Amplitude Distributions of Telephone Channel Noise and a Model for Impulse Noise

By J. H. FENNICK

(Manuscript received June 30, 1969)

The noise waveforms found on voice bandwidth telephone channels are generally recognized to be non-gaussian in their amplitude distribution. This paper presents data which suggests that a simple exponential is a good function to describe amplitude densities in the extreme tails.

A comprehenisve model of impulse noise as viewed on trunk groups is then presented. The model relates the distributions of impulse noise levels and impulse noise counts.

I. INTRODUCTION

Noise on telephone channels has been measured for years with instruments which are constructed to enable reasonably good correlations between the reading obtained and the annoyance of the noise during a telephone conversation.[1] Fluctuations of the meter pointer during a measurement are either ignored or mentally averaged by the observer, depending upon their frequency of occurrence and their magnitude. With the introduction of data transmission on the telephone network, the relatively frequent high amplitude excursions of the noise waveform were viewed as a "new" kind of noise, primarily because they were generally not annoying in voice communication and it was recognized that no meaningful measure of them could be obtained with the standard noise measuring sets. The term "impulse noise" was applied to these high excursions and new instruments were designed to measure them.[2]

The significance of impulse noise in data transmission has given rise to a great deal of effort devoted to its measurement, characterization, and evaluation as a transmission impairment.[3–6] (For an extensive bibliography, see Ref. 3.) Several models have been suggested to describe the erratic behavior and clustering phenomena associated with this type of noise. The Pareto model of Berger and Mendelbrot and the generalized hyperbolic model proposed by Mertz appear to be the best presented to date.[3,7] A more mathematically tractible (than the hyperbolic) model has recently been applied to error rate data by Fritchman. He proposed a partitioned Markov chain model which would seem to show promise in this area although it does not seem to have been applied to impulse noise data as yet.[8] The model presented here does not deal specifically with the intervals between occurrences of noise pulses but is concerned directly with the number of occurrences per unit time above any threshold (in decibel) of observation. Extrapolation of occurrences of noise pulses to errors created in data transmission is a function of many parameters besides the occurrence of noise and will not be discussed here although good prediction techniques exist.[5]

In order to set the background for the discussion of impulse noise as a separate phenomenon, as opposed to the background noise or as a part of the composite noise waveform on a channel, data are first presented on the amplitude probability density function of the noise as observed and comparisons made with gaussian noise. The data reflect only the range of variables encountered and should not be considered as statistically describing the amplitude distributions of noise on telephone channels.

II. IMPULSE NOISE AS A DISTINCT PROCESS

Typical oscillograph noise waveforms from a random noise generator and from a telephone channel are shown in Fig. 1. Each trace is 200 ms long and both have the same rms value. The upper one is from the noise generator, the lower one from a telephone channel. The occurrence of two "impulses" are shown near the left end of the lower trace. It is primarily the occurrence of such "pulses" that make real channel noise decidedly different from band-limited white gaussian distributed noise (the upper trace).

Figures 2a and b show two such impulses extracted from a noise recording, sampled at a 15 kHz rate and analyzed to determine their amplitude and phase characteristics in the frequency domain. In both cases, the phase characteristic is shown to be relatively smooth, but the frequency content highly variable. Similar analyses on about 2000 noise pulses verified these observations. However, if a large sample of pulses, on the order of 200, is taken from a given channel, the average spectrum appears to be approximately the shape of the channel gain-frequency characteristic—not a very surprising result. Such an averaging is shown in Fig. 3.

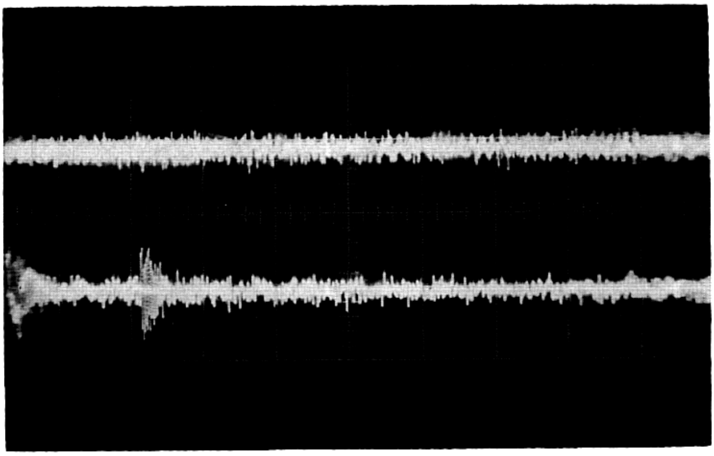

Fig. 1 — 200 ms samples of random noise and telephone channel noise with equal rms levels.

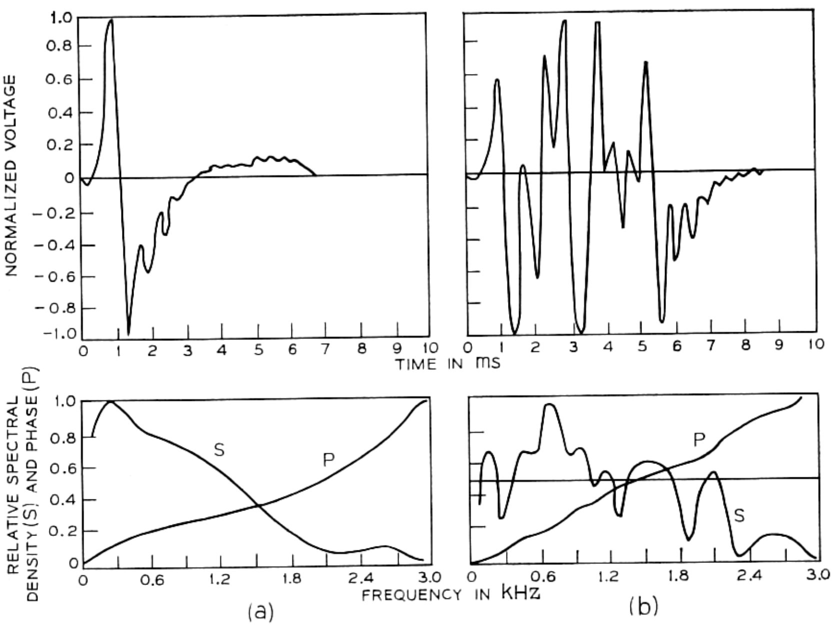

Fig. 2 — Samples of impulses extracted from telephone channel noise with their amplitude and phase characteristics in the frequency domain.

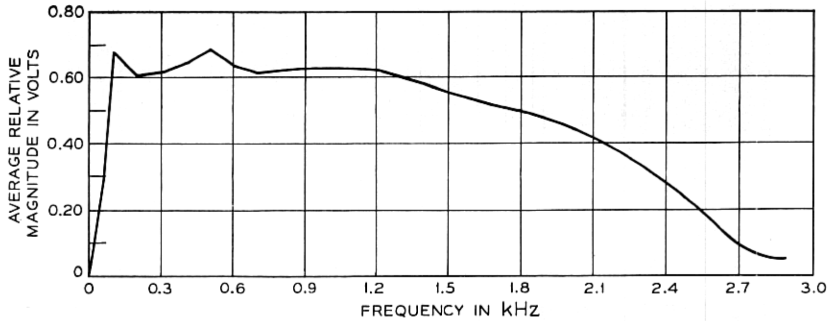

Fig. 3 — Average spectral content of about 200 impulses from a single telephone channel.

Figures 1 and 2 serve as partial justification for treating impulse noise as a separate phenomenon. The pulses shown in Fig. 1 do not rise to strikingly high amplitudes compared to the rest of the noise waveform. Those in Fig. 2, however, are so large that the scale prohibits viewing the background noise waveform which continues beyond that shown. This extreme peaking will become more apparent in Section III.

III. PERCENT OF TIME WAVEFORM IS WITHIN AN INTERVAL

The percentage of time that the noise is within a given interval (± 0.5 dB in this case) is a useful means of describing a random waveform. Data in the form of histograms were obtained by sampling, at a 10 kHz rate, 30 minute tape recordings of telephone channel noise. Equipment limitations imposed a usable dynamic range of 30 dB, so the apparatus was adjusted to examine only the extreme peaks of the noise. In practice this usually required that the noise be examined at levels corresponding to percentages of 10^{-2} or less. This approach was also consistent with the nature of the problem—the relatively high noise amplitudes were of greatest interest. Logarithmic compression and decibel scaling were used and resulted in a unique presentation of the data. Instead of the usual scaling in voltage, the abscissa is scaled in decibels removed from the rms value of the noise. A negative sign preceding an abscissa value refers simply to one polarity of noise waveform, a positive sign refers simply to the opposite polarity. Zero on the abscissa corresponds to the rms value of the noise waveform. For convenient comparison, the equivalent data for a gaussian distribution are also shown in each of the figures presented. The ordinate, proportion of time the waveform is within $\pm\frac{1}{2}$ dB of the indicated level, is presented in powers of 10 from 10^{-2} to 10^{-8}.

Figures 4 and 5 show the histograms as measured on two different channels. Figure 4 was taken from data recorded on a coaxial cable system and Fig. 5 from a microwave radio system. The striking departure from a stationary gaussian process is obvious. The sampling rate of 10 kHz over a 30 minute period resulted in 18×10^6 samples. Values on Fig. 4 of 5.5×10^{-8} represent one sample in 18 million and can hardly be considered significant. The values of 10^{-8} shown on Fig. 4 represent voids in the data. Figure 6 shows a histogram constructed by combining seven 30 minute recordings, and so represents an "average" histogram over 3.5 hours of real time. The result is surprisingly linear for values below about 5×10^{-5} and suggests that the tails of the amplitude distribution of real channel noise are approximated quite well by a simple exponential.

A total of 37 half hour recordings were analyzed in this fashion. Seventeen of these were taken from microwave radio channels and 20

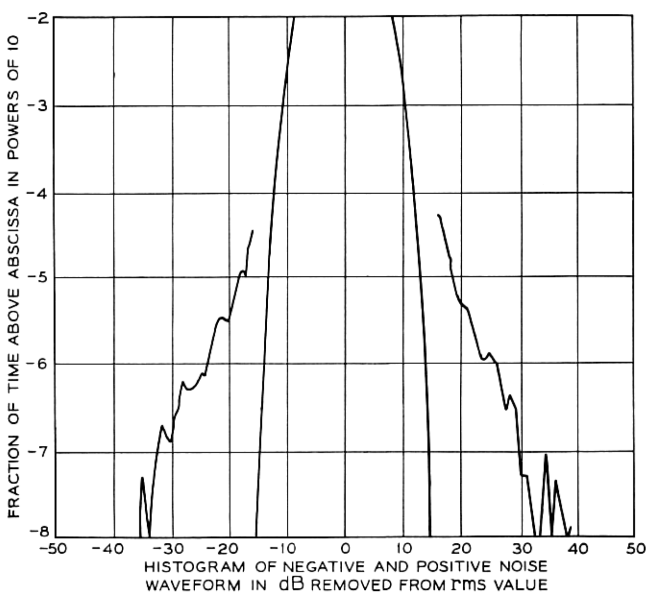

Fig. 4 — Histogram of telephone channel noise amplitudes compared with gaussian distribution. Sample from a coaxial transmission system. Proportion of time that the noise waveform is within ±0.5 dB of abscissa value. Abscissa is decibel removed from rms.

from various types of cable or coaxial carrier systems. The variability observed on the microwave systems is much greater than that on cable systems so the two sets of data are treated separately.

Since no data are available on the amplitude histograms at values in excess of 10^{-4}, it is assumed here that the histogram for such values is represented by a truncated normal function. The observed data suggest that, if the noise is stationary, its amplitude density function then may be written:

$$p(x) = \begin{cases} 0; & x < -b \\ ce^{kx}; & -b \leqq x \leqq -a \\ \hat{\Phi}; & -a < x < a \\ ce^{-kx}; & a \leqq x \leqq b \\ 0; & x > b \end{cases} \tag{1}$$

where

$\pm b$ = realistic bounds on the voltage waveform (channel saturation),

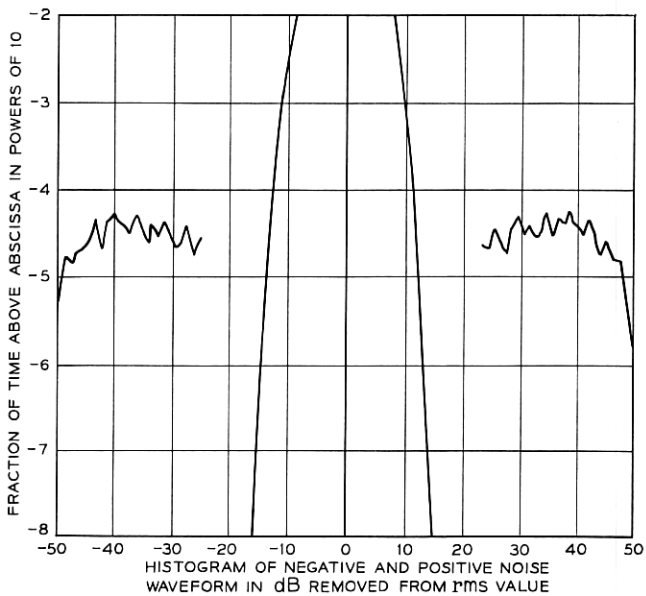

Fig. 5 — Histogram of telephone channel noise amplitudes compared with gaussian distribution. Sample from a microwave radio system. Constructed as in Fig. 4.

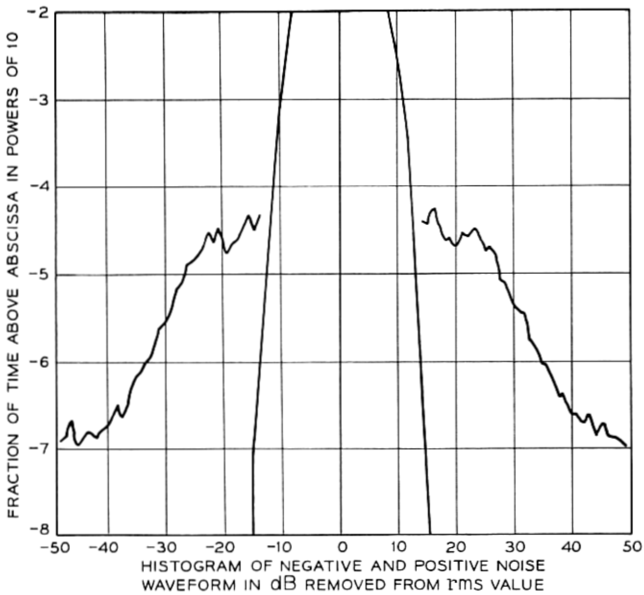

Fig. 6 — Histogram of noise amplitude from cable and coaxial systems taken over 3½ hours. Constructed as in Fig. 4.

$\pm a$ = points of departure of the density function from an assumed underlying Gaussian,

$\hat{\Phi}$ = gaussian density truncated at $\pm a$.

c, k = parameters describing the exponential density function.

The value of b ranges from about 30 to 50 dB*, the value of a ranges from about 10 to 15 dB*, and k may be negative as illustrated in Fig. 5. The variable c ranges over 14 orders of magnitude from 10^{-7} to 10^{7}. No significant correlations were found between any of the variables in the data analyzed. The point of departure from the assumed underlying truncated gaussian distribution a is given by the positive solution of the quadratic

$$a = k \pm \{k^2 - 2 \ln [c(2\pi)^{\frac{1}{2}}]\}^{\frac{1}{2}}.$$

Some values of k and c are given below.

3.1 *Histograms on Cable and Coaxial Carrier Systems*

As stated earlier, the histograms on cable and coaxial carrier systems showed less variability than those on microwave radio systems. In fact,

* That is, above the rms value.

two of the 20 observations tracked the assumed gaussian distribution to within less than $\frac{1}{2}$ dB over the entire range from 10^{-3} to 10^{-7}. These two observations lend credence to the assumption of an underlying gaussian process and also show that at least two channels had no impulse noise in the sense of the term as defined in Section I.

The data are summarized in two ways. First, the values of the variables k and c were examined, and then the intercepts (abscissa values) for various values of proportion were studied.

For cable and coaxial carrier systems (20 samples)* the mean of k was 0.45 and the estimated standard deviation s was 0.46. Because of the extreme range of c, as mentioned in Section III, only the median appears to be of interest; it was found to be 0.0028. A probability density function using the mean value of k and the median of c is shown in Fig. 7. The resultant exponential departs from the guassian distribution at about 4.5×10^{-6} on the ordinate.

The second method of examining the data is considered to be more meaningful in terms of a representative average. The intercepts at proportion values of 10^{-4} to 10^{-7} were studied. The mean $\langle x \rangle_{av}$ (in dB), estimated standard deviation s, median, and 90 percent confidence intervals (CI) about the mean, are shown in Table I. The average and median functions so derived are also shown in Fig. 7. The distributions of intercepts were found to be very nearly log-normal for all four proportion values (10^{-4} through 10^{-7}). This explains the differences between the means and medians as in Table I and Fig. 7. A skew distribution of the intercepts is of the form to be expected. A lower bound on the intercept is imposed by the gaussian assumption and a gradual tailing off of the values at the high range might be expected.

Taking the median value of the exponential distribution as being a representative value of conditions on cable carrier systems, it is of interest to compare tail values of the resultant cumulative distribution function (CDF) with the gaussian distribution. The median exponential intercepts the gaussian at an x value 12.6 dB above the rms. This corresponds to the $\log^{-1}(12.6/20) = 4.26\sigma$ point. If the noise amplitude were truly gaussian, only 0.004 percent of the waveform would lie beyond the $\pm 4.26\sigma$ points. However, 0.0134 percent of the area lies below the exponential portion of the density function, nearly a full order of magnitude difference. This sheds some light on the predicted performance of data systems, for instance, in the presence of gaussian noise, a typical analysis situation, and that actually observed in a working version of the system over real channels.

* Each "sample" is 30 minutes of time.

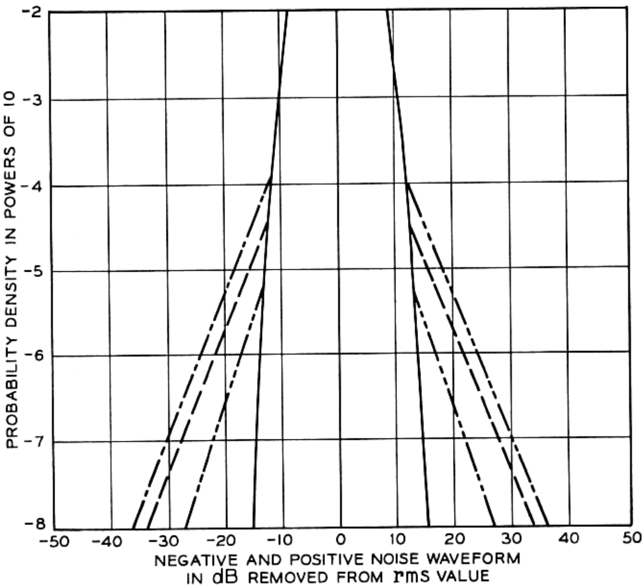

Fig. 7 — Amplitude probability density functions. Estimated average and median taken over constant density values and derived from mean parameter k and median parameter c, compared to gaussian. Averages taken over 10 hours of noise on cable carrier systems. — gaussian; — - — average; — — — median, — - - — mean k, median C.

3.2 *Histograms on Microwave Radio Systems*

The data for the microwave systems were analyzed in the same way as the second method for the cable systems. The individual values of k and c were not computed because of the dubious value of such an effort. Averaging over the intercepts for constant values of density (method 2) illustrates the greater variability in the microwave systems. The results, presented in Table II, show this by the larger estimated standard deviations and wider 90 percent confidence intervals about the

Table I—Estimated Probability Densities for Cable and Coaxial Carrier Systems

Probability Density	$\langle x \rangle_{av}$ (dB)	s (dB)	Median (dB)	90% CI (dB)
10^{-5}	19.3	4.4	15	1.5
10^{-6}	24.6	4.9	23	1.8
10^{-7}	30.2	5.5	28	2.0

TABLE II—ESTIMATED PROBABILITY DENSITIES FOR MICROWAVE RADIO SYSTEMS

Probability Density	$\langle x \rangle_{av}$ (dB)	s (dB)	Median (dB)	90% CI (dB)
10^{-4}	17.7	10.7	12.6	5.5
10^{-5}	23.2	12.8	18.5	5.2
10^{-6}	24.7	9.7	22	4.1
10^{-7}	29.6	9.4	27	4.1

estimated means (compare with Table I). The median function so derived is shown in Fig. 8. The distributions of intercepts were again found to be closely approximated by the log-normal distribution and the median curve examined as for the cable carrier systems. The median exponential intercepts the gaussian distribution at 11.6 dB above the rms value. This corresponds to $\log^{-1}(11.6/20) = 3.8\sigma$, or 0.0165 percent of the noise waveform that would lie beyond $\pm 3.8\sigma$ of a gaussian distribution. The values of k and c for the median curve on Fig. 8 are 0.48 and 0.042. Integration of the resultant exponential function over the

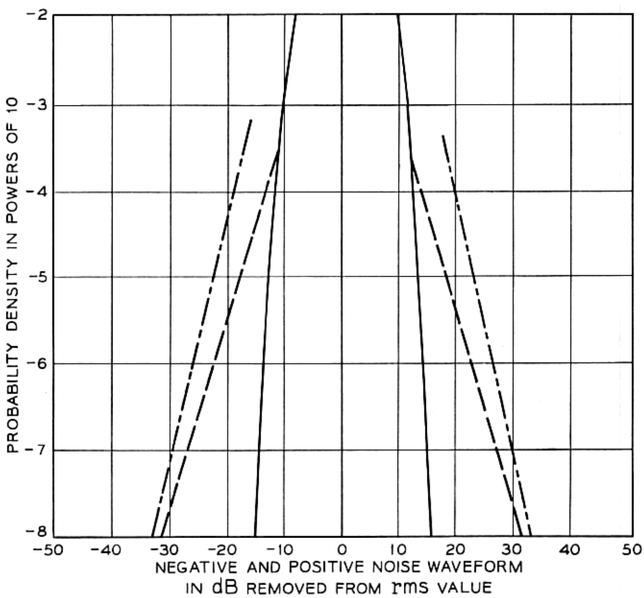

Fig. 8—Amplitude probability density functions. Estimated median taken over constant density values from 8½ hours of noise from microwave radio systems. — gaussian, - - - average, — — — median.

appropriate intervals yields 0.068 percent of the median waveform in excess of the 3.8σ points of an assumed gaussian. In this case, the median difference is about a factor of four.

IV. FORMAL DEFINITION OF AN IMPULSE AND SOME PULSE LENGTH DATA

In the process of impulse noise analysis a formal definition is required. The definition is illustrated in Fig. 9 and was first proposed by Kaenel, and others.[9] The waveform illustrated in Fig. 9 represents an ideally rectified noise waveform being sampled by an A/D converter. All portions of the noise waveform that remain below a variable slicing level, designated level 2, are considered as part of the underlying band-limited white gaussian or background noise until level 2 is exceeded. Once level 2 is exceeded, the noise pulse, or impulse, is measured starting at the point where level 1 was exceeded as indicated in the figure until it returns below level 1 and remains for a specified amount of time referred to as a guard interval. The function of the guard interval is to distinguish between nodes of a single impulse and two impulses which occur close together in time. Various guard intervals have been used in the analysis of voiceband impulse noise, from 0.3 ms to 0.8 ms. The choice is somewhat arbitrary, but on the basis of the author's unpublished interpulse time distributions, his choice is 0.6 ms as an optimum value. This is preferred because interpulse gap length histograms commonly exhibit a null at about 0.6 ms. The adjustment of levels 1 and 2 vary, but level 1 is typically 10 dB above the rms value, and level 2

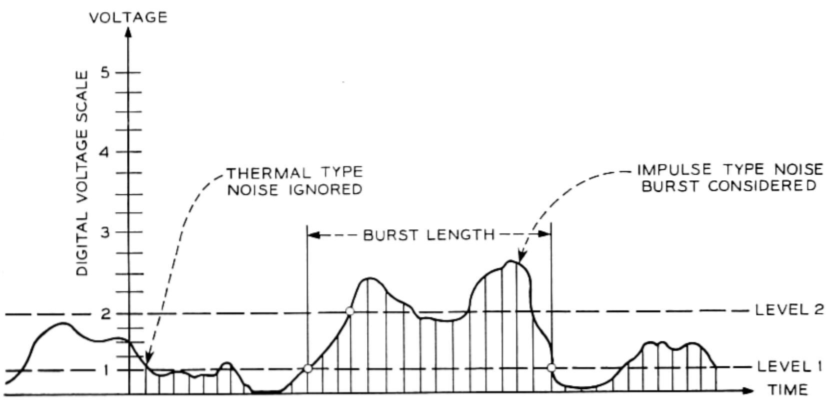

Fig. 9 — Ideally rectified noise waveform illustrating definition of pulse length.

from 13 to 16 dB above rms. In the context of this formal definition, the impulse has been referred to as a burst.[9]

Under the rules of the definition, just given, frequency functions were constructed for the lengths of several thousand impulses. A set of these are shown in Fig. 10; the set is shown as "envelopes of all the observed frequency functions." Only two points appear to be significant. The modes of the functions occur at about 1.2 ms, and lengths in excess of 10 ms are almost never observed. The remainder of this paper discusses an impulse noise model.

V. A MODEL FOR IMPULSE NOISE ON TELEPHONE CHANNELS

This section describes impulse noise as viewed on a trunk group as it is used by a switched network subscriber. The distributions of the peak amplitudes of individual impulses have been of interest for some time, and extensive data concerning them have been collected.[2-4,6] Methods of relating such distributions to data system performance have also been derived.[5]

The data are most frequently collected by means of simple threshold detectors. Excursions of the noise waveform above the threshold are recorded on electromechanical counters.[2] Such measuring devices have finite counting rates which may be exceeded at times by the rate of occurrence of impulses in clusters. For this reason, from this point on, "counts" referring to values recorded by the instruments will be used instead of the word "impulse." The count process necessarily differs in some respects from the impulse noise process for the reasons just cited.

5.1 *Terminology and Definitions*

Some jargon has accumulated in the area of impulse noise studies; it is sometimes conflicting as well as confusing. The following terminology is adopted here.

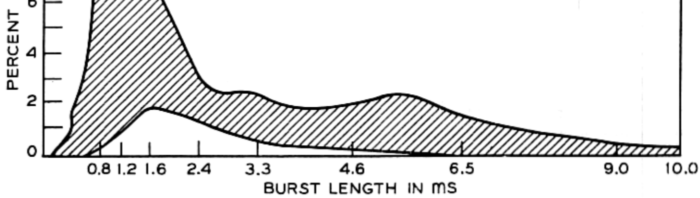

Fig. 10 — Envelopes of length density functions derived under definition of Fig. 9.

(*i*) Count—Refers to a number registered on the counter of an impulse noise threshold detecting type of measuring instrument, set at a specified level, during a specified measurement interval. (The count may be less than the actual number of impulses which exceeded the measurement threshold during the interval because of the finite maximum counting rate of the instrument.) Upper case C denotes the random variable count.

(*ii*) Impulse Noise Level—A level, expressed in decibels, at which the recorded count in a specified measurement interval is equal to some specified count denoted C_o.

(*iii*) Level Distribution—A distribution of levels, expressed in decibels (dBm, dBrn, and so on), taken across a number of channels, at which a specified count C_o is recorded in a specified measurement interval. Script "ℓ" denotes the random variable level.

(*iv*) Count Distribution—A distribution of counts observed in measurements on a number of channels taken at a specified level.

(*v*) Log-Count Distribution—A distribution of the logarithms of counts, expressed in decibels. Upper case "D" denotes the random variable log-count and is defined: $D = -10 \log_{10} (C/C_o)$, where C_o is an arbitrarily "specified reference count" greater than zero. C_o is arbitrary, but once picked it must be held constant for its associated level distribution.

(*vi*) Amplitude Distribution—A cumulative distribution of the peak amplitudes of individual impulses on a single channel. The average complementary distribution is linear on semi-log paper for counts in the range of interest; that is; $C < \approx 300$ in 30 minutes.

(*vii*) Slope—When spelled with "S", Slope refers to the slope of the peak amplitude distribution. Through common usage, the number assigned to Slope is the negative reciprocal of the slope of the peak amplitude distribution, designated "m", and expressed in decibels per decade of counts.

5.2 *General Comments on Level and Count Distributions*

Sample level distributions are constructed from data obtained through the use of multilevel impulse counters which record the number of counts at several levels, each separated by 2 to 6 dB, occurring during a prescribed measurement period. Level distributions may be constructed from the data depending upon the specific number of counts C_o in which one is interested. Suitable interpolation between the levels actually observed permits one to estimate the level at which some specified number of counts C_o actually occurred. Thus each level distribution has a number C_o associated with it, as well as a specific measurement interval. The primary data used in this study consists of level distributions of 15 counts in 15 minutes and 90 counts in 30 minutes.[4,6]

The number of counts observed in a multilevel measurement tends to decrease exponentially as the level, in decibels, increases. Some departures from this rule are observed in individual measurements, but the average amplitude distribution taken over a large number of measurements in a single class of trunks appears to be exponential.[6] The number of counts C, at any level ℓ, may be estimated from the number of counts C', at level ℓ', by the empirically derived relation

$$C = C' \exp [(\ell' - \ell)/(Mm)]. \tag{2}$$

where $M = (\log_e 10)^{-1}$. Because different types of transmission facilities exhibit different impulse noise properties, the average noise level and average Slope vary over an appreciable range as facilities change. However, within a given type of facility or within a class of trunks, greater homogeneity is observed.[6] The model is therefore directed at a description of the noise as observed within populations of transmission channels on a single type of facility which are common to some larger grouping such as a trunk group.

5.3 *Assumptions*

The following two assumptions, supported by studies of available noise data, are basic to the model which is presented in Section 5.4.

(*i*) Level distributions for a specified count C_o are normal with mean ℓ_o and standard deviation σ_ℓ.

(*ii*) σ_ℓ is independent of C_o within a given trunk class. The first assumption is the most reasonable in view of the data; there are conflicting data concerning the second and it appears to be more valid for compandored facilities than for noncompandored facilities.[4,6] Under these assumptions, and one more stated below, it is shown below that the count and level distributions are completely described by the parameters associated with one level distribution: C_o, ℓ_o, σ_ℓ, and the Slope m. The Slope is estimated by the straight line connecting the mean of the level distributions for different choices of C_o.[6]

5.4 *The Model*

Any number of level distributions may be obtained from the data by choosing different values of C_o. As C_o increases, the corresponding level ℓ_o will decrease and trace a path in the count-level plane given by equation (2). A family of such level distributions form a probability density surface above the count-level plane with normal cross sections parallel to the level axis. Such a surface is illustrated in Fig. 11. Under assumption (*ii*), lines parallel to the mean Slope are projections of constant probability density with the same functional form as equation (2). One of two cross sections may be taken which will define a probability density function. If the cross section is parallel to the count axis, a count distribution results. To see this more clearly, consider an experiment where impulse noise measurements are made on a group of similar trunks. A value for C_o is chosen and the associated level distribution with mean ℓ_o is found. The distribution will be normal with standard deviation σ_ℓ. Another value of C_o is chosen and a second level distribution is constructed. It will have the same standard deviation as the first. The experiment may be repeated any number of times to construct the family of distributions illustrated in Fig. 11.

In the experiment just described, the noise level ℓ associated with C_o was the random variable. Now suppose one wishes to let the count, or log-count, be the random variable while holding ℓ fixed. It is noted that equation (2) is the relationship between the means ℓ_o and C_o. Assume for the moment that equation (2) holds completely and is indeed a fixed relation between the two possible random variables, ℓ and C. Equation (2) may be rewritten, with $\ell' = \ell_o = 0$ and $C' = C_o$ as this constitutes an arbitrary shift in the decibel scale to define $\ell' = 0$:

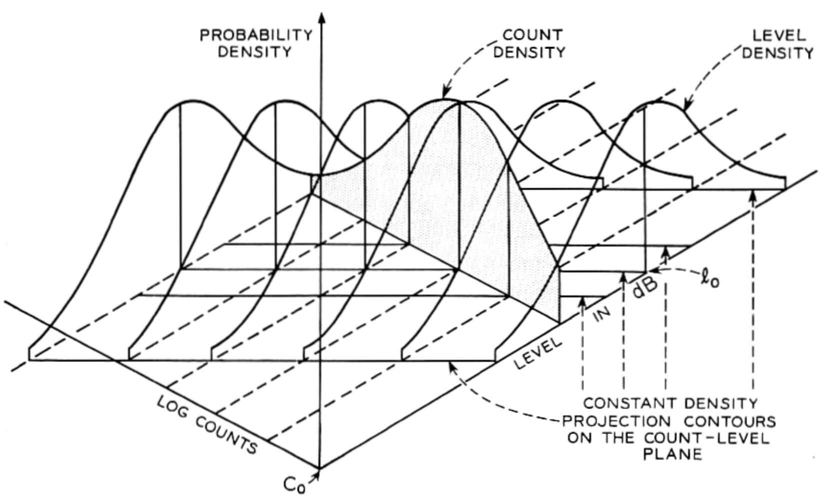

Fig. 11 — Probability density surface for the impulse noise count process on trunk groups.

$$\ell = -m \log_{10} (C/C_o). \tag{3}$$

Define $D = -10 \log (C/C_o)$. Then $\ell = Dm/10$, and the log-count distribution, the probability that D is $\leqq$ some value x, is by assumption 1,

$$P[D \leqq x] = P[\ell \leqq xm/10] = \frac{1}{\sigma_\ell (2\pi)^{\frac{1}{2}}} \int_{-\infty}^{xm/10} \exp(-\ell^2/2\sigma_\ell^2)\, d\ell. \tag{4}$$

The density function $f(x)$ is found to be

$$f(x) = \frac{m}{10\sigma_\ell (2\pi)^{\frac{1}{2}}} \exp[-m^2x^2/200\sigma_\ell^2]; \qquad -\infty \leqq x \leqq \infty. \tag{5}$$

Thus D is approximated by a normal distribution with mean zero and standard deviation $\sigma_D = 10\sigma_\ell/m$.*

In the previous derivation, equation (2), a relationship between expected values was assumed to hold as a mapping between the random variables ℓ and C or ℓ and D. To check the validity of this assumption a second experiment can be performed on the data collected in the first. The level ℓ can be held fixed at ℓ_o, and the count distribution at ℓ_o obtained by interpolation as described earlier. The observed log-count distribution may be compared with that derived in equation (5). This is done in Section 5.5.

5.5 *A Check on the Model Using Count Distribution Data*

Figure 13 is an example of count distributions derived in three different ways from a set of data consisting of 127 measurements on non-compandored carrier facility trunks 1,000 to 2,000 miles in length. The level distribution for these data, with $C_o = 15$, is slightly skew, the mean is 6.129 dBrn and the median 61.8 dBrn. The count distribution at 61.8 dBrn, obtained by interpolation between levels measured, is shown by the circled points on the figure. A point-by-point mapping from the level distribution by use of equation (2) is shown, as well as the log-normal one predicted by equation (5). The coincidence of all three sets of data is striking.

VI. THE TIME VARIABILITY OF IMPULSE NOISE

An additional check on the validity of this model is provided by its implications in the time variability of the noise. To see this, one additional assumption is made, and predicted and observed results serve to

* As a matter of interest, values of σ_D calculated from the 1964 Intertoll Trunk Survey (Ref. 6), are shown in Fig. 12.

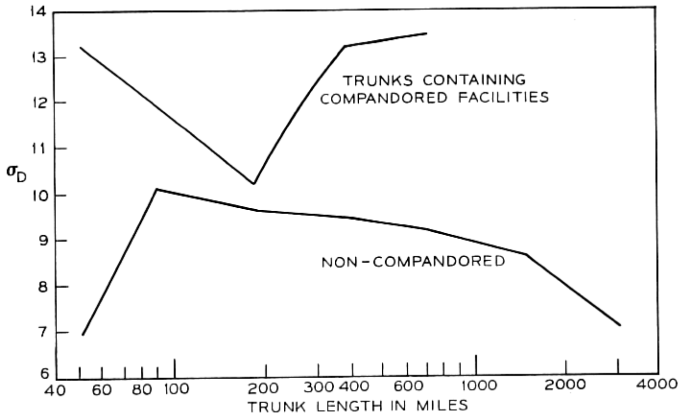

Fig. 12 — Values of standard deviation of log-count distributions $F(D)$ as observed on Bell System trunks.

validate both this additional assumption and the preceding model.

Consider making impulse noise measurements on a large number of channels at a fixed level ℓ_o and recording the cumulative count on the ith channel C_{ni} at times nT, $n = 1, 2, \cdots$. Now assume that the accrued count is a linear function of time so that each total count C_{ni} after time nT may be estimated by $C_{ni} = nC_{1i}$, where C_{1i} is the count in the first interval T on the ith channel. If the same reference count C_o is retained in the definition of D (log-counts), for all time intervals, then the mean value of D will increase as log (n) but the variance of the count distribution (as opposed to the log-count) behaves differently however, as shown by the following.

Under the assumption that equation (2) holds as a mapping between ℓ and C, the distribution of C (counts) may also be derived:

$$P[C \leqq y] = P[\ell \geqq -m \log (y/C_o)]$$

$$= \frac{1}{\sigma_\ell (2\pi)^{\frac{1}{2}}} \int_{-m \log y/C_o}^{\infty} \exp (-\ell^2/2\sigma_\ell^2)\, d\ell, \qquad (6)$$

and the density of C is approximated by the log-normal:

$$f(y) = \frac{Mm}{\sigma_\ell (2\pi)^{\frac{1}{2}}} y^{-1} \exp \left[-\frac{M^2 m^2}{2\sigma_\ell^2} \ln^2 (y/C_o) \right];$$

$$0 \leqq y \leqq \infty,$$

$$\ln \equiv \log_e .$$

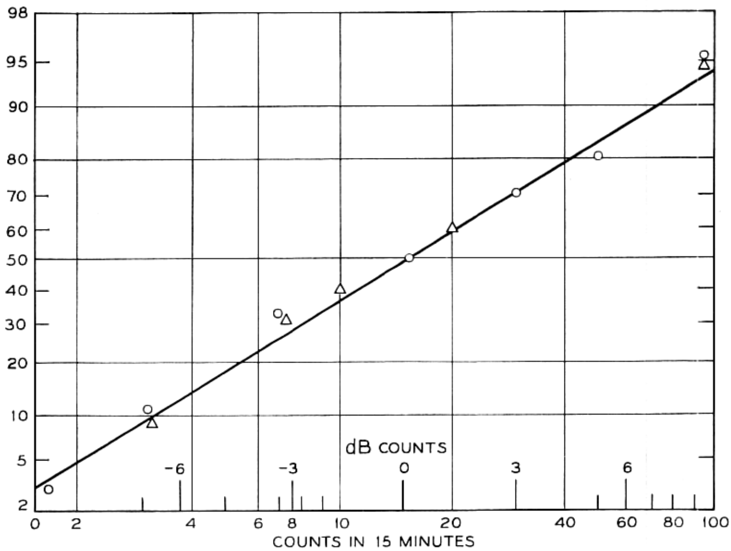

Fig. 13 — Empirical verification of the distribution of counts as derived from distribution of levels of constant counts. Distribution of counts in 15-minute measured periods on one class of trunks. ○ as measured; △ constructed from level distribution; — predicted from equation (5).

The rth moment of C is then found to be

$$E[C^r] = C_o^r \exp [r^2/(4a)]; \qquad a = \frac{M^2m^2}{2\sigma_\ell^2} \approx \frac{9.5}{\sigma_D^2}, \tag{7}$$

and the variance, $\sigma_C^2 = C_o^2(e^{1/a} - e^{1/2a}) = C_o^2A$. Now, as the measurement interval is increased as above, C_o is replaced by nC_o and $\sigma_C^2(nC_o) = n^2C_o^2A$. The variance of the count distribution increases as the square of time if the mean increases linearly.

Note from equation (7), that the mean of the count distribution is not equal to the reference count C_o which is associated with the level distribution. The two are related as*

$$\langle C \rangle_{av} = C_o \exp [\sigma_\ell^2/(2m^2M^2)]$$
$$\approx C_o(1.027)^{\sigma_D^2}. \tag{8}$$

Thus, the mean of the count distribution at level ℓ_o is always greater than the reference count C_o. Furthermore, from the definitions of the

* Note from Fig. 12 that σ_D may be as large as 13.5 so $\langle C \rangle_{av}$ may be as large as 130 times C_o.

level distribution and the quantity D, C_o is equal to the median of the count distribution. Solving equation (8) for σ_D yields

$$\sigma_D \approx 8.7(\log \langle C\rangle_{av}/C_o)^{\frac{1}{2}},$$

and an estimate of the variance of the log-count distribution may be made from the mean and median of the count distribution. This relation should be very useful in practice.

Now consider measurements of length K_2T taken on a number of channels with the counts recorded after K_1T and K_2T, $K_2 > K_1$. Let x be a random variable that takes the value of the count at K_1T, and y one that takes the value of the count at K_2T. If it were true that the count on each channel is a linear function of time, then for the ith channel measurement, $y_i = (K_2/K_1)x_i$ and the coefficient of correlation $\rho_{xy} = 1$. Such correlation coefficients were calculated for several sets of data. The results are presented in Table III. T was equal to 5 minutes in all cases. The notation ρ_{ij} indicates the correlation between the counts at the end of i 5-minute intervals with that after j 5-minute intervals. The mean ratio of the count after j intervals to the count after i intervals and the ratio of the variance after j and i intervals is also given. The expected values, derived from the model, are given in each case (in parentheses), as well as the observed values. While the correlation coefficients are not all as close to unity as one might hope, especially for the 5-minute versus 30-minute measurements ($i = 1, j = 6$), the mean and variance do appear to increase directly and as the square of time respectively.

On the basis of the data shown in Table III and Fig. 13, the model appears to be an adequate description of the observed behavior of the impulse noise on transmission facilities as viewed through impulse noise measuring sets.

TABLE III—CORRELATION COEFFICIENTS AND RATIOS OF MEANS AND VARIANCES*

i, j	ρ_{ij}	μ_j/μ_i	s_j^2/s_i^2	Sample Size
1,2	(1) 0.87	(2) 2.04	(4) 3.70	87
1,2	(1) 0.92	(2) 1.97	(4) 5.00	76
1,2	(1) 0.90	(2) 1.98	(4) 3.98	216
1,3	(1) 0.96	(3) 3.10	(9) 9.90	161
1,3	(1) 0.98	(3) 2.90	(9) 8.50	168
2,4	(1) 0.97	(2) 2.06	(4) 3.60	93
1,6	(1) 0.58	(6) 6.76	(36) 46	161

* For counts observed after i and j 5-minute intervals. Predicted values in parentheses are followed by observed values.

VII. SUMMARY

The following relations and conclusions come from the model presented and the data upon which it is based.

(*i*) Level distributions are normal with mean ℓ_o and variance σ_ℓ^2.

(*ii*) Count distributions are log-normal with mean $\langle C\rangle_{av}$ which is linearly related to the length of the measurement interval, and variance, σ_C^2, which is proportional to the square of the interval. Equivalently, log-count distributions are normal with mean proportional to the logarithm of the measurement interval and variance, σ_D^2, independent of interval.

(*iii*) σ_ℓ is dependent upon the class of trunk but is independent of C_o, an arbitrary reference count greater than zero.

(*iv*) $\sigma_D = 10\sigma_\ell/m$, m is a measure of the slope of the distribution of noise peak amplitudes.

(*v*) $\sigma_D \approx 8.7(\log_{10} \langle C\rangle_{av}/C_0)^{\frac{1}{2}}$.

(*vi*) $\langle C\rangle_{av} \approx C_o(1.027)^{\sigma_D^2}$

(*vii*) The mean of the count distribution, $\langle C\rangle_{av} = C_o e^{1/(4a)}$ and the variance

$$V(C) = C_o^2 e^{1/(2a)}[e^{1/(2a)} - 1]; \qquad a = \frac{m^2M^2}{2\sigma_\ell^2}; \qquad \frac{1}{M} = \log_e 10.$$

(*viii*) The median of a count distribution, taken at level ℓ_o, is equal to C_o and the mean $\langle C\rangle_{av}$, may be 100 times C_o. Expected count by itself is accordingly a very poor statistic for describing impulse noise. However, the mean and the median completely describe the count or log-count distributions.

The model helps to explain the apparent erratic behavior of impulse noise measurements. Any measurement is a sample taken from the bivariate sample space illustrated in Fig. 11. The fact that the distribution of counts is log-normal also accounts for the great fluctuation in the count observed on successive measurements on a given channel. It is shown however that the average rate of occurrence is reasonably constant with time for intervals from 5 to 30 minutes.

VIII. ACKNOWLEDGMENTS

Numerous people have contributed to the collection and analysis of the data that has resulted in the preparation of this paper. Special accord is to be given to Messrs. D. A. Lewinski and I. Nasell, whose efforts of several years ago gave rise to the system-wide statistically sound surveys which provided the fundamental data; and to the many people from Bell Telephone Laboratories, the American Telephone and Telegraph Company, and associated companies who participated in those surveys. My appreciation of the efforts of Messrs. D. P. Dumas, C. R. Ellison, and B. F. Hellrigel must also be noted for their significant contributions in the data processing, reduction, and analysis. I wish to express my thanks for the many stimulating discussions with N. A. Marlow during the course of this work and for a critical review of the first draft.

REFERENCES

1. Aikens, A. J., and Lewinski, D. A., "Evaluation of Message Circuit Noise," B.S.T.J., *39*, No. 4 (July 1960), pp. 911–931.
2. Favin, D. L., "The 6A Impulse Noise Counter," Bell Laboratories Record, *41*, No. 3 (March 1963), pp. 100–102.
3. Mertz, P., "Impulse Noise and Error Performance in Data Transmission," AD 614 416, U. S. Department of Commerce, Clearing House for Scientific and Technical Information, April 1965.
4. Fennick, J. H., and Nasell, I. E. O., "The 1963 Survey of Impulse Noise on Bell System Carrier Facilities," 1966 IEEE Int. Conv. Record, Part I, Wire and Data Communication, March 1966, pp. 149–156.
5. Fennick, J. H., "A Method for the Evaluation of Data Systems Subject to Large Noise Pulses," 1965 IEEE Int. Conv. Record, Part I, Communications 1, Wire and Data Communication, March 1965, pp. 106–110.
6. Ellison, C. R., Holmstrom, J. R., and Nasell, I. E. O., "The Transmission Performance of Bell System Intertoll Trunks," B.S.T.J. *47*, No. 8 (October 1968), pp. 1561–1614.
7. Berger, J. M., and Mandelbrot, B., "A New Model for Error Clustering in Telephone Circuits," IBM J. Res. and Development, *7*, (July 1963), pp. 222–236.
8. Frichtman, B. D., "A Binary Channel Characterization Using Partitioned Markov Chains," IEEE Trans. Inform. Theory, *IT-13*, No. 2 (April 1967), pp. 221–227.
9. Kaenel, R. A., and others, "Burst Measurements in the Time Domain," Trans. Audio and Electroacoustics, *AU-14*, No. 3 (September 1966), pp. 115–121.