similar in shape to that of the instantaneous waveform, although they differ in absolute values. Because of the similarity of distributions, the technique developed later in this paper for measuring levels would work equally well with either the envelope or the original speech waveform.

2.7 *Choice of the Time Constant*

The distribution of a speech wave envelope will depend on the choice of the time constant of the RC filter. A family of unnormalized cumulative functions for the 67 second continuous speech sample of subject AD with time constant as the parameter is shown in Fig. 3. Also included is the computer-obtained cumulative function of the instantaneous amplitudes.

With large values of RC, the speech amplitude peaks become smeared out in time, and more low level energy is evident. This is shown in Fig. 3, which is in agreement with the data of Shearme and Richards.[4] With an RC of 2.5 msec, the envelope levels are spread over almost as great a range as are the instantaneous levels. In Section III it will be shown that the level measurement should be taken with the threshold fairly close to the lower end of the linear range of the distribution. Fig. 3 shows that large values of RC compress the distribution, narrowing the allowable threshold range. An RC of 2.5 msec is chosen to avoid this difficulty.

The VU meter (discussed more fully in Section IV) is constructed so that the needle follows the speech level at a "syllabic rate" and has a time constant of about 140 msec.* The longer time constant is necessary to allow an observer to follow the meter movement; he would be at a loss to keep track of the needle if the time constant were 2.5 msec. The

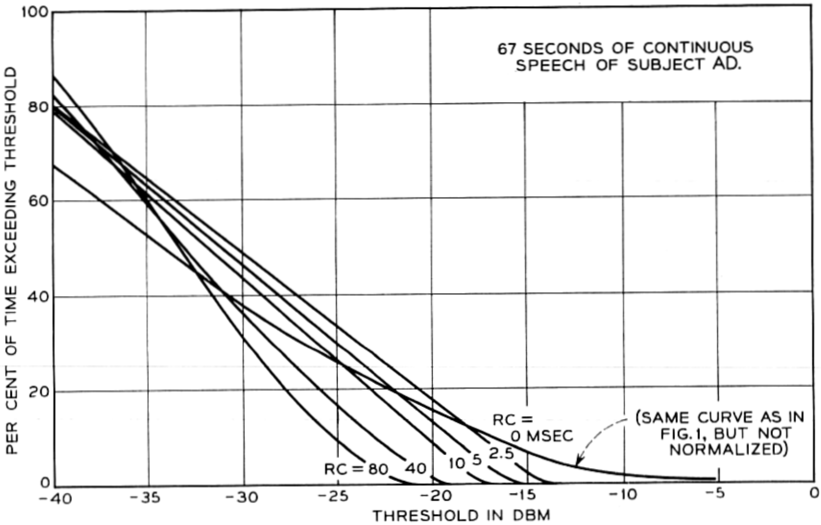

Fig. 3 — Effect of changing time constant on envelope distribution.

smaller value can be used in this study because the human observer limitation is not present.

2.8 *Results of Envelope Distribution Measurements*

The cumulative distribution of the envelope of the combined speech of all 16 talkers is shown in Fig. 4. This represents about 25 minutes of speech exceeding the −40 dbm threshold, with a total elasped "real time" of about 103 minutes. This distribution is again linear over most of the range, but a better approximation is to use two straight lines, as

* Based on measurements made of the 63 per cent rise and decay times for three different VU meters.

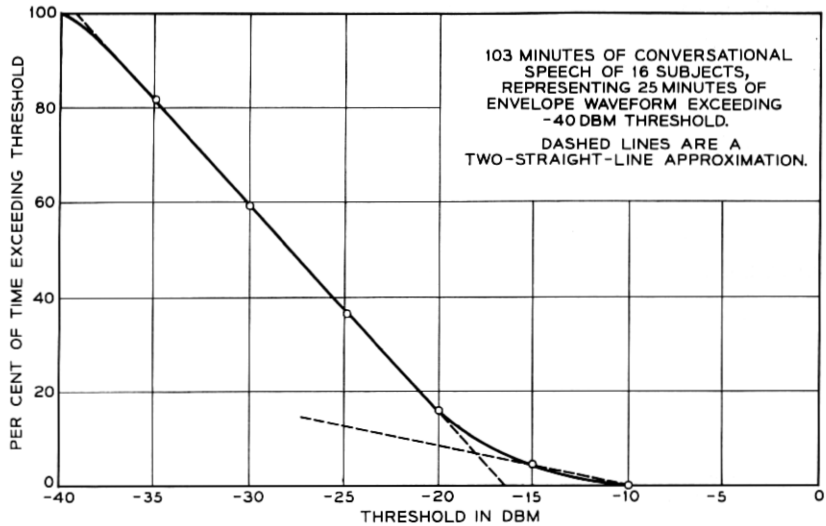

Fig. 4 — Envelope distribution of combined speech of all subjects.

shown in the figure. (The two-line approximation will be discussed in Section 3.3.)

Regarding the distributions of the individual speakers, the curves can be placed into three categories, as shown in Fig. 5. The curves of half the speakers were distributed log-uniformly; an example is the speech of NS. The speech of seven others had curves which seemed to be composed of two log-uniform functions; similar to that of BS. Most of the break points occurred very close to the bottom of the curve; the one illustrated is in fact the most pronounced case of a double valued distribution. One speaker, JM, had a noticeable downward break point near the top of the curve. This effect was also present to a very small degree in two or three of the other speakers. It will be shown in Section 3.4 that such low level break points have little effect on level measurements, and for this reason this distribution will not be considered in subsequent analysis.

2.9 *Length of Sample*

The statistical speech level measure to be proposed in this study is based on the assumption that the speech sample has an approximately log-uniform distribution. It is therefore of interest to learn: (1.) what length speech sample is required to yield a log-uniform distribution, and (2.) what length is required before the sample is representative of the long term distribution of a particular speaker. To answer these questions, the *continuous* speech tapes of four men and four women were analyzed as follows: The distribution of a one-second (real time, not time over a threshold) sample of speech for a subject was obtained. Then, a two-second sample was analyzed such that the two-second sample included the one-second sample. This was done in like manner for 4, 8, 16 and 32 seconds. The whole process was repeated for each subject. Fig. 6 is an example of the cumulative function obtained with this technique for subject MB.

For six of the subjects, practically every cumulative function was a straight line. An exception was the one-second sample which was occasionally curved. For the two other speakers, the 8 second sample was the shortest sample which appeared linear. This represents between four and five seconds of speech exceeding the −40 dbm threshold. It appears therefore that at least four or five seconds of "over the threshold" speech is required to achieve a log-uniform distribution.

In general, no conclusion could be drawn regarding a desirable sample length for a representative result because the data are inconsistent on this point. For example, for some speakers the one-second segment happened to be loud, while for others, it was quiet. Thus as the sample became longer, for some speakers the distribution settled downward,

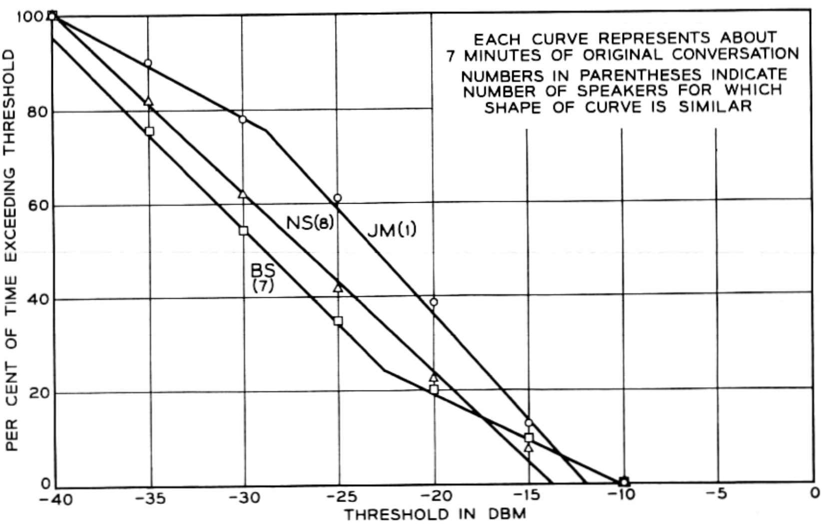

Fig. 5 — Representative speech envelope distributions for individual talkers.

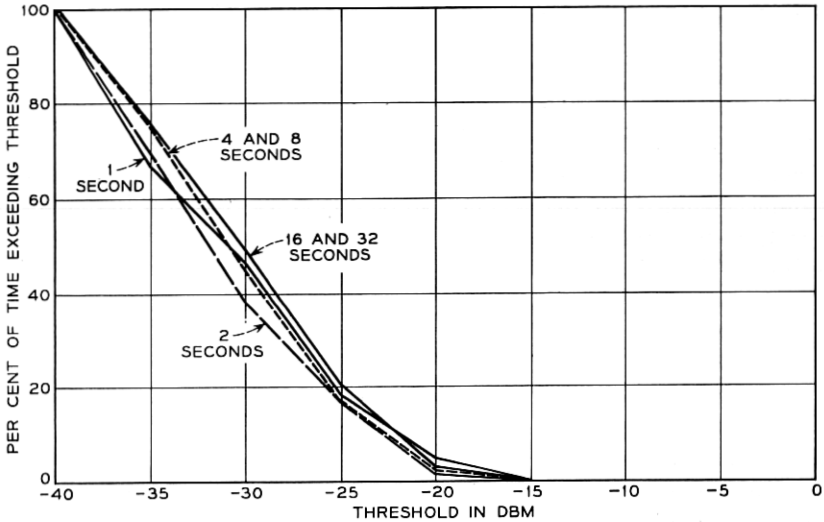

Fig. 6 — Envelope distributions for various sample lengths of continuous speech of subject MB. Data points are 5 db apart.

for others it went up, and for a few it fluctuated. Some distributions were "stabilized" at 8 seconds (that is, the 8 second function was the same as that for 16 and 32 seconds), others never stabilized.*

III. ESTABLISHING A MEASURE OF SPEECH LEVEL

3.1 *Properties of the Simple Log-Uniform Distribution*

Let X be a random variable, already defined by (1), which is uniformly distributed between a and b, where a and b are expressed in dbm. The peak value of X, at b, will be denoted X_{peak}. The density function is

$$p(x) = 1/(b - a) \tag{3}$$

and is shown in Fig. 7, along with the cumulative function. The lower limit, a, could be considered the threshold for a log-uniform speech distribution. This distribution, having a single constant value for $p(x)$, will be called the *simple log-uniform distribution* to distinguish it from the composite distribution, which will be defined in Section 3.3.

* The instability of the level distribution of a speaker is called *speech variation* and is further treated in Section 6.4.

The mean, or average value of X, is equal to

$$X_{\text{ave}} = (a + b)/2. \tag{4}$$

This quantity may be measured by obtaining the time average of X sampled and averaged only over those time intervals where X exceeds the threshold a.

The above-the-threshold rms voltage, denoted V_{rms}, is shown in Appendix B to be equal to

$$V_{\text{rms}} \text{ (in dbm)} = 6.38 + 10 \log_{10} (\Delta\text{mw}) - 10 \log_{10} (b - a) \tag{5}$$

where

$$\Delta\text{mw} = \log^{-1} \frac{b}{10} - \log^{-1} \frac{a}{10}. \tag{6}$$

That is, Δmw is the difference in milliwatts between the end points of the log-uniform distribution.

The average absolute voltage, again measured above the threshold, is denoted V_{ave} and is given by (see Appendix B)

$$V_{\text{ave}} \text{(in dbm)} = 10 \log \frac{(1000)(\bar{V})^2}{600} \tag{7}$$

where

$$\bar{V} = \left(\frac{8.686}{b - a}\right)(0.775)\left(\sqrt{\log^{-1}\frac{b}{10}} - \sqrt{\log^{-1}\frac{a}{10}}\right). \tag{8}$$

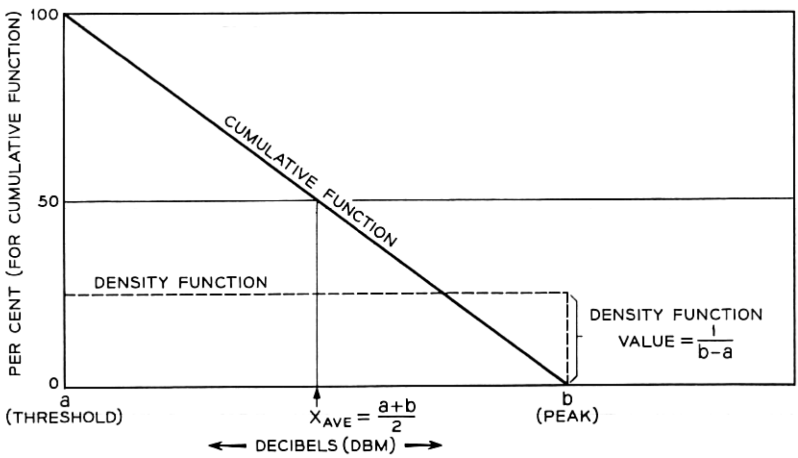

Fig. 7 — The log-uniform distribution.

(The quantity $(V)\ (b - a)/(8.686)$ is the difference, in volts, between the voltages to produce b and a dbm.)

Consider now what would happen to the density function of the random variable shown in Fig. 7 if the threshold a were raised (moved to the right) while the level of the pre-threshold random variable were held fixed (i.e., b remains fixed). The density function would increase in height, since $(b - a)$ would be smaller, but it would still be uniform over the range from threshold to peak. Although the peak b does not change, the quantities X_{ave}, V_{rms}, and V_{ave} do, as shown in Fig. 8. (The curves in the figure were calculated from (4), (5), and (7).)

It is clear from Fig. 8 that X_{peak} is the only quantity shown which is not dependent on the threshold setting. In fact, the other quantities vary so strongly with threshold that they would be completely meaningless were the threshold not specified.

Fig. 9 is also a plot of X_{peak}, X_{ave}, V_{rms}, and V_{ave} except that in this case the threshold is held fixed and the pre-threshold level is varied, in effect changing the value of b. The peak is seen to be the only quantity which varies on a db-for-db basis with level changes.

3.2 *The Average Peak Level*

If it were guaranteed that the levels in every speech sample were log-uniformly distributed, then our search for an ideal level measure

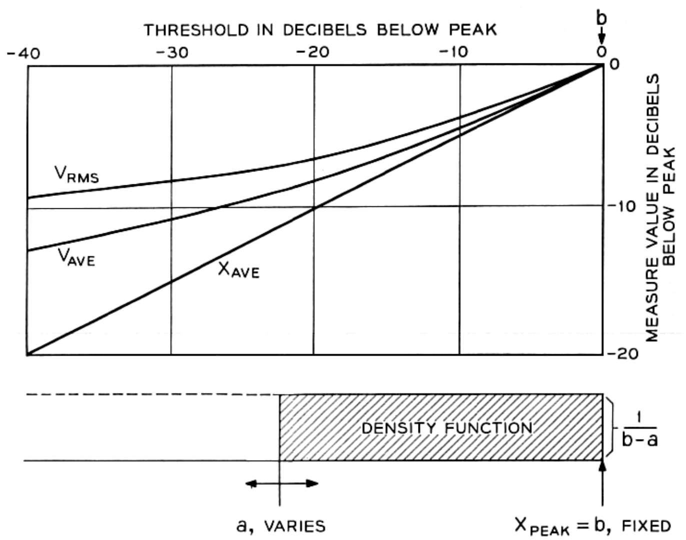

Fig. 8 — Measures of the log-uniform distribution as a function of varying threshold. Pre-threshold signal level remains unchanged.

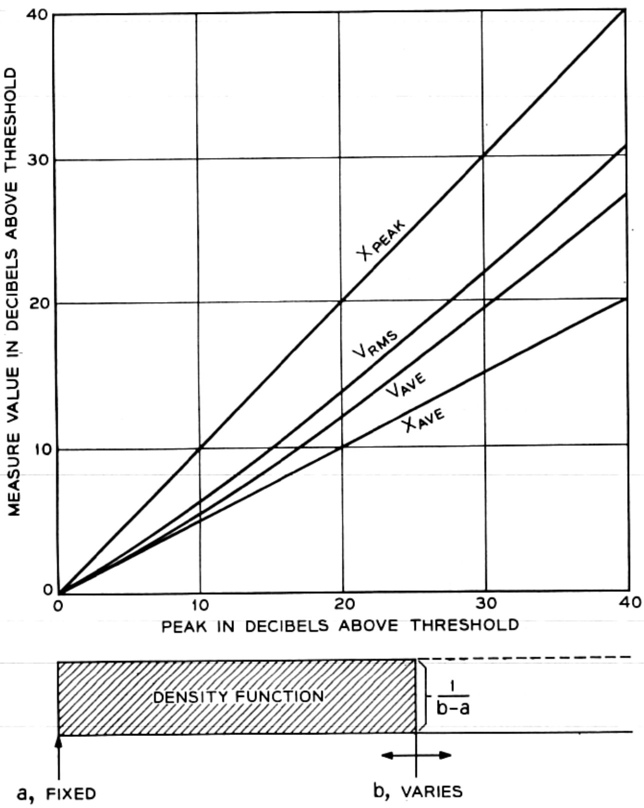

Fig. 9 — Measures of the log-uniform distribution as a function of varying pre-threshold signal level. Threshold is fixed.

would indeed be over. We would simply record the peak voltage (or peak envelope voltage) which occurs in a speech sample, and use this quantity to specify the level of the sample. It was already noted, however, that the peak is generally unsatisfactory as a level measure because it is too sensitive to isolated disturbances. Another approach to the problem is evident if we rewrite (4), solving for b:

$$b = a + 2(X_{\text{ave}} - a). \tag{9}$$

The peak can now be obtained by building a device to measure X_{ave} above some threshold, a, and then substituting X_{ave} in (9).* X_{ave} will of course be a function of a, but this is unimportant since the a dependence

* The Digital Speech Level Meter, described later in this paper, is actually constructed to measure the quantity ($X_{\text{ave}} - a$) and then substitute this difference into (9).

will cancel out upon solving for b. We shall denote the quantity obtained applying (9) as the *average peak level*, or apl:

$$\text{apl} \equiv a + 2(X_{\text{ave}} - a). \tag{10}$$

The apl is the peak of a hypothetical simple log-uniformly distributed variable which would have produced the same X_{ave} as was actually obtained. If the speech sample levels are in fact log-uniformly distributed, the apl will be equivalent to the sample peak and will possess all the properties of the peak. If the distribution is log-uniform except for some loud extraneous sound, the apl may deviate slightly from some of the stipulated requirements of an "ideal" measure, but it will be fairly immune to the extraneous sound since the measurement is taken over the entire speech sample.

The peak can also be obtained from V_{rms} or V_{ave}. Assume a device is built to measure V_{rms} above a known threshold a. Once V_{rms} is known, (5) might be solved for b, but this is a rather difficult task. A simpler method would be to read the peak from the V_{rms} curve of Fig. 9. The resulting measure would be the peak of a simple log-uniform distribution which would yield the same V_{rms} as was actually measured.

In this study, the peak is computed with X_{ave} rather than V_{rms} or V_{ave} because the apl has a simple, linear relationship to X_{ave} (10), whereas one must resort to graphs, tables, or involved computation with the other measures. The instrumentation required to apply (10) is straightforward, as will be shown in Section V.

3.3 *The Composite Log-Uniform Distribution*

Certain speech samples have cumulative distribution functions which are markedly different from a single straight line and are therefore not from a log-uniform density function. They can, however, be approximated by log-uniform functions in the following way. Consider a process in which a random variable X_1, log-uniformly distributed between a and b_1, is observed for five minutes, and is followed by X_2, log-uniformly distributed between a and b_2, for 10 minutes. (This discussion will be restricted to two variables, but any number of X_i could be considered.) To obtain the distribution for the entire 15 minute process, one adds the two separate density functions, each weighted by a suitable factor. The fraction of time X_1 is present will be called θ_1, and θ_2 will be the fraction of time for X_2.

The structure of the composite distribution is shown in Fig. 10. The density functions are in the upper drawing and the lower drawing shows the composite cumulative function. Given the cumulative function, one

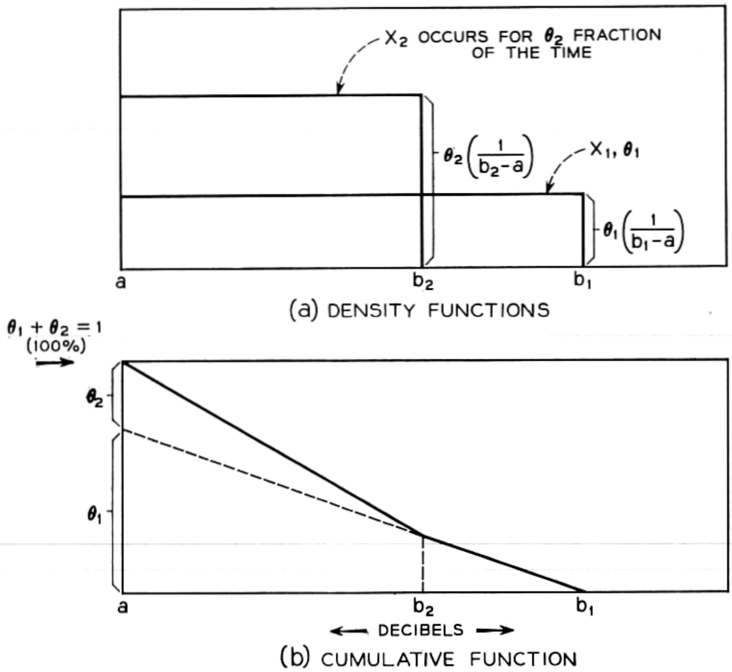

Fig. 10 — Distribution of a random variable X which is a composite of X_1 (b_1, θ_1) and X_2 (b_2, θ_2).

can immediately obtain b_1, b_2, θ_1, and θ_2. The manner in which this is done is shown on the drawing.

Routine calculation shows that the mean of the composite distribution is

$$X_{\text{ave}} = (X_1,X_2)_{\text{ave}} = \frac{a + \theta_1 b_1 + \theta_2 b_2}{2}. \tag{11}$$

This is the same mean which would result from a simple log-uniform distribution which has a peak at $\theta_1 b_1 + \theta_2 b_2$. The apl of such a distribution would therefore be a weighted average of the peaks of the log-uniform random variables which generate the composite distribution. That is,

$$\text{apl} = \theta_1 b_1 + \theta_2 b_2 . \tag{12}$$

Unfortunately, the apl is no longer threshold-invariant, since θ_1 and θ_2 are themselves dependent upon the threshold. This can be seen by letting the threshold a in Fig. 10 approach b_2. The variable X_2 will become less evident as it vanishes below the threshold, and θ_2 will approach zero while θ_1 approaches unity. The apl will therefore move towards b_1 from some point inbetween b_1 and b_2.

The amount of apl variation caused by changing the threshold will depend on the values of all the parameters involved. A theoretical analysis of this effect is included in Appendix C. It is shown that for most of the speech samples in this study, apl variations in the order of one db could occur if the threshold were allowed to become close to b_2. (If the threshold is too close to b_2, the variation is more severe.) Some experimental measures of this effect, included in Section 6.3, support the theoretical estimates.

3.4 *Suitability of Log-Uniform Approximation to Speech Levels*

Several of the speech samples analyzed here have cumulative functions which are quite linear. For a few others, the functions can be very well fitted by two lines, indicating a two variable composite distribution.

Now consider Fig. 4 which shows a distribution which slopes off gradually and for which the two line approximation introduces a noticeable error at the break point. This error can be reduced if a three line fit is made, and if ten lines are used, the error all but vanishes. The Fig. 4 curve can therefore be considered a composite of a large number of log-uniform distributions, all having a common threshold and having successively higher peaks.* In general, the composite distribution is valid if the cumulative function exhibits the following two properties:

(1.) For all points above the threshold, the curve cannot break downward (its second derivative cannot be negative).† This guarantees that the composite distribution contains no simple distribution which exists entirely above threshold.

(2.) If the curve breaks upward, the lowest break point (in dbm) must be above the threshold. Thus there are no density function peaks below threshold.

Every speech distribution noted by the author obeys both of the above rules if proper care is taken in determining the threshold. For the composite log-uniform distribution to be valid, the threshold may be set anywhere in the linear range of the curve between the downward and upward break points. This is generally a broad range of at least 15 db.

---

* The suitability of a Gaussian model is discussed in Appendix D.

† This rule is violated in Fig. 4 if the −40 dbm threshold is used. In making measurements from the curve, the threshold is raised until this rule is obeyed, as is done in Appendix A. For later reference, we might note here that the Digital Speech Level Meter uses a −30 dbm threshold, which is sufficiently high to clear all downward break points of the speech used in this study.

The threshold should, however, be set in the lower part of this range to minimize the threshold dependence of the apl.

IV. THE VU METER

4.1 *Technique of Using the VU Meter*

The VU meter is a widely accepted speech level measuring instrument. Its basic design consists of an amplifier, full-wave rectifier, and meter. The characteristics of the unit, especially the meter movement, are standardized and may be found in several references.[6] A standard procedure for reading the VU Meter (when monitoring a telephone conversation) has been adopted and is described by Carter and Emling:[7]

"The volume used by the party selected is the arithmetic average in VU of a series of individual volume measurements made on a selected party's speech throughout the conversation.

"An individual volume measurement provides a single figure based on a portion of speech several seconds in length (say 3 to 10 seconds). It is . . . the visual or inspection mean of the highest meter deflections, exclusive of the one or two very highest deflections, observed during the measuring period.

"[Typically], in a 5-second measuring interval, for example, there may be about 25 syllables, with a meter deflection or swing resulting from everyone of these. These swings can be divided roughly into two types: a large group of relatively small swings from the weaker syllables and a small group of high swings from the six or seven loudest syllables. It is on this second class of strong swings that the volume measurement is based; the highest one or two are excluded, however, since these may be somewhat special as to emphasis or accent and are not related closely to the five or six remaining strong swings."

One could regard the above process as a method of estimating the "average peak" of the meter response. Judging from the work reported in the previous sections of this paper, this measure is ideally a very good indication of speech level. In practice, however, it exhibits variability first because an observer's readings of the same speech sample are not repeatable, and secondly because different observers show different biases in reading the meter. Measurements of these variabilities were made by the author in a brief unpublished study. The standard deviation of a single observer was found to be as much as 1.5 VU (db) and a range of observer bias of almost 3 VU occurred among observers.

Shearme and Richards[4] report similar findings. They find that a "trained observer will yield 5 per cent of readings as much or greater than 2 db away from the mean value." This corresponds to a standard deviation of 1 VU, obtained with our most experienced observer. Shearme and Richards also report that "even with trained observers a total range [of observer bias] of 4 db is encountered."

4.2 *Relationship of this Study to VU Measurements*

It is apparent that the VU meter yields imprecise readings when used in an attempt to make objective speech level readings. A major source of the variability is due to the human observer, and one way of removing this variation is to instrument the reading process. This could be done by constructing a device which would follow a set of rules similar to those stated by Carter and Emling.[7] But these rules were tailored for an observer and perhaps there could be a better measure of speech level when the human limitation is removed.

Indeed, the apl has been shown to have a direct relationship to the underlying speech level distributions. Principally for this reason, the author chose to construct a device which measures the apl and not the VU level of a speech sample. It will be shown that this device, called the Digital Speech Level Meter, yields readings of less variation than VU readings and is therefore potentially a more precise instrument for measuring levels.

This does not imply, however, that the Digital Speech Level Meter is a total replacement for the VU meter. The VU meter is generally adequate for setting a "good" recording level, and its readings are often considered to be an indication of subjective loudness. This is usually argued on the basis of the design of the needle movement. Further reasoning is based on the ground that the meter is read by an observer who himself has some ideas about the loudness of the signal. Thus the VU and apl readings reflect somewhat different properties of speech. They may be compared with reference to objective level measurements, but in other respects each measure must be judged on its own merits.

V. THE DIGITAL SPEECH LEVEL METER

5.1 *Obtaining the Log-Average Voltage*

It is shown in Section 3.2 that the apl is easily obtained once the average of the log-uniform distribution (X_{ave}) is known. Fig. 11 illustrates the technique used to obtain X_{ave}. Speech is full-wave rectified, filtered, and applied to a log voltage to frequency converter. That is, the output

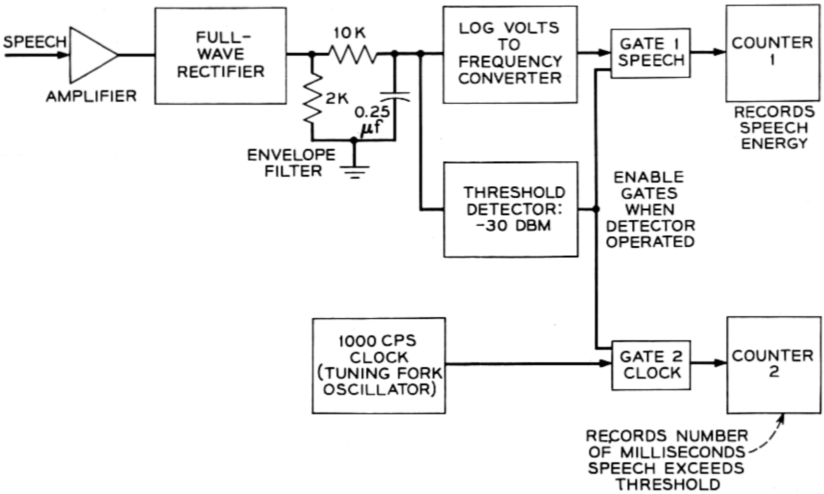

Fig. 11 — Basic design of digital speech level meter.

frequency exhibits uniform incremental changes with uniform changes in the decibel level of the input. The linearity extends over about a 30-db range.

The filtered signal is also applied to a speech detector consisting of solid state circuitry and having insignificant pickup and hangover times. The threshold of the detector is −30 dbm re OTL. This value, rather than −40 dbm, was chosen to keep the speech in the linear range of the detector. This threshold setting is still low enough to keep the apl nearly independent of the threshold for those speech samples which do not have a simple log-uniform distribution.*

The speech detector operates two gates. The "speech gate" sends pulses from the voltage controlled oscillator to counter 1, whose reading may be interpreted as the accumulated energy of the log voltage. The "clock gate" sends pulses from a 1000-cps clock to a second counter, whose reading specifies the amount of time the speech level has exceeded the threshold.

To obtain a level reading for a speech sample, the counters are first

* The statement that "the apl is nearly independent of threshold" does not imply that one may randomly vary the Speech Level Meter threshold without affecting the meter reading. Since the meter measures X_{ave}, which itself *does* depend on threshold (4), the threshold must be taken into account in solving for the apl (10). Thus consider two meters with thresholds of −35 and −30 dbm, respectively. If each is properly calibrated with respect to its own threshold, then each should read approximately the same apl for the same speech sample.

reset to zero and then the speech sample is played. When finished, the counter 1 reading is divided by the counter 2 reading to obtain an average frequency. This of course may be directly converted to average log voltage since the frequency is a linear function of the log voltage.

The instrumentation up to this point is very similar to the method used by P. D. Bricker to obtain a measure of speech level.[8] Bricker's circuit is almost identical to that of Fig. 11 except that his speech detector has a 200-msec hangover time and his first counter is driven by a *linear* voltage to frequency converter. Thus his average frequency obtained from the two counter readings is an approximate measure of V_{ave} rather than X_{ave}. Bricker's success in estimating VU readings with his technique provided considerable encouragement for the present study.

5.2 *Obtaining a Direct Reading of the APL*

The above procedure requires that the observer write down two numbers, divide them, and apply a conversion to yield the apl value. One way of instrumenting these operations is as follows. In Fig. 11, a flip-flop is installed on counter 2 in such a way as to shut down the whole device upon observing 1000 msec of speech. This automatically accomplishes the necessary division. Counter 1 is constructed to count toward zero starting from some negative number whose value depends on the calibration of the voltage to frequency converter.

The voltage to frequency converter is adjusted so that for a 1 db increment in the level of a sine wave input signal (thereby increasing the mean of the logarithm of its envelope by 1 db), the frequency converter changes by 2.0 cps. (This is actually 20 cps, but a decimal point is inserted in the read out.) Recall now that if *speech* has its over-all level increased by 1 db, the mean of its logarithm is increased by only 0.5 db. The converter, having a 2 to 1 "frequency to db" conversion, will exhibit a frequency change of 1.0 cps, correctly reflecting the change in speech level.

Because of the nature of the speech level meter calibration, it will not work properly in its present form if used to measure the levels of a tone or other signals which do not have a log-uniform distribution.*

The meter can be set to read speech over intervals of time other than

* Consider a random variable Y having a probability distribution such that the peak is linearly related to the above-the-threshold average (denoted $\bar{y}$) by $\bar{y} = a + [(b - a)/k]$ where a is the threshold, b is the peak, and k is a constant independent of a. The (log-) uniform distribution, in which $k = 2$ (see (4)), is one of a large class of such distributions. Although the technique described in this paper for measuring speech levels might be suitable for measuring other random variables, the speech level meter is calibrated for $k = 2$ and requires a uniform distribution.

one second without subsequent division. This is accomplished by inserting flip-flops just ahead of the counters. For example, if one flip-flop is placed in front of each counter, the counting rate will be halved and the meter will read directly for 2 seconds of speech.

Fig. 12 is a photograph of the speech level meter. To obtain a reading, the observer presses the reset key which turns off the display and starts the internal counters integrating the speech energy. When the lower counter reaches 1000 (this display is usually not illuminated), the upper display is turned on, the observer records the number, and again pushes the reset key if another reading is desired.

It is possible to modify the meter so it does not stop after a fixed time interval but continues counting in the manner described in the previous section. This and several other options are provided by various front panel controls.

VI. RESULTS OBTAINED WITH THE SPEECH LEVEL METER

6.1 *Scope of the Results*

The data presented here represent measurements made on 16 samples of telephone speech, each about 7 minutes long. All of the samples were

Fig. 12 — Digital speech level meter.

recorded on the same circuit. Any conclusions which are based on these data must be regarded as limited in scope and can be broadened only through further data acquisition. The data included here should, however, suggest the general limits of performance which can be expected.

6.2 *Measuring Technique*

All level measurements reported here were made by taking the average of a succession of 4 second readings. If the meter is reset immediately after each reading, this technique yields a result which is equivalent to that obtained by allowing both counters (Fig. 11) to run continually and forming a ratio at the end of the sample. The present method was adopted because it is easy for the observer to use, and reads directly, without conversion.

6.3 *Response of the APL to Changes in Level*

The requirement that the apl be invariant with threshold is equivalent to the requirement that it vary on a db-for-db basis with attenuation or amplification of the voice signal. This is true because a signal attenuation of, say, 5 db will yield the same shape probability density function as will raising the threshold by 5 db, although the resulting distributions will differ in absolute levels. The apl's of the two new distributions should ideally differ by 5 db.

The following experiment illustrates the effect of level changes on the apl. Four 7-minute samples of speech were each played through the speech level meter at three different levels, each 5 db apart. The readings were as follows:

Table I
Effect of Over-All Level Changes on APL Readings

Level	Speaker			
	AD	JS	MH	CB
−5 db	−20.01	−20.57	−15.26	−17.57
	Δ = 5.28	Δ = 4.68	Δ = 5.48	Δ = 4.85
Normal	−14.73	−15.89	−9.78	−12.72
	Δ = 5.63	Δ = 6.06	Δ = 4.06	Δ = 4.61
+5 db	−9.10	−9.83	−5.72	−8.11

With one exception, the apl readings for each speaker reflect the speech level variations with an error of less than 1 db over the 5 db increments, and all of the speakers are within 1 db for the 10 db increments. This is in general agreement with the theoretical results of Appendix C, namely, that the apl is threshold invariant to within about 1 db for most speech samples.

6.4 *Repeatability of Meter Readings*

The speech samples of three talkers were each played ten times to determine the variation which might be expected in obtaining repeated measurements. The estimations of the standard deviations of the levels for the samples were 0.080, 0.154, and 0.043 db. The meter readings are therefore highly repeatable.

A sample of the readings taken during one of the runs is shown in Table II. Only 5 of the original 10 columns are shown. These data are included to illustrate two very different sources of variation which occur in taking readings.

The first source is *speech variation*, which exists because the speaker varies his level as he talks. This variation is reflected in the range which exists in the numbers in a single vertical column. For example, one concludes from the data in the first column that the apl for the entire speech sample is -10.99 dbm, with an estimated standard deviation for any randomly chosen 4-second sample of 3.57 db.

Speech variation does not enter into the repeatability of measuring the level of a *particular* speech sample. In this case, the variation of this measure would be determined in part by the variability of reading the same 4-second sample and by the number of samples taken. A rough idea of the repeatability of a 4-second sample reading is found by reading across the top horizontal row of Table II. (Other rows are not suitable for comparison because of timing errors in resetting the meter. That is, the fifth reading may not be taken for exactly the same speech sample every time the tape is played.) A rough guess at the standard deviation of a particular sample is 0.3 db, based on cursory inspection of data taken on short speech samples (not included here).

If this value of 0.3 db is divided by $\sqrt{N}$, where N is the number of 4-second readings, one might expect to obtain the standard deviation of the average of the entire speech sample. For the data of subject SK, as shown in part in Table II, N equals 30, which would lead us to expect a σ of 0.055 db. The measured value of σ was 0.154 db. The data from the other two speakers having deviations of 0.080 and 0.043 db are more in line with the expected value of 0.06 db. For each of these speakers, $N \approx 25$.

TABLE II

Repeated measurements of a seven minute sample of speech of Subject SK. (Only 5 of the original 10 columns are shown. All readings are negative numbers.)

1	2	3	4	5
19.3	19.1	19.2	19.0	19.2
11.5	11.8	12.0	11.4	11.1
18.8	18.7	18.9	18.7	18.0
12.4	12.5	12.4	12.5	12.5
15.4	15.3	15.5	15.1	15.5
9.0	9.3	8.9	9.5	8.9
9.5	10.8	7.4	11.4	9.1
5.2	4.4	5.1	1.5	5.2
4.7	4.9	4.5	7.7	4.6
13.4	11.4	13.6	5.8	10.0
12.4	12.0	12.2	13.6	12.1
15.0	14.6	15.4	11.2	14.1
10.7	9.9	10.7	12.4	9.5
13.5	15.8	13.7	14.8	15.7
9.8	8.2	8.2	9.7	8.5
9.3	9.7	9.7	10.6	9.4
8.2	8.7	8.7	7.9	7.8
12.6	9.2	10.3	8.8	12.5
10.7	12.0	12.0	13.0	10.7
13.7	10.8	10.6	11.8	13.2
11.9	14.0	13.8	13.7	11.6
10.3	10.4	10.1	12.2	10.5
8.1	8.2	8.0	6.2	8.9
10.8	10.8	10.8	11.5	9.0
10.4	10.2	10.6	10.4	11.0
8.9	8.7	9.0	10.0	6.8
12.0	12.1	9.5	6.7	11.5
9.4	9.7	10.0	11.1	8.6
2.6	2.6	5.4	7.6	2.4
10.1	9.0	6.4	0.7	10.5
		Column Averages		
−10.99	−10.83	−10.75	−10.55	−10.61

The variation in any one column is predominantly due to speech variation. For example, $\sigma = 3.57$ db for column 1. Differences in the column averages are due to measurement variation. For all 10 columns, $\sigma = 0.154$ db.

6.5 *A Comparison of Different Types of Measurements*

The apl readings made by the speech level meter were compared with the apl estimates based on the graphical technique described in Appendix A. The results are shown in Table III. Also included in this table are the VU readings for the samples taken by Miss Kathryn L. McAdoo, an experienced VU reader.[9]

The meter and graphical apl levels are plotted against each other in Fig. 13. The linear least mean squares fit passes through the two averages

Table III

APL Readings for All Speech Samples*

Subject	Meter APL	Graphical APL	VU
RT	−19.56 dbm	−19.76 dbm	−24.68 vu
AD	−15.49	−17.5	−23.00
MB	−17.49	−18.25	−24.24
JS	−16.56	−19.0	−23.00
PF	−14.20	−15.0	−19.67
JM	−8.79	−11.75	−16.45
ES	−21.11	−21.75	−24.66
PR	−18.24	−19.55	−23.67
MH	−10.00	−12.3	−17.73
SR	−9.22	−11.07	−14.41
SK	−10.77	−12.85	−19.14
CB	−13.28	−14.0	−18.60
SS	−16.97	−20.64	−22.77
BS	−13.96	−15.56	−20.33
NS	−13.35	−13.7	−20.08
VB	−18.05	−21.76	−27.12
Averages	−14.82	−16.53	−21.22

Rank Order Correlations: Meter APL vs Graphical APL = 0.944
Meter APL vs VU = 0.949

* These readings should not be directly compared with those in Table I because of a difference in calibration used for the two sets of readings.

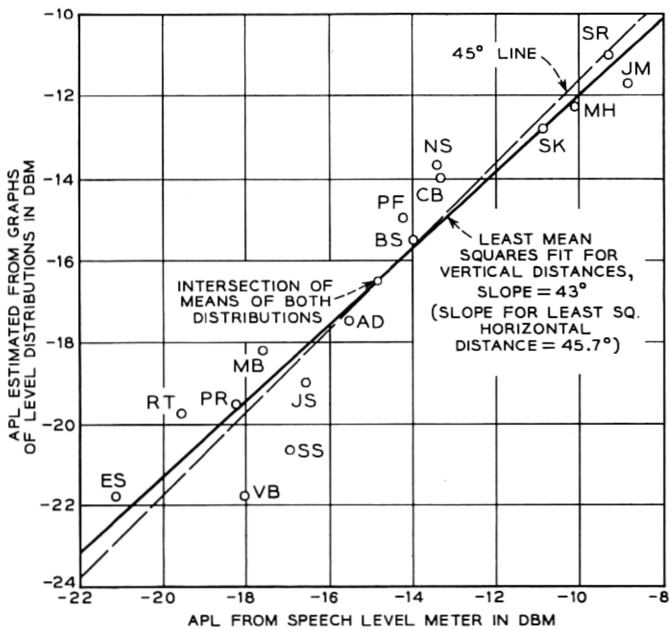

Fig. 13 — A comparison of graphical and meter APL readings.

of both distributions and has a calculated slope of 43°. Because the slope is so close to 45°, we conclude that on the average, the methods are consistent with each other in comparing relative levels among speakers.

The means of the distributions do not coincide, showing an over-all bias such that the meter reads about 1.7 db higher than the graphs, with a variation of about 1 db. This can be attributed to several factors, such as differences in the instrumentation used in the meter and in the equipment which generated the graphs, and the inadequacy of the two-line approximation used in the graphical analysis. Another factor is the threshold dependency of the apl; the meter had a threshold of −30 dbm while the threshold in the graphical analysis was closer to −40 dbm.

The VU readings and meter apl levels for the 16 speech samples are plotted against each other in Fig. 14. The slope of the linear least mean squares fit is 41.3°, showing that the apl and VU readings tend to differ by within 2 db of a constant over the range of the speech samples. (A 15-db change in meter level readings produces a 13 db change in the least mean squares fit.)

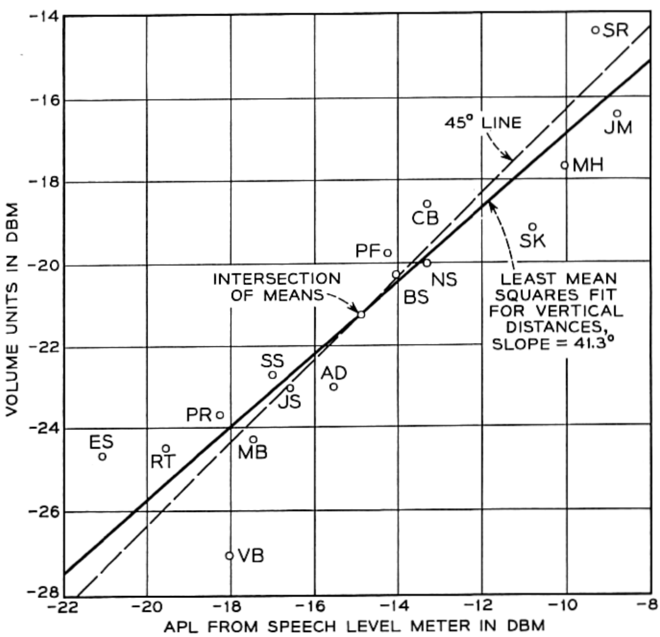

Fig. 14 — A comparison of VU and meter readings.

VII. CONCLUSION

In this study we have shown that the apl is a one-dimensional objective measure of speech level and that it satisfies, within certain bounds, a stipulated set of requirements of an "ideal" measure. The Digital Speech Level Meter is presently undergoing further tests which will help to determine more precisely the properties of the apl.

One unanswered question is that of determining a relationship between meter readings and subjective impressions of loudness. Other areas of further study include measuring levels of clipped or volume limited speech, high-fidelity speech (as opposed to telephone speech), and possibly other types of signals such as noise. Note that the demonstrated correspondence between level distributions of telephone and high-fidelity speech (Figs. 1 and 2) implies that the meter would work equally well with either type of speech.

The level measuring technique described here has many possible applications if further experimentation indicates the method to be suitable. The limited data already available show that the technique is promising.

VIII. ACKNOWLEDGMENT

I am most grateful to Miss Donna Mitchell for the many long and tedious hours she spent gathering and analyzing almost all of the data obtained for this study. I am also indebted to F. S. Fillingham and S. E. Michaels for their assistance in the design and construction of the experimental meter.

APPENDIX A

Procedure Used to Obtain Graphical APL Measurements

The cumulative function for the syllabic waveform of the speech of SR is shown in Fig. 15. This curve is chosen for demonstration because it has two break points, and obtaining the apl reading for it involves more steps than for most other graphs.

Notice that the break point near −35 dbm is below the −30 dbm threshold used in the speech meter and therefore does not affect the meter reading. This is the case for all speakers having such break points. The break point is therefore ignored and the curve is extended to 100 per cent as a linear extrapolation. In computing θ_1 and θ_2, the vertical

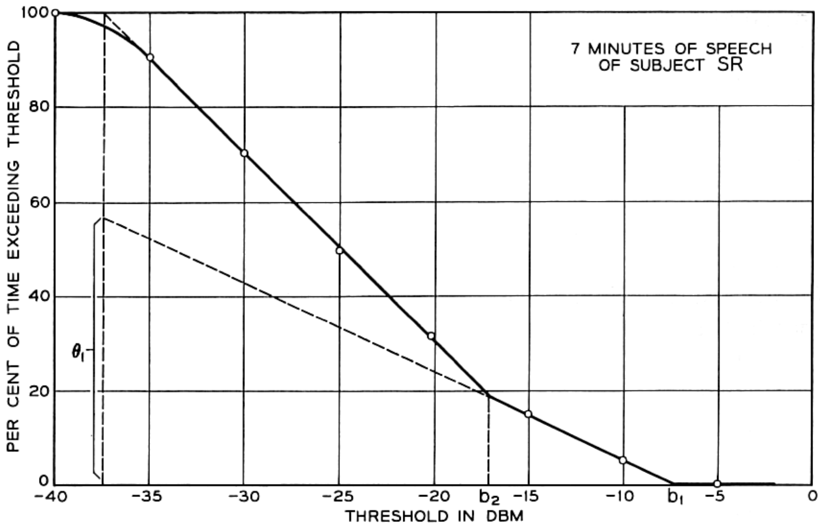

Fig. 15 — Estimating APL from a distribution having a low-level break point.

line from which θ_1 is chosen must be the line at which the cumulative function reaches 100 per cent. This no longer occurs at -40 dbm but rather at -38 dbm.

Having found θ_1, θ_2, b_1, and b_2, the apl is computed from (12). Table IV is a tabulation of these quantities for all of the 16 speakers.

TABLE IV

GRAPHICAL CALCULATION OF THE APL FOR 16 SPEAKERS

Speaker	b_1 (dbm)	θ_1	b_2	θ_2	Graphical APL
RT	−15.5	.31	−22	.69	−19.76
AD	−17.5	1	—	—	−17.5
MB	−18.25	1	—	—	−18.25
JS	−19.0	1	—	—	−19.0
PF	−15.0	1	—	—	−15.0
JM	−11.75	1	—	—	−11.75
ES	−21.75	1	—	—	−21.75
PR	−15.0	.30	−21.5	.70	−19.55
MH	−9.0	.50	−15.6	.50	−12.3
SR	−7.5	.58	−16.0	.42	−11.07
SK	−9.3	.52	−16.7	.48	−12.85
CB	−14.0	1	—	—	−14.0
SS	−13.5	.27	−23.3	.73	−20.64
BS	−10.2	.56	−22.4	.44	−15.56
NS	−13.7	1	—	—	−13.7
VB	−17.2	.57	−27.8	.43	−21.76

APPENDIX B

Derivation of V_{rms} and V_{ave} for the Log-Uniform Distribution

B.1 V_{rms}

Let X be a random variable uniformly distributed between a and b (Fig. 7) such that

$$x_{(\text{dbm})} = 10 \log \frac{1000\,(v^2)}{600}\,. \tag{13}$$

Let Y be a random variable which represents the power in milliwatts dissipated in a 600 ohm resistor. Then

$$x = 10 \log y. \tag{14}$$

From Fig. 7,

$$\text{Prob } X \leqq x = \int_a^x \frac{1}{b-a}\,dx = \frac{x-a}{b-a}\,. \tag{15}$$

Substituting (14) into (15),

$$\text{Prob } Y \leqq y = \frac{(10 \log y) - a}{b-a}\,. \tag{16}$$

Recall that for any variable Z,

$$\log_{10} Z = (0.4343) \ln_e Z. \tag{17}$$

This is used in differentiating (16),

$$\begin{aligned} p(y) &= \frac{4.343}{y(b-a)} \text{ for } y \text{ between } \log^{-1}\frac{a}{10} \text{ and } \log^{-1}\frac{b}{10} \\ &= 0 \text{ elsewhere.} \end{aligned} \tag{18}$$

Equation (18) tells us that the density function of the power in milliwatts is of the form of a hyperbola, not an exponential as might be guessed from the uniform distribution of log y.

To obtain V_{rms}, obtain the average power, that is, the expectation of y

$$E(y) = \int_{\log^{-1}(a/10)}^{\log^{-1}(b/10)} y\,p(y)\,dy = \int_{\log^{-1}(a/10)}^{\log^{-1}(b/10)} \frac{4.343}{b-a}\,dy. \tag{19}$$

Define

$$\Delta \text{ mw} = \log^{-1}\frac{b}{10} - \log^{-1}\frac{a}{10}\,. \tag{20}$$

Then, integrating (19),

$$E(y) = \frac{4.343}{b - a} (\Delta \text{ mw}) . \tag{21}$$

The rms voltage is the voltage required to produce this average power. Expressed in dbm,

$$V_{\text{rms}} = 10 \log E(y) = 6.38 + 10 \log (\Delta \text{ mw}) - 10 \log (b - a). \tag{22}$$

B.2 V_{ave}

Let V be a random variable representing the *absolute* voltage which would generate X. This voltage is monotonically related to the power by

$$y = \frac{(1000)\,(v)^2}{600}. \tag{23}$$

Taking logarithms,

$$10 \log y = 10 \log \frac{10}{6} + 20 \log v. \tag{24}$$

Substitute into (16),

$$\text{Prob } (V \leqq v) = \frac{\left(10 \log \frac{10}{6} - a + 20 \log v\right)}{b - a}. \tag{25}$$

Differentiating,

$$\begin{aligned} p(v) &= \frac{8.686}{(b - a)v} \text{ for } v \text{ between } 0.775 \sqrt{\log^{-1} \frac{a}{10}} \\ &\qquad\qquad \text{and } 0.775 \sqrt{\log^{-1} \frac{b}{10}} \\ &= 0 \text{ elsewhere.} \end{aligned} \tag{26}$$

Define

$$\Delta v = (0.775) \left(\sqrt{\log^{-1} \frac{b}{10}} - \sqrt{\log^{-1} \frac{a}{10}} \right). \tag{27}$$

Then in an identical manner of obtaining (19),

$$\bar{V} = E(V) = \left(\frac{8.686}{b - a}\right) \Delta v. \tag{28}$$

Expressed in dbm,

$$V_{\text{ave}} \text{ (in dbm)} = 10 \log \frac{(1000)\,(\bar{V})^2}{600}. \tag{29}$$

APPENDIX C

Variation of the APL With Threshold

Fig. 10 shows a composite log-uniform distribution of two variables, X_1 and X_2, each occurring above threshold for θ_1 and θ_2 fractions of the total time, respectively. The apl equals $\theta_1 b_1 + \theta_2 b_2$, and since θ_1 and θ_2 vary with threshold, the threshold will also influence the apl.

Let the threshold be increased to some new a', which is somewhere between a and b_2. We define

$$\varphi_1 = \left(\frac{b_1 - a'}{b_1 - a}\right)\theta_1 \tag{30}$$

$$\varphi_2 = \left(\frac{b_2 - a'}{b_2 - a}\right)\theta_2 . \tag{31}$$

The variables φ_1 and φ_2 represent the respective proportions of X_1 and X_2 which remain above threshold, each weighted by the original value of θ_i. Since $\varphi_1 + \varphi_2 \neq 1$, new values for θ_i are obtained by letting

$$\theta_1' = \frac{\varphi_1}{\varphi_1 + \varphi_2}, \qquad \theta_2' = \frac{\varphi_2}{\varphi_1 + \varphi_2}. \tag{32}$$

Knowing θ_1', θ_2', b_1, and b_2, it is possible to calculate a new apl′ and subtract from it the original apl to determine the variation produced by the threshold change. The general relationship between apl variation and threshold change is rather involved, and will be omitted here. From Fig. 10, however, one can see that if a is moved a short distance to the right, the effect upon the apl will vary, depending upon whether the move was made very near to b_2 or some distance from it. Assume, for example, θ_2 is very large (say 0.95), causing b_2 to dominate the apl for low thresholds. A 2 db change in a, if a is low, may hardly affect the apl, but if a is very near b_2, the 2 db change could eliminate the X_2 variable and cause the apl to shift rapidly toward b_1.

Calculations were made to determine what the graphical apl's in Table IV (Appendix A) would have been had a threshold of −25 dbm been used instead of −40 dbm. This is a severe test, as −25 dbm is a

TABLE V
VARIATIONS IN THE APL WITH RESPECT TO THRESHOLD

Speaker	−40 dbm apl	−25 dbm apl	Differences, db
RT	−19.76	−18.59	1.17
PR	−19.55	−18.42	1.13
MH	−12.30	−11.83	0.47
SR	−11.07	−10.36	0.71
SK	−12.85	−12.20	0.65
SS	−20.64	−17.31	3.33*
BS	−15.56	−12.48	3.08*
8 other speakers	No difference, since $\theta_2 = 0$		

* For a −30 dbm threshold (instead of −25 dbm), these differences are: SS, 0.93 db; BS, 1.24 db.

somewhat unreasonable threshold for these particular speech samples. (In fact, since the new threshold clears b_2 for subject VB, this sample is not considered in this comparison). The apl comparisons are as shown in Table V.

Based upon the results in the table, the statement is made that for most speech samples in this study, the apl is, within about 1 db, invariant with threshold.

It might be possible to reduce the apl threshold dependence by subtracting a small correction factor from fairly low readings, in which the peaks are close to the threshold. The value of the correction would taper off for higher apl's. Further study may determine whether such a procedure is advisable or feasible.

APPENDIX D

The Gaussian Distribution as a Speech Model

The Gaussian distribution, in addition to the log-uniform distribution, may serve as a model for the speech data. The cumulative speech function of Fig. 4 is replotted in Fig. 16, along with the (log-) Gaussian cumulative function with mean (μ) of −27.9 dbm and standard deviation (σ) of 8.3 db. The mean was set equal to the speech median, while σ was derived by setting $\sigma/2$ equal to the +19.2 per cent point above the median (−32.0 dbm).

The two curves are very similar* and might be even more so if the

* The fact that the cumulative function for all speakers is nearly Gaussian is *not* a consequence of the central limit theorem. The theorem states that the sum (or average) of n independently distributed variables will have a nearly Gaussian dis-

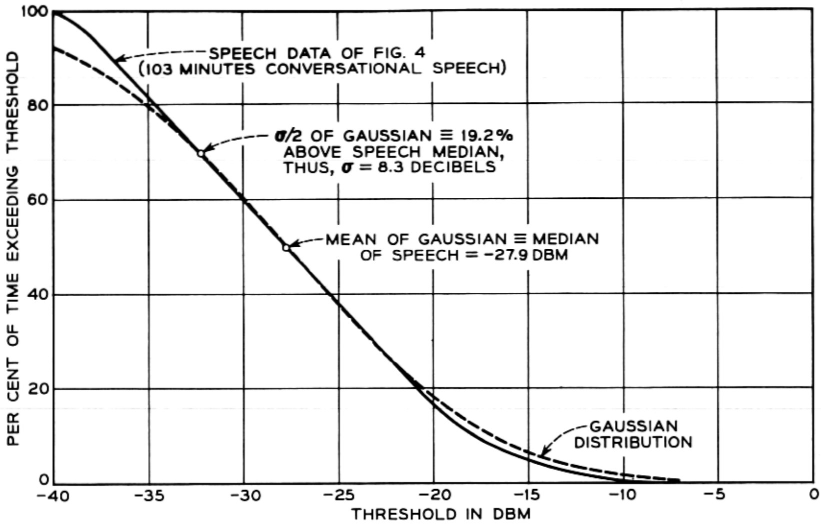

Fig. 16 — Speech distribution compared with Gaussian distribution.

speech function had not been normalized to 100 per cent at −40 dbm. (If the unnormalized speech curve were used, the Gaussian parameters would need readjustment.) The Gaussian model is, of course, most familiar and is of great help in analysis. For specifying speech levels, however, it is inferior to the log-uniform model for the following reasons:

(1.) It is not unidimensional; both μ and σ are required to specify one distribution curve.

(2.) There seems to be no clear-cut method of obtaining either μ or σ. In Fig. 16, μ was equated to the speech median, but the median is dependent upon the threshold. If the threshold were removed, the circuit would be under constant observation and the silent intervals would introduce data of uncertain significance. In Fig. 16, σ was obtained from a quantile point (19.2 per cent above median), and this also varies with threshold.

(3.) For some speech distributions, the simple log-uniform model is a better fit than the Gaussian model (see the NS curve of Fig. 5). And even when the Gaussian fit is better, as in Fig. 16, the composite log-

tribution when n becomes very large. The speech distribution in Fig. 16 is of *one* variable: the waveform of the envelope of a 103 minute speech sample. If each speaker had a simple log-uniform distribution with an apl of −10 dbm, then the 103 minute sample would have precisely that distribution. It may be that the overall level distribution for many speakers is approximately Gaussian, but this is a result of the nature of the speakers and not of a limiting theorem.

uniform model is still valid, if the threshold falls in the linear range of the cumulative function.

We are actually in the favorable position of not caring whether the distribution is Gaussian or log-uniform, as our only concern is that there exists a (quasi-) linear part of the cumulative function, and either of the above models provides for this. For this reason, the composite log-uniform distribution, which embraces both of these models, is used as a basis for specifying a unidimensional speech level.

REFERENCES

1. Sivian, L. J., Speech Power and Its Measurement, B.S.T.J., *8,* Oct., 1929, p. 646.
2. Dunn, H. K., and White, S. D., Statistical Measurements on Conversational Speech, JASA, *11,* Jan., 1940, p. 278.
3. Davenport, W. B., An Experimental Study of Speech-Wave Probability Distributions, JASA, *24,* July, 1952, p. 390.
4. Shearme, J. N., and Richards, D. L., The Measurement of Speech Level, Post Office Electrical Eng. J., *47,* 1954, p. 159.
5. Brady, P. T., A Technique for Investigating On-Off Patterns of Speech, B.S.T.J., *44,* Jan., 1965, p. 1.
6. Beranek, L. L., *Acoustic Measurements*, John Wiley, New York, 1949, p. 504.
7. Carter, C. W., and Emling, J. W., Unpublished report on making volume measurements on telephone message circuits, Bell Telephone Laboratories, 1950.
8. Bricker, P. D., A Technique for Objective Measurement of Speech Level, J. Acoust. Soc. Amer., Aug., 1965, letter.
9. McAdoo, K. L., Speech Volumes on Bell System Message Circuits 1960 Survey, B.S.T.J., *42,* Sept., 1963, p. 1999.