THE BELL SYSTEM TECHNICAL JOURNAL
Vol. 51, No. 4, April, 1972
Printed in U.S.A.

Objective Measures of Peak Clipping and Threshold Crossings in Continuous Speech

By PAUL T. BRADY

(Manuscript received September 1, 1971)

This study reports data on the statistics of instantaneous speech levels in continuous speech samples, with special emphasis on threshold crossings and other quantities related to peak clipping. All clipping thresholds for each speech sample are defined with respect to the individual speech level for that sample, specified in equivalent peak level (epl). Speech clipping is also treated as speech-correlated noise by assuming that it is caused not by a voltage limiting process, as actually occurs, but by an additive "phantom signal" that will cause the original signal to appear to be clipped. Empirical measures are obtained for the percent time a clipping level is exceeded, for the relation between phantom signal (i.e., noise) power and clipping level, and for the loss in signal power resulting from clipping. The relation is also established between epl and average (rms) power to allow signal power and signal-to-clipping noise levels to be specified.

I. INTRODUCTION

This study reports data on the statistics of instantaneous speech levels in continuous speech samples, with special emphasis on threshold crossings and other quantities related to peak clipping. Since a standard measure of specifying the amount of peak clipping is lacking, the study begins by arbitrarily defining the clipping level in terms of an objective speech level measure, the equivalent peak level (epl).[1] Then, measures of time spent above various thresholds (or clipping levels) are reported. Finally, a method is suggested for interpreting clipping as speech-correlated noise. Empirical measures are given for the amount of noise introduced as a function of the amount of clipping, and signal power lost because of clipping.

This paper does not address itself to subjective measurements related to clipping of the speech signal. The subjective effects of clipping must eventually be considered in a treatment of the broad problem of determining clipping performance objectives. It is also necessary to examine the objective effects of clipping on the transmitted signal so that guidelines can be established for design and operation of transmission systems. This paper is directed toward providing data to assist in engineering considerations of speech circuits.

II. DEFINITION OF A MEASURE OF PEAK CLIPPING

Peak clipping, produced by an abrupt limiting of a waveform when its amplitude attempts to exceed a clipping voltage, is possible in virtually any speech transmission system if the speech level becomes too high. It is quite easy to specify the limiting voltage in an absolute sense (e.g., ±0.5 V), but the speech impairments caused by clipping are a function of the relation between the clipping voltage and the speech level.

One clipping measure sometimes used is the difference between the clipping level and the volume unit (VU). This difference is quite variable because of the variability in the VU measure, as determined in experiments by Shearme and Richards,[2] and Brady.[3] A study related to VU and speech peaks was done by Noll,[4] who measured the difference between the highest instantaneous peak of a speech sample and the VU for possible application to peak clipping. This difference was also quite variable among samples, causing Noll to conclude that "volume (i.e., VU) distributions cannot easily be converted to peak distributions." He states further that the variability "might have been caused by inaccuracies inherent in reading a VU meter."

Another clipping measure sometimes used is the difference between the highest instantaneous peak in the speech sample and the clipping voltage. Although this measure is objective and fairly easily obtained, it is too dependent on the near chance value of the highest peak. Should this peak be due to an unusually loud segment of speech or even a click or spurious signal, the clipping measure has little relationship to the major part of the speech sample.

The present study begins by defining a new measure of clipping as the difference between the epl of a speech sample and the clipping level. This difference is measured in dB. It promises to be more stable than VU-related clipping measures since the epl is an objective measure with considerably less variability than the VU.[1,3] It is emphasized that this new clipping measure is not proposed here as a new standard; it is simply adopted here as a more stable basis than previous measures so that subsequent derived measures will be more precisely defined.

There are three principal derived measures discussed here. One measure is a count of the percent of time that the clipping level is exceeded. For example, if one clips 3 dB below the epl, what percent of the waveform would be affected? The second is the measure of power in a "phantom" interference signal. The clipping is assumed to be caused not by a voltage limiting process, but rather by a second interference signal added to the first. By estimating power in the phantom signal, a signal/noise ratio can be defined for different clipping levels. The third measure is power lost due to clipping.

III. USE OF CONTINUOUS SPEECH

Many objective measures of speech are strong functions of the "activity factor" or the percent of time a person is talking. For example, a person speaking at a fixed level can change his average power by varying his activity factor. In the present study, the quantities to be measured are dependent on the time base chosen.

In studies on clipping, there is little interest in what happens when speech is not present. It makes little sense to examine long silent intervals, which would occur during a telephone conversation. Therefore, this study's measures are restricted to only those times when speech is present.

The author knows of no speech detecting technique that will define intervals of speech activity in a manner that is insensitive to arbitrary choices of parameters such as detector threshold setting. This is shown, for example, in two previous studies in which substantial variation in detected speech patterns occurred with fairly small changes in detector parameters.[5,6] In the present study, the detection process is bypassed by using "continuous speech" material that contains a negligible number of perceived gaps.

The method of recording continuous speech is documented in Ref. 6. These recordings were prepared from recordings of experimental telephone conversations by manually splicing together sections containing a continuous flow of speech. There were eight male and eight female speakers each providing two separate recordings, except for two women who each made only one recording. In all, there were 16 male and 14 female continuous speech samples, the average length of each sample being about 55 seconds.

Measures made on these samples will be considered valid measures made "during speech." For example, if a threshold is exceeded here 3 percent of the time, this will imply that in noncontinuous "real" speech, the threshold is exceeded 3 percent "during speech" or "while the speaker is talking." The continuous speech tapes were edited by the author using a highly empirical procedure and personal judgment. Therefore, all measures reported here are dependent on the author's personal judgment, to the extent that the selection of continuous speech material is dependent on this judgment.

The 30 samples of continuous speech provide a basis for studying the variability of the desired measures over different *speech samples.* Note again that only 16 speakers are represented in the 30 samples, so that measures of variability are somewhat confounded if they are to be applied to variability over *different speakers.* This confounding is probably slight, since two samples from the same speaker are sometimes quite different in measures such as speech level and dominance of conversation (whether the person is basically listening or talking).

The speech samples of all the men were played back at a level high enough to use nearly the full range of a 12-bit A-to-D converter. Once the level adjustment for males was set, it was not changed from sample to sample. The level adjustment was then reset for the women, and all female speech samples were played at the new adjustment.

IV. RELATION OF EPL TO PEAKS

The epl reads the peak of an *assumed* log uniform distribution of instantaneous voltage, whose rms voltage above threshold is the same as that measured in the sample.* If speech were distributed exactly according to the log uniform distribution, then the epl would be the highest level or the "peak." In practice, the speech distribution does not abruptly and neatly terminate at some fixed level. Occasionally there are voltages exceeding the range of levels of the speech. The expected epl would be a "nominal" upper voltage limit, which would be exceeded occasionally during loud speech passages.

To get a quantitative measure of the relation between the epl and the highest instantaneous peaks, all 30 continuous speech samples were played into an A-to-D converter connected to a PDP-8 computer. The computer stored a histogram count of the measured voltages sampled at 1 kHz. After calculating the epl at the end of each speech sample, the number of A-to-D readings that exceeded the epl was counted.

On the average, the epl was exceeded 2.0 percent of the time ($\sigma =$ 0.29 percent) for female speech samples and 2.9 percent of the time

* The epl as defined in Ref. 1 has been revised as follows. Step 5a should read, "if $D \leq 6.75$, set $\Delta = (D - 2.75)/0.4$."

(σ = 0.46 percent) for male speech samples. The average of all 30 samples was 2.5 percent. In general, in noncontinuous or "standard" speech, the epl would be exceeded about 2.5 percent of the "time that speech is present."

The speech voltages that did exceed the epl occasionally reached levels of 7 or even 8 dB above the epl, but such events were very rare. The epl + 6 dB level was exceeded about one-tenth as often as the epl. That is, the epl + 6 dB was typically exceeded between 0.2 and 0.3 percent of the time. Thus, the epl + 6 dB appears to be a practical upper limit of the instantaneous speech levels.

V. PERCENT TIME CLIPPING LEVEL IS EXCEEDED

The log-uniform speech distribution model predicts that as a threshold (i.e., clipping level) is lowered, equal dB decreases in the threshold produce equal increments in percent time that speech exceeds the threshold.[7] This model is somewhat inaccurate in the immediate region of the epl, but it has been found to hold up over a wide range (30 or 40 dB) below the epl of most speech samples tested.[1,7] Eventually, for very low thresholds, the model must break down because one could continue lowering the threshold in equal dB increments to minus infinity; and one certainly cannot indefinitely add equal percent time increments.

Two experiments with continuous speech yielded data relevant to percent time above clipping level. The first showed that the epl itself is cleared about 2.5 percent of the time. This is one point on the "percent time vs clipping level" curve.

The second experiment sought to establish another point on the curve by obtaining a scatter plot of percent time over a range of thresholds for different speech samples. In an earlier study, percent time over a fixed threshold of −25 dBm was measured for 30 continuous speech samples. However, the samples were played at their original recorded levels rather than at levels equalized for A-to-D converter range. The epls had a sigma of 3.6 dB, which was a larger range than those obtained for the samples with equalized levels. If the epls had a fairly large range, it follows that epl minus threshold also had the same large range for a fixed threshold.

The scatter plot of percent time over threshold vs threshold (re epl) is shown in Fig. 1. The mean value of the 30 points occurs at the coordinates 15 dB for epl minus threshold, and 30 percent for time above threshold. Perhaps the simplest way to provide a linear fit to

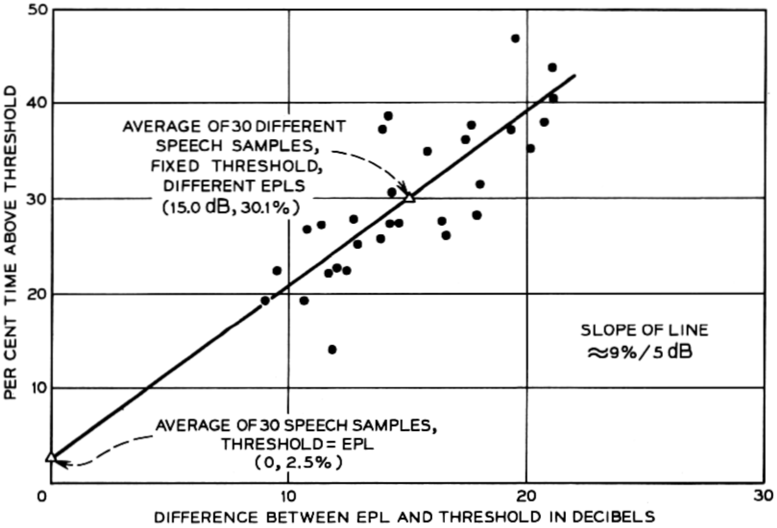

Fig. 1—Percent time threshold is cleared vs difference between epl and threshold for continuous speech.

the Fig. 1 data is to draw a straight line between the mean values for the two experiments.

Each of these two points determining the linear fit is an average of 30 samples. An alternative means of fitting a line to the data could be to use all 60 points in a least-mean-square fit. This might yield a more "accurate" fit to these data, but it must be remembered that the vertical axis, the "percent time" measure, is based upon an empirical definition of continuous speech. The straight line connecting the means of the two data sets is probably just as reasonable a fit if the curve is to be regarded as an approximate guide for application to speech circuit design. Note that in Fig. 1, the line passes within 5 percent divisions of most of the points, tending to justify the linear fit predicted by the idealized log-uniform speech distribution model. Previous work suggests that the linearity will be maintained until 30 or 35 dB below the epl.[7]

The line has a slope of 1.83 percent increase in time for every 1 dB drop in threshold, or approximately 9 percent for every 5 dB drop. This result can be combined with the conclusion of Section IV as a guideline for threshold clearance percentages:

Instantaneous voltages rarely occur at a level higher than 6 dB above the epl. During *continuous speech,* the epl is exceeded roughly 2.5 percent of the time. A threshold is cleared an additional 9 percent of the time for every 5 dB drop in threshold below the epl.

VI. RELATION OF EPL TO AVERAGE POWER

While obtaining the epls of the 30 speech samples, measures were also obtained of the long-term rms power in the continuous speech. The average epl minus rms for the 30 samples was 10.0 dB, with a standard deviation of 0.8 dB.

Therefore, during continuous speech, the speech power is about 10 dB below the epl. This result will be useful in the next section in relating speech power to clipping signal power.

VII. POWER IN A PHANTOM CLIPPING WAVEFORM

Instead of thinking of peak clipping as a limiting process, imagine it to be caused by the addition of a second phantom signal that will cause the original waveform to be restrained at some voltage. This process is illustrated in Fig. 2, showing the original signal, the phantom signal, and the result of adding them (i.e., the clipped signal).

The phantom signal can be considered as speech-correlated noise. If its power is known, then a signal/noise (S/N) ratio for clipped speech can be determined. In applying the S/N ratio, be careful to remember that the noise is speech-correlated, and hence may produce very different

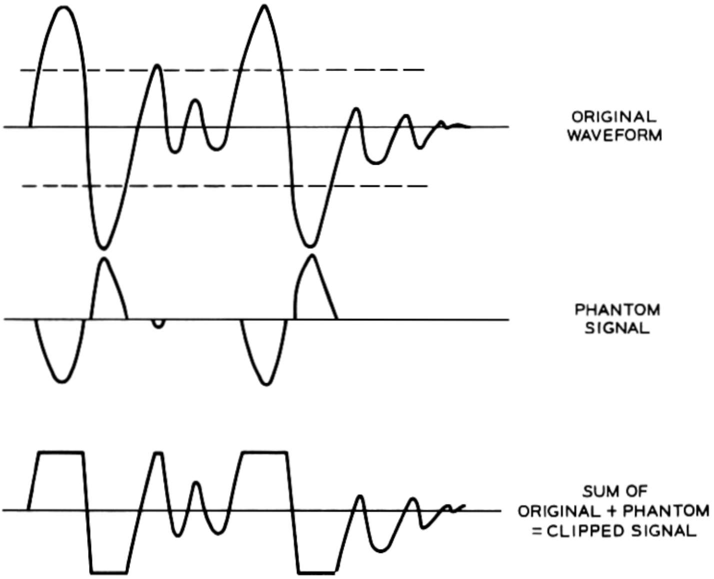

Fig. 2—Illustration showing how clipping can be thought of as due to an additive phantom signal.

effects on intelligibility, detectability, etc., than would a steady background noise with the same S/N ratio.

Note that this "power" of the phantom signal is a fictitious quantity, and no such *power* is added to the signal, even though a *voltage* is added. In fact, clipping produces a net power *loss* since it removes part of the signal.

The phantom signal power was calculated by playing a speech sample into the PDP-8 computer, and recording a histogram of the absolute magnitudes of all voltage levels obtained. When finished, the computer chose a fixed threshold level, T, and for each voltage level above this computed (Voltage − Threshold)^2. An accumulator was updated with this quantity n_T times, where n_T was the number of counts for that threshold.

The computer then divided the total $\sum (V - T)^2$ by the total number of A-to-D samples for the *entire* speech sample. The time base for averaging was the total length of the continuous speech sample, not the number of voltages exceeding T. Thus, the power for the phantom signal (the $V - T$ voltage waveform) was an average "during the time speech is present." The lower the threshold, the greater the power in the $(V - T)$ phantom noise signal would be, since $(V - T)$ was greater for each voltage and more voltages exceeded T.

The above process was done for six thresholds: −5, −10, $\cdots$, −30 dBm. For each speech sample a curve of phantom power vs clipping voltage was produced. These curves for all 30 speech samples are plotted in Fig. 3. Each speech sample curve has been normalized to its epl.

The author knows of no speech distribution model that would be appropriate for predicting the curves of Fig. 3. The log-uniform model, which works well for relating power to epl, is a poor fit in the voltage region close to the epl, since it incorrectly predicts an abrupt upper limit to the speech voltages at the epl. Lacking a model, the curves of Fig. 3 were treated as a scatter plot of independent points (each curve produced 5 or 6 such points) which were fitted with a third-degree polynomial. The curve fitting program, supplied by E. A. Youngs of Bell Laboratories, yielded the curve described by the equation below. In this equation, P = power in the phantom signal in dB (re epl) and T = clipping threshold in dB (re epl). (E.g., for clipping level 1 dB below the epl, $T = -1$.)

For all speakers,

$$P = -21.86 - 1.193T - 0.0443T^2 - 0.00057T^3.$$

This curve is plotted in Fig. 4.

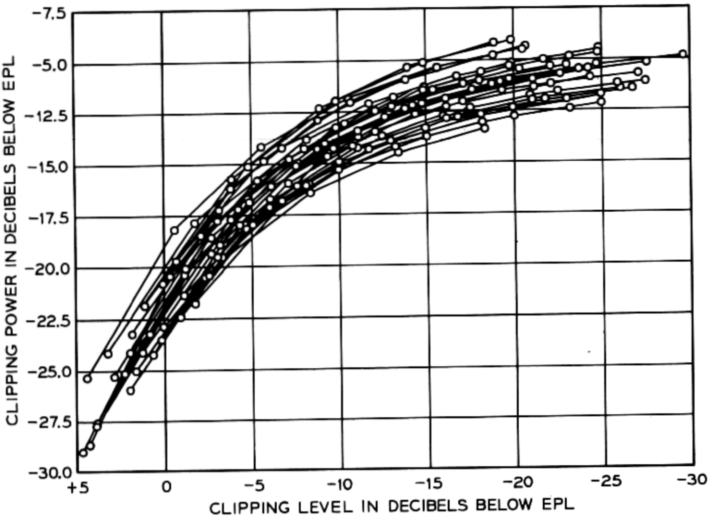

Fig. 3—Clipping power in a phantom clipping signal as a function of clipping level, for 30 continuous speech samples. Each curve is normalized to the epl for the associated speech sample.

VIII. POWER LOST FROM PEAK CLIPPING

If a signal is clipped, the resulting signal has less power than the original signal. The power loss in dB was calculated for all 30 continuous speech samples using six clipping levels for each sample. The 30 curves of power loss vs clipping level are superimposed in Fig. 5. The polynomial fit to these data, shown in Fig. 6, is given by

$$P = -1.084 + 0.373T - 0.012T^2,$$

where P = power loss (if 5 dB is lost, $P = -5$) and T = clipping threshold (if 5 dB below epl, $T = -5$). (A second-order polynomial was sufficient to meet the "fit criterion" of the polynomial curve fitting program.)

Also plotted in Fig. 6 is a curve of power loss vs clipping obtained by Wathen-Dunn and Lipke,[8] after an empirical instantaneous speech level probability distribution function developed by Davenport.[9] Wathen-Dunn defines his zero dB clipping level reference point as that point exceeded by only 0.1 percent of Davenport's speech signal, which consisted of having speakers read aloud from a book. Although such speech must have contained perceptible pauses, it is closer to "continuous speech" than a telephone conversation. To compare Wathen-Dunn's work with this, let us consider his 0.1 percent peak point for book

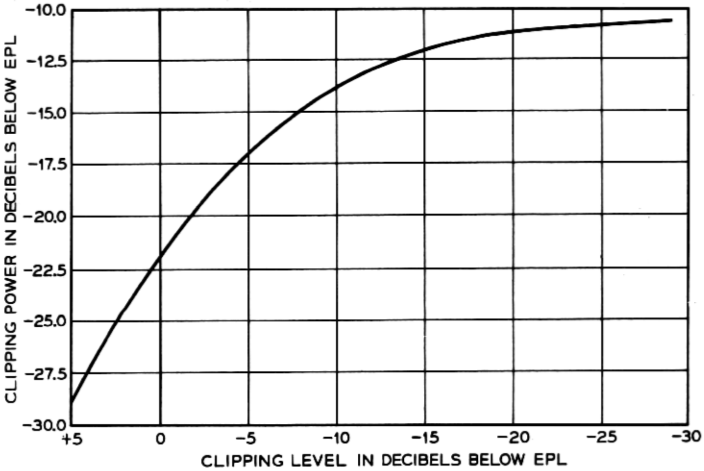

Fig. 4—Third-order polynomial fit to data of Fig. 3.

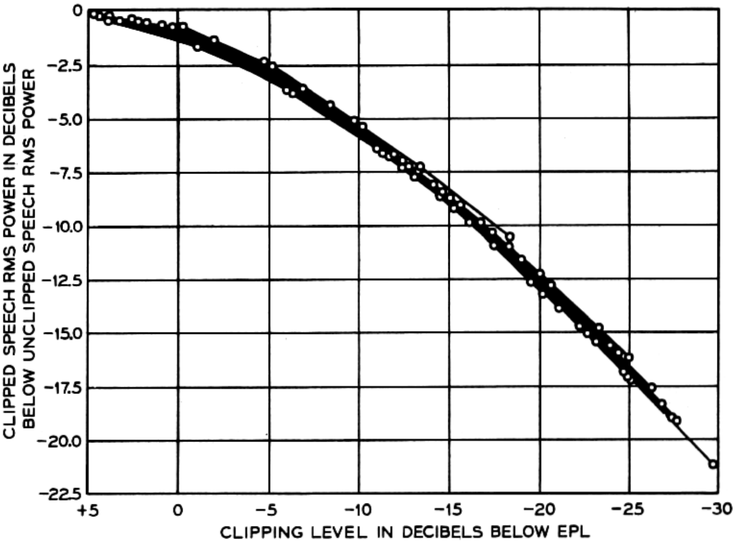

Fig. 5—Power loss vs clipping level for 30 samples of continuous speech.

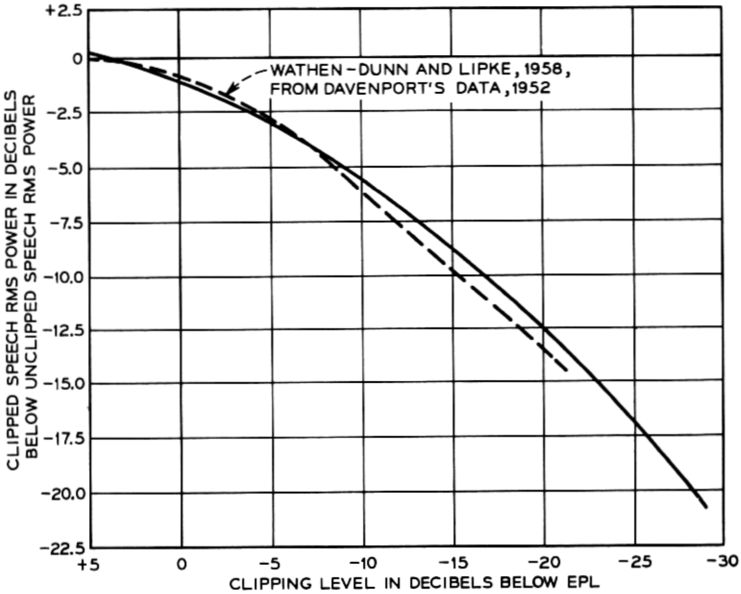

Fig. 6—Solid curve is a polynomial fit to data of Fig. 5. Dashed curve is taken from Wathen-Dunn and Lipke.[8]

reading as equivalent to the epl + 6 dB "peak" criterion established here for continuous speech. Fig. 6 of the present report is Wathen-Dunn's Fig. 6 with the horizontal axis shifted 6 dB; that is, what he calls 9 dB of clipping is plotted here as 3 dB below the epl.

In this study, the calculation of power loss is very close to Wathen-Dunn's and is somewhat a repeat of his work, except that the present study is a direct empirical calculation of the loss, rather than a theoretical calculation based on an empirical probability function.

IX. DIFFERENCES BETWEEN MALE AND FEMALE CONTINUOUS SPEECH

Some of the results of this study have been reported separately for male and female speech because there appears to be a real difference in the speech activity factor in the two sets of speech samples.

The average epls for men and women were adjusted to be roughly equal. The final average epls were measured as −5.86 dBm for women and −7.18 dBm for men. The women's levels therefore averaged 1.3 dB higher than the men's. However, measurements of average power (true rms) show that women averaged −16.46 dBm, and men averaged

−16.02 dBm. While the women's speech samples had higher levels, they had a lower average power than the men's.

This can be explained if during the editing process to produce continuous speech, the author allowed more short gaps in the women's speech than the men's, resulting in a lower activity factor for women. This is borne out by measurements of percent of A-to-D samples crossing a fixed −25 dBm threshold, which show an average of 35.9 percent for women, and 48.3 percent for men. This difference would cause the higher level women's speech to have less average power.

Once again, we are back to the problem of determining "when speech is present." Even though all editing was done in a two month period by the author, and even though the same telephone circuit, recording equipment, and editing equipment was used for both sets of data, consistent differences exist between the data for men and women. This finding should be a further warning that the results found here are intended only as guidelines and not as rigid specifications, since they can be influenced by the arbitrary design of the speech detection process.

X. RELATION OF PEAK CLIPPING TO VU

In an earlier unpublished study, the author suggested that VUs might be obtained from epls by subtracting 11 dB from the epl. This was based on extensive data from one highly trained observer, plus sample data from two other observers.

In the present study, the epl + 6 dB is suggested as a practical upper bound to the speech waveform. Thus, using the 11 dB epl to VU conversion, the expected VU would be about 17 dB below the highest peaks. This result is remarkably close to Noll's estimate of 17.2 dB as the peak-VU factor.[4]

It would appear that the data presented here might be applied to VU data by using the 11 dB conversion for epl to VU. A large body of VU data exists from field trials and the present results might be applied to these data. The author has also shown that VU data taken from many observers can be extremely erratic and at times can seem almost random.[3] Because of the variability in VU measurements, care should be taken in combining the results of the present study with existing VU data.

XI. SUMMARY

This study sought to examine speech signal statistics relevant to peak clipping and threshold crossings. All measurements were made on continuous speech. The following results were obtained:

(i) Instantaneous voltages rarely occur at levels higher than the epl + 6 dB.

(ii) The epl is exceeded by the instantaneous waveform about 2.5 percent of the time.

(iii) An additional 9 percent of time above threshold is gained for every 5 dB drop in threshold below the epl.

(iv) The average power in continuous speech is about 10 dB below the epl.

(v) The power in a "phantom noise signal", representing speech-correlated clipping noise, can be approximated by the curve in Fig. 4.

(vi) The power loss in continuous speech resulting from clipping can be approximated by the curve in Fig. 6.

(vii) VUs can be roughly approximated by subtracting 11 dB from epls. The other results might then be appropriately modified for VU data.

REFERENCES

1. Brady, P. T., "Equivalent Peak Level: A Threshold-Independent Speech-Level Measure," J. Acoust. Soc. Amer., *44*, No. 3 (September 1968), pp. 695–699.
2. Shearme, J. N., and Richards, D. L., "The Measurement of Speech Level," Post Office Elec. Eng. J., *47*, 1954, pp. 159–161.
3. Brady, P. T., "Need for Standardization in the Measurement of Speech Level," J. Acoust. Soc. Amer., *50*, No. 2 (August 1971), pp. 712–714.
4. Noll, A. M., unpublished work.
5. Brady, P. T., "A Technique for Investigating On-Off Patterns of Speech," B.S.T.J., *44*, No. 1 (January 1965), pp. 1–22.
6. Brady, P. T., "A Statistical Analysis of On-Off Patterns in 16 Conversations," B.S.T.J., *47*, No. 1 (January 1968), pp. 73–91.
7. Brady, P. T., "A Statistical Basis for Objective Measurement of Speech Levels," B.S.T.J., *44*, No. 7 (September 1965), pp. 1453–1486.
8. Wathen-Dunn, W., and Lipke, D. W., "On the Power Gained by Clipping Speech in the Audio Band," J. Acous. Soc. Amer., *30*, No. 1 (January 1958), pp. 36–40.
9. Davenport, W. B., "An Experimental Study of Speech-Wave Probability Distributions," J. Acoust. Soc. Amer., *24*, No. 4 (July 1952), pp. 390–399.