#4

A STUDY OF SPEECH PROBABILITY DISTRIBUTIONS

W. B. DAVENPORT, JR.

LOAN COPY

TECHNICAL REPORT NO. 148

AUGUST 25, 1950

RESEARCH LABORATORY OF ELECTRONICS
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
CAMBRIDGE, MASSACHUSETTS

The research reported in this document was made possible through support extended the Massachusetts Institute of Technology, Research Laboratory of Electronics, jointly by the Army Signal Corps, the Navy Department (Office of Naval Research) and the Air Force (Air Materiel Command), under Signal Corps Contract No. W36-039-sc-32037, Project No. 102B; Department of the Army Project No. 3-99-10-022.

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

RESEARCH LABORATORY OF ELECTRONICS

Technical Report No. 148 August 25, 1950

A STUDY OF SPEECH PROBABILITY DISTRIBUTIONS

W. B. Davenport, Jr.

This report is substantially the same as a Doctoral thesis
in the Department of Electrical Engineering, M.I.T.

Abstract

The problem of measuring a probability concerning a random time function is shown to reduce generally to that of measuring a time average. A study is made of the error due to averaging over only a finite interval of time. A description is given of apparatus, which, with pulse techniques, can measure certain probabilities concerning the instantaneous amplitude and the zero-crossing periods of any time function whose statistics lie in the range of the voice wave statistics.

Measurements of the first probability distribution density of the speech-wave instantaneous amplitude for several speakers indicate that this distribution varies exponentially for large values of amplitude. Small but definite increases in conditional probability were noted for delay times of the order of magnitude of the pitch period. These measurements were of the stationary, or "long time", speech statistics. Most of the measurements were made on substantially undistorted speech; some were made to show qualitatively the effect of reflections.

A STUDY OF SPEECH PROBABILITY DISTRIBUTIONS

Introduction

Statistics and communication

In recent years, the methods and concepts of mathematical statistics have been applied to the study of the basic principles of communication. This method of attack is due primarily to Wiener (1, 2) and Shannon (3, 4). An application of statistical methods to the study of communication systems enables the improvement of such systems through bandwidth reduction, decreased average power requirements, better noise suppression, and similar factors. In general, for a full achievement of the possible improvements, systems radically different from present systems will have to be developed, incorporating such features as message storage and coding. Such drastic steps need not be taken in order to achieve some improvements in existing systems. For example, under certain conditions, the average power requirements of present day pulse code modulation systems may be reduced by basing the choice of the quantization levels and pulse codes on certain of the probability distributions of the ensemble of messages to be transmitted.

The statistical theory of communication assumes that the messages to be transmitted are stochastic (5, 6) phenomena; that is, all messages are random functions of time; and all a priori knowledge about any particular message is to be found only in the statistical properties of the set of messages from which that particular message was drawn. Thus, one of the basic postulates of this theory is that neither the designer of a new communication system, nor the receiver in any system, can have an a priori knowledge of the exact structure of any message to be transmitted.

The mere knowledge that all a priori message information is contained in the statistical properties of the message ensemble enables an extensive theoretical study to be made of the general properties of communication systems. As soon as the problem of the design of a specific system arises, a detailed knowledge must be obtained of the statistical properties of the ensemble of messages to be transmitted over that system.

At the present time, there are two types of statistical data that are of particular interest: the correlation functions (2), and the probability distributions (6, 7). The probability distributions are more inclusive than the correlation functions in the sense that the correlation functions may be derived analytically from the probability distributions, but not vice versa (2). In many practical cases, it might be a more economical use of time to determine the correlation functions experimentally and directly than it would be to determine the probability distributions experimentally and then from this data to derive the correlation functions analytically. Thus, on an experimental basis at least, it is worthwhile to make experimental studies of both the correlation functions and the probability distributions. The problem of experimentally obtaining the correlation functions of various random phenomena has been studied by Cheatham (8), Singleton (24), and others. One of the purposes of this thesis is to study the problem of the experimental determination of certain probability distributions.

Purpose of this investigation

There were two main objectives for this study: the design and construction of apparatus which could measure certain probability distributions of physical random processes; and the application of this apparatus to a preliminary study of some particular random process appearing as a message function in some communication system.

Although it is possible to construct theoretical models of a large number of physical random processes and from these models to obtain statistical characteristics through analytic methods, there are many other physical random processes that are not amenable to such an attack. For these latter phenomena it is necessary to determine the statistics experimentally. Also, in the case of phenomena that are amenable to an analytic study, a complete solution requires an experimental verification of the analytic results in order to determine the validity of the model. Thus there is a definite need for apparatus capable of measuring statistics.

The design of apparatus for the experimental determination of statistics may generally proceed in one of two directions. The apparatus might be designed so that its capabilities fit the characteristics of the phenomenon to be studied, or the apparatus might be designed so that its capabilities are the most expansive obtainable with the then available equipment techniques. The former method has the advantage of greater equipment economy; the second has the advantage of greater equipment utility. It was felt that, for the purposes of this investigation, it would be sufficient to design apparatus to fit the characteristics of some specific random process.

In view of the importance of voice communication systems, it was decided to study certain of the probability distributions of the voice. Unfortunately, the physiological and psychological processes involved in the transmission of an idea from the brain of one person to the brain of another person via the voice are by no means completely known to man at the present time. It is thus impossible to state definitely which are the most important voice characteristics but some judicious conjectures may be made.

Under certain conditions it is desirable to transmit the instantaneous amplitude of the voice wave substantially undistorted. For such systems it would be important to study the probability distributions of the instantaneous amplitude.

Studies by Licklider and others (9, 10) have shown that severe amplitude limiting of the voice wave, so as to produce a constant-amplitude rectangular wave whose periods are the zero-crossing periods of the voice wave, produces a comparatively small reduction in word intelligibility. Thus it is of interest to study some of the probability distributions concerning the zero-crossing periods of the voice wave.

It was decided to confine this investigation to a study of the first and second probability distributions and to study only the "long time", or time stationary, statistics of the voice, rather than the "short time", or time variable, statistics. The first probability distribution of a random process provides a measure of the probability that the process has a given value, independent of the time of measurement. The second probability distribution gives a measure of the probability of occurence of two specified values separated in time by a specified interval, and thus provides a measure of the statistical correlation between two points separated in time.

Certain questions now arise. How long a sample of the voice wave must we observe in order to be reasonably certain of obtaining stationary statistics? Which speech characteristics (i.e. vowels, consonants, etc.) produce significant effects in the probability distributions? Is there an appreciable difference between the probability distributions of different speakers? Is there a significant effect on these types of measurement due to the speaker's environment? One of the purposes of this investigation thus becomes the obtaining of at least partial answers to these and similar questions.

Previous work

Studies of the first probability distribution of the voice wave instantaneous amplitude have been made by various investigators. Among them are Sivian (11), Braunmühl (12), Thierbach and Jacoby (13), Dunn and White (14), Jacobsen (15), and Sacerdote (16). In general, their results seem to be subject to considerable doubt in the region of small amplitudes. This point will be discussed in sec. I.

To the writer's knowledge, there has been no prior publication concerning the second probability distributions of the voice wave instantaneous amplitude.

The average density of zero-crossings has been studied by the Northeastern University Visual Message Presentation group (17) but, as far as the writer knows, there has been no prior publication concerning either the first or second probability distributions of the voice wave zero-crossing periods.

I AMPLITUDE DISTRIBUTIONS

A. The First Probability Distribution Density – Theory

We will present the theoretical background for the measurement of the first probability distribution density of the voice wave instantaneous amplitude. In particular, we will define the necessary statistical quantities, develop a method of measurement, and make a statistical study of the measurement errors. The results that will be presented are implicitly contained in the voluminous literature on mathematical statistics. However, because of some misconceptions that have arisen in the past in similar studies, it was felt worthwhile to develop explicitly those results that apply to our problem.

Postulates and definitions

For the sake of concreteness, all of the following discussions will relate to the study of a specific random process: the sound wave associated with the human voice. However, it should be kept in mind that the basic concepts and methods are equally applicable to the study of other random physical processes; only equipment details need be changed.

Our first postulate is that the voice wave is the result of a random, or stochastic (i.e. chance) process. By this we mean in essence, that no matter how much we know of the past history of a voice wave we cannot predict exactly what its future will be.

Our second postulate is that we are dealing with a statistically stationary process; that is, we assume that the statistical parameters concerning the process of voice wave formation do not change with time. Whether we consider the process under study to belong to a single person or to a large segment of the human race, a doubt arises as to the logical justification of this stationarity postulate. For example, the person ages and the race evolves. However, we can presumably confine our studies to some interval of time during which sensibly stationary characteristics might be obtained. Even though the logical foundation of this postulate is somewhat shaky, it is a desirable postulate in that its use considerably simplifies the mathematics involved in our study. It is well to mention that the same logical doubts apply to the postulation of the stationary character of almost any physical process.

With the acceptance of the above postulates, it is possible to define time invariant statistics for the voice wave. One such statistic is the "first" probability $P(x_1 \leqslant x(t_1) < x_2)$, that the random variable, $x(t)$, the instantaneous amplitude of the voice wave, lies in the interval (x_1, x_2) at the time t_1. The adjective "first" expresses the fact that we are concerned with a single instant of time, t_1. Later we will define "second" probabilities concerning the values of $x(t)$ at two distinct times, t_1 and t_2. Since we are dealing with stationary random processes, such a probability is, by definition, independent of the value of t; hence

$$P(x_1 \leqslant x(t_1) < x_2) = P(x_1 \leqslant x(t) < x_2) \tag{1}$$

where t may be any arbitrarily chosen instant of time.

It will frequently be desirable to define a "derived random variable", $\xi(t)$, related to the random variable $x(t)$ and an associated interval (x_1, x_2) by the relations

$$\xi(t) = \begin{cases} 1 \text{ if } x_1 \leqslant x(t) < x_2 \\ 0 \text{ otherwise} \end{cases} \quad . \tag{2}$$

Such a pair of random variables is shown by Fig. 1. Thus we have that

$$P(\xi = 1) = P(x_1 \leqslant x < x_2) \quad . \tag{3}$$

If we take the upper limit of the interval (x_1, x_2) to be

$$x_2 = x_1 + \Delta x_1 \tag{4}$$

we may then define the "first probability distribution density", $W_1(x)$, by the equation

$$W_1(x_1) \equiv \lim_{\Delta x_1 \to 0} \frac{P(x_1 \leqslant x < x_1 + \Delta x_1)}{\Delta x_1} \tag{5}$$

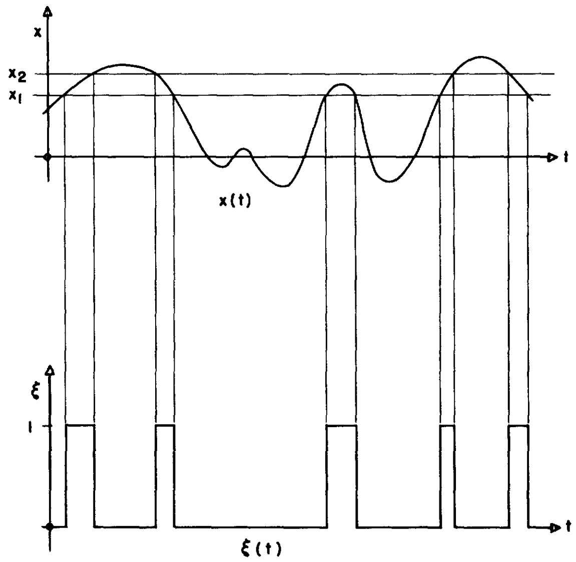


Fig. I The random variable, $x(t)$, and its derived random variable, $\xi(t)$.

or by the equation

$$P(x_1 \leqslant x < x_2) = \int_{x_1}^{x_2} W_1(x)\,dx \quad . \quad (6)$$

As $P(x_1 \leqslant x < x_1 + \Delta x_1)$ may not always approach zero when Δx_1 approaches zero (for example if $x(t)$ assumes only discrete values of x) we must include the possibility that the first probability distribution density, $W_1(x)$, might consist in part, or entirely, of impulse functions (i.e. Dirac delta functions).

Method of measurement

The starting point for the measurement of the first probability distribution density, $W_1(x)$, lies in the definition of $W_1(x)$ as given by Eq. 5. If $W_1(x)$ is a reasonably smoothly varying function of x, we may choose Δx_1 small enough so that

$$W_1(x_1) \doteq \frac{P(x_1 \leqslant x < x_1 + \Delta x_1)}{\Delta x_1} \tag{7}$$

to a high degree of approximation. Thus our problem becomes that of measuring the probability of occurrence of the event $(x_1 \leqslant x < x_1 + \Delta x_1)$.

Let us now define a derived random variable, $\xi(t)$, by the relations

$$\xi(t) = \begin{cases} 1 \text{ if } x_1 \leqslant x(t) < x_1 + \Delta x_1 \\ 0 \text{ otherwise} \quad . \end{cases} \tag{8}$$

The statistical average, $\bar{\xi}$, of this new derived random variable is given by the equation (Cramér, p. 170, ref. 7)

$$\bar{\xi} = \int_{-\infty}^{+\infty} \xi W_1(\xi)\,d\xi \quad . \tag{9}$$

Now ξ may assume only the discrete values zero and one; hence the distribution density, $W_1(\xi)$, consists of two impulses. The integral in Eq. 9 then becomes a summation and the statistical average of the derived random variable is given by

$$\bar{\xi} = \sum_{j=0}^{1} \xi_j\, P(\xi = \xi_j) \quad . \tag{10}$$

Evaluating the summation gives

$$\bar{\xi} = P(\xi = 1) \quad . \tag{11}$$

Then, remembering the defining relations for the derived random function, we see that

$$P(x_1 \leq x < x_1 + \Delta x_1) = P(\xi = 1) = \overline{\xi} \quad . \tag{12}$$

Thus the desired probability, $P(x_1 \leq x < x_1 + \Delta x_1)$, is equal to the statistical average of the derived random variable related to the original random function, $x(t)$, and the interval $(x_1, x_1 + \Delta x_1)$.

The voice mechanism is capable of generating an infinitude of different voice waves, $x(t)$. To each of these different waves, there exists a corresponding derived random variable, $\xi(t)$. The set of all of the different forms of $\xi(t)$ is called the ensemble of the random functions $\xi(t)$. By definition, the statistical average of ξ involves, in essence, the simultaneous measurement of the values of all of the member functions of the ensemble at a time t. Operationally, it is usually simpler to measure the time average of a single member function of the infinite ensemble than it is to measure the statistical average of the entire ensemble. A question then arises, What is the relation between the time average and the desired statistical average?

The time average of $\xi(t)$ over the finite interval (O, T) is given by

$$< \xi(t) >_T = \frac{1}{T} \int_O^T \xi(t)\, dt \quad . \tag{13}$$

We are dealing with a stationary random process, hence a shift in time origin is immaterial, and we may define the time average of $\xi(t)$ over an infinite interval by

$$\begin{aligned} < \xi(t) >_{av} &\equiv \lim_{T\to\infty} \frac{1}{T} \int_O^T \xi(t)\, dt \\ &= \lim_{T\to\infty} < \xi(t) >_T \quad . \end{aligned} \tag{14}$$

The ergodic theorem (Wiener, sec. 0.8 ref. 2) then tells us that

$$< \xi(t) >_{av} = \overline{\xi} \quad . \tag{15}$$

That is, for a stationary random process, the ergodic theorem says that (except for a subset of functions whose total probability of occurrence is zero) the time average over an infinite interval of a single member of an ensemble of random functions is equal to the statistical average taken over the entire ensemble. Thus $P(x_1 \leq x < x_1 + \Delta x_1)$ may be measured by taking the time average of the derived random variable

$$P(x_1 \leq x < x_1 + \Delta x_1) = < \xi(t) >_{av} \tag{16}$$

or in terms of the time average over the finite interval

$$P(x_1 \leq x < x_1 + \Delta x_1) = \lim_{T\to\infty} < \xi(t) >_T \quad . \tag{17}$$

Experimentally then, we may measure the time average of $\xi(t)$ over a finite interval (O, T). As the length of the interval is allowed to increase, the average should approach a limiting value. We may then take this limiting value of the average as the desired probability.

The time average of $\xi(t)$ over a finite time interval may be experimentally determined either by a process of continuous time integration or by a process of sampling and summation. If we were concerned only with the determination of $W_1(x)$, the continuous integration method would be simpler. However, the choice of a sampling and summation method considerably simplifies the equipment required for the studies of the conditional probability distributions and the zero-crossing period distributions discussed in succeeding sections.

Let us now consider the problem of determining $< \xi(t) >_T$ through sampling procedures. Suppose that at $(n-1)$ arbitrarily chosen instants of time within the interval (O, T), and at the time T, we measure the value of $\xi(t)$. Then defining Δt_k as the time interval between the $(k-1)$ st and the kth sample, the time average of $\xi(t)$ over the finite period (O, T) is given by

$$< \xi(t) >_T = \lim_{\Delta t \to 0} \frac{1}{T} \sum_{k=1} \xi_k \Delta t_k \tag{18}$$

where ξ_k is the value of $\xi(t)$ at the kth instant of sampling, where Δt is the largest of the values of Δt_k, and where $n \to \infty$ as $\Delta t \to 0$. The integral is thus defined in the usual Riemann sense. A knowledge of the values of ξ at the instants of sampling and of the time intervals between samples thus gives $< \xi(t) >_T$ through a limiting process.

If we take the simplest case of time periodic sampling, the values of Δt_k all become equal to the sampling period, τ_s, and are given by

$$\Delta t_k = \tau_s = \frac{T}{n} \quad . \tag{19}$$

Thus in the case of periodic sampling we have

$$< \xi(t) >_T = \lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} \xi_k \quad . \tag{20}$$

Let us now define a quantity ν as the number of occurrences of the event $(\xi = 1)$ in the measurement of n samples. We then have that

$$\sum_{k=1}^{n} \xi_k = \nu . 1 + (n - \nu) . 0 \quad . \tag{21}$$

Thus Eq. 20 becomes

$$< \xi(t) >_T = \lim_{n \to \infty} \frac{\nu}{n} \tag{22}$$

where ν/n is thus the relative frequency of occurrence of the event ($\xi = 1$). Then from Eqs. 17 and 22 we finally see that

$$P(x_1 \leq x < x_1 + \Delta x_1) = \lim_{T \to \infty} \lim_{n \to \infty} \frac{\nu}{n} \quad . \tag{23}$$

Thus our desired probability is equal to the limiting value of the relative frequency of occurrence of the event ($\xi = 1$), and hence of ($x_1 \leq x < x_1 + \Delta x_1$), as both the number of samples and the time interval over which the samples are taken increase without limit. The errors involved in not passing to the limit are discussed under Statistical Accuracy. The result expressed by Eq. 23 is a seemingly obvious result in that it conforms with our intuitive conception of a probability. However, a review of the derivation of this result would show that such a simple result would not generally be obtained. In fact, Eq. 23 is obtained only if the times of sampling are chosen at random, or are uniformly spaced.

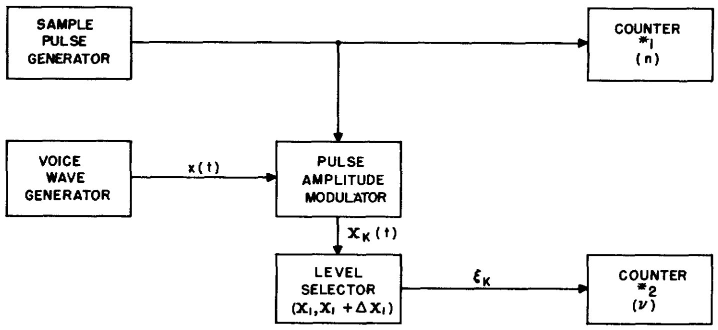


Fig. II Measurement of first probability distribution density, $W_1(x)$.

A block diagram of the equipment used to measure the ratio ν/n is given in Fig. II. The detailed circuits of these units are discussed later. The output of the sampling pulse generator consists of a repetitive chain of constant amplitude and constant duration pulses. These pulses are applied to a pulse amplitude modulator simultaneously with the voice wave, x(t). The output, X(t), of the modulator consists of a chain of pulses whose duration and repetition period are determined by the sampling pulses, and whose amplitudes are linearly related to the amplitudes of the voice wave at the times of sampling. These varying amplitude pulses are then applied to a level selector. The level selector generates an output pulse only when the amplitude of the input pulse lies within a predetermined amplitude interval ($X_1 \leq X(t) < X_1 + \Delta X_1$), and hence only when the voice wave (at the times of sampling) lies within a predetermined interval ($x_1 \leq x(t) < x_1 + \Delta x_1$). The output of the level selector then consists of the samples, ξ_k, of the derived random variable, $\xi(t)$.

To measure the relative frequency of occurrence of the event ($x_1 \leq x(t) < x_1 + \Delta x_1$), the sampling pulse generator is switched on for a time interval T. The number of sample pulses, n, generated in the interval is determined by counter number one. The number of level selector output pulses, ν, is determined by counter number two. The ratio of the readings of these two counters is then ν/n.

In order to measure a complete distribution density, $W_1(x)$, the expected range of variation of the amplitude-modulated pulses, X_k, is divided into approximately fifty to one hundred adjacent amplitude intervals. A measurement is then made of the value of the relative frequency, ν/n, for each of these intervals. If the resulting data indicate that ν/n does not vary too rapidly from one amplitude interval to the next, a close approximation to $W_1(x)$ has been determined. If it is felt that the relative frequency does vary too rapidly from one amplitude interval to the next, the width, ΔX, of the amplitude interval is reduced and data are retaken until satisfactory results are obtained.

Statistical accuracy

Let us now investigate the error to be expected as the result of not passing to the infinite limits required by Eqs. 17 and 23. First, we will derive an expression for the upper bound of the probability that the magnitude of the error (i.e. of the difference between the measured relative frequency and the true probability) exceeds a specified constant. We will find that this upper bound is a function of the variance, or mean square fluctuation, of the relative frequency. An expression for this variance will then be derived. Finally, we will evaluate this expression for several simple cases.

The value, ξ_k, of the kth sample of the derived random variable, $\xi(t)$, is obviously itself a random variable. The sum, ν, of these samples then is also a random variable, as is the relative frequency of occurrence, ν/n, of the event $(x_1 \leqslant x < x_1 + \Delta x_1)$.

The statistical average, or mean, of the relative frequency is given by

$$\overline{\left(\frac{\nu}{n}\right)} = \overline{\left(\frac{1}{n}\sum_{k=1}^{n} \xi_k\right)} = \frac{1}{n}\sum_{k=1}^{n} \overline{\xi}_k \quad . \tag{24}$$

(In this paper, the word "mean" will always signify "statistical average" and never "time average".) The operations of summation and statistical averaging may be interchanged as the mean of a sum of random variables is equal to the sum of the means (Cramér p. 173, ref. 7). The statistical average of the kth sample of the derived random variable is, by definition, the statistical average of the random variable itself. Thus

$$\overline{\xi}_k = \overline{\xi} = P(\xi = 1) \quad . \tag{25}$$

We then see that the mean of the relative frequency is thus the mean of the derived random variable

$$\overline{\left(\frac{\nu}{n}\right)} = \overline{\xi} \quad . \tag{26}$$

Direct application of the Bienaymé-Tchebycheff inequality (Cramér, art. 15.7, ref. 7) then shows that

$$P\left[\epsilon \leqslant \left| \frac{\nu}{n} - P(x_1 \leqslant x < x_1 + \Delta x_1) \right| \right] \leqslant \frac{\sigma^2\left(\frac{\nu}{n}\right)}{\epsilon^2} \quad . \tag{27}$$

That is, the probability that an arbitrary positive constant, ϵ, is equal to or less than the magnitude of the difference between the relative frequency of occurrence of the event $(x_1 \leqslant x < x_1 + \Delta x_1)$ and the probability of occurrence of that event is equal to, or less than, the variance of the relative frequency divided by the square of the constant. The Bienaymé-Tchebycheff inequality places an upper limit on the probability of the difference magnitude exceeding an arbitrary constant, and thus might be expected to provide a pessimistic estimation of the sampling accuracy obtained.

Let us now consider the evaluation of the mean square fluctuation, or variance of the relative frequency. This quantity is important not only because of its appearance in Inequality 27, but also because of its significance as the mean square value of the measurement error. Direct expansion of the defining equation for variance shows that

$$\sigma^2\left(\frac{\nu}{n}\right) = \overline{\left(\frac{\nu}{n}\right)^2} - \bar{\xi}^2 \quad . \tag{28}$$

As shown by Costas (25), the mean square of the relative frequency may be expressed as

$$\sigma^2\left(\frac{\nu}{n}\right) = \frac{\overline{\xi^2}}{n} - \bar{\xi}^2 + \frac{2}{n^2} \sum_{k=1}^{n-1} (n-k)\, \phi_\xi\left(\frac{kT}{n}\right) \tag{29}$$

where $\phi_\xi(\tau)$ is the autocorrelation function of the derived random variable and is defined by

$$\phi_\xi(\tau) \equiv \langle \xi(t)\, \xi(t+\tau) \rangle_{av} = \overline{\xi(t)\, \xi(t+\tau)} \tag{30}$$

as we are dealing with stationary random processes. The limit of the autocorrelation function as τ approaches zero is

$$\phi_\xi(0) = \overline{\xi^2} = \sigma^2(\xi) - \bar{\xi}^2 \tag{31}$$

and the limit as τ approaches infinity is

$$\phi_\xi(\infty) = \bar{\xi}^2 \tag{32}$$

assuming no periodicities in the derived random variable.

Let us now define the alternating component of the derived random variable by

$$\xi_A(t) \equiv \xi(t) - \bar{\xi} \quad . \tag{33}$$

Direct expansion of the autocorrelation function of this alternating component shows that

$$_A\phi_\xi(\tau) = \phi_\xi(\tau) - \bar{\xi}^2 = \phi_\xi(\tau) - \phi_\xi(\infty) \quad . \tag{34}$$

A possible form for the autocorrelation function is shown in Fig. III.

Substitution of the autocorrelation of the alternating component in Eq. 29 then gives

$$\sigma^2\left(\frac{\nu}{n}\right) = \frac{\sigma^2(\xi)}{n} + \frac{2}{n^2}\sum_{k=1}^{n-1}(n-k)\ {}_A\phi_\xi\left(\frac{kT}{n}\right) \quad . \tag{35}$$

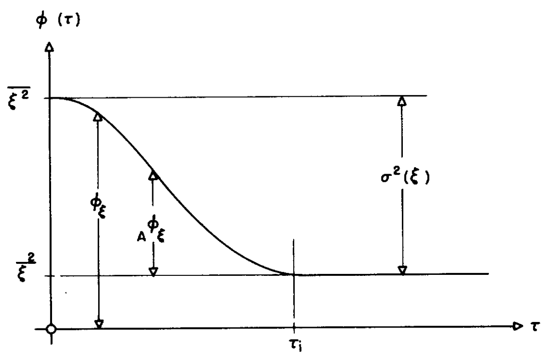


Fig. III Autocorrelation function of the derived random variable.

We have thus obtained a general expression for the mean square expected error involved in the measurement of a probability by the method of taking a finite number of samples. Further information may now be obtained by the consideration of certain specific cases.

If our original speech wave, x(t), is a truly stochastic process, then as we increase τ, the autocorrelation of the alternating component of the derived random variable may be made to approach arbitrarily closely to zero. Let us then define a value τ_i such that (see Fig. III)

$${}_A\phi_\xi\ (\tau) \doteq 0, \text{ if } \tau_i \leqslant \tau \quad . \tag{36}$$

We may then say that any two samples separated in time by more than τ_i are uncorrelated (i.e. they are statistically independent samples). Then if we make our sampling period greater than τ_i, the variance of the relative frequency becomes

$$\sigma^2\left(\frac{\nu}{n}\right) = \frac{\sigma^2(\xi)}{n}, \text{ when } \tau_i < \tau_s = \frac{T}{n} \quad . \tag{37}$$

This is the well-known result concerning independent samples.

It is not always possible, however, nor is it always desirable, to use independent samples. Let us first investigate the case where the number of samples in a measurement period is allowed to increase without limit. We then have that

$$\lim_{n\to\infty} \sigma^2\left(\frac{\nu}{n}\right) = \sigma^2(<\xi(t)>_T) \quad . \tag{38}$$

Allowing $n \to \infty$ in Eq. 35, gives us

$$\sigma^2(<\xi(t)>_T) = \frac{2}{T}\int_0^T \left(1 - \frac{\tau}{T}\right){}_A\phi_\xi(\tau)\, d\tau \quad . \tag{39}$$

Thus, in the case of an infinite number of samples (i.e. in the case of averaging by integration), an error is still probable. For the random processes under discussion, the autocorrelation function of the alternating component of the derived random variable becomes essentially zero for $\tau > \tau_i$. Thus the integral in Eq. 39 remains finite as T approaches infinity and we see that

$$\lim_{T \to \infty} \sigma^2 (< \xi(t) >_T) = 0 \qquad . \tag{40}$$

As an example of this case, let us assume a hypothetical autocorrelation function having the form

$$A\phi_\xi (\tau) = \begin{cases} \sigma^2(\xi) \left(1 - \frac{\tau}{\tau_i}\right) & \text{if } 0 \leq \tau < \tau_i \\ 0 & \text{if } \tau_i \leq \tau \end{cases} \tag{41}$$

which is shown by Fig. IV-A. Substitution in Eq. 39, and evaluation of the integral gives

$$\sigma^2 (< \xi(t) >_T) = \sigma^2 (\xi) \frac{\tau_i}{T} \left(1 - \frac{\tau_i}{3T}\right) \qquad . \tag{42}$$

If we now consider the case where the period of measurement is considerably larger than the independent value of τ(i.e. τ_i), the variance becomes

$$\sigma^2 (< \xi(t) >_T) \doteq \sigma^2(\xi) \frac{\tau_i}{T} \tag{43}$$

and is thus inversely proportioned to the time interval over which measurements are made.

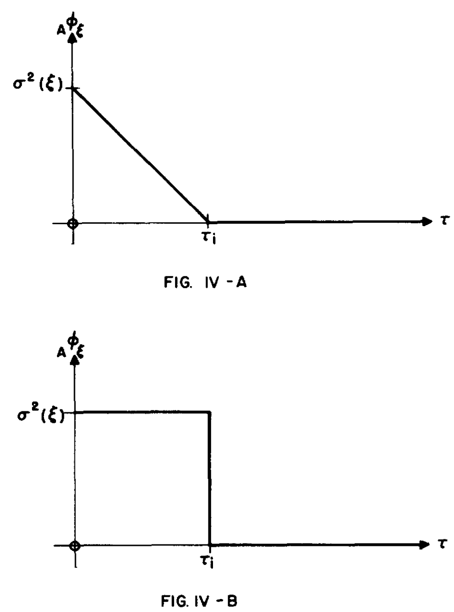


Fig. IV Hypothetical autocorrelation functions.

Of main interest is the case where neither integration nor independent sampling is used. To consider it, we must return to the expression for the variance of the relative frequency given by Eq. 35. Examination of this expression shows that we must know the particular form taken by the autocorrelation function of the derived random variable in order to obtain the variance of the relative frequency for any specific problem. However, an approximate idea of the behavior of the variance may be obtained by arbitrarily assuming some idealized form for the autocorrelation function.

For simplicity, let us assume that

$$A\phi_\xi (\tau) = \begin{cases} \sigma^2 (\xi) & \text{if } 0 \leq \tau < \tau_i \\ 0 & \text{otherwise} \end{cases} \tag{44}$$

which is shown by Fig. IV-B. It should be realized, of course, that this is not a physically realizable autocorrelation function.

If we assume that the period of measurement is very much larger than the independent value of τ, that is, if we assume

$$\tau_i << T \tag{45}$$

we then obtain from Eq. 35

$$\sigma^2\left(\frac{\nu}{n}\right) \doteq \sigma^2(\xi)\left(\frac{\tau_s + 2\tau_i}{T}\right) \tag{46}$$

where τ_s is the sampling period.

In essence, the choice of a different form for the atuocorrelation function has as its main effect the variation of the constant multiplying τ_i in Eq. 46, as may be seen by a comparison of Eqs. 43 and 46. In conclusion, we may say that it is desirable to increase the number of samples per measurement period until the sampling period is a small fraction of the independent value of τ(i.e. of τ_i). Further increases in the number of samples decrease the expected error, but at a diminishing rate.

B. The First Probability Distribution Density – Experimental Results

One of the purposes of this study was a preliminary investigation of certain of the voice wave statistics. In view of the preliminary nature of the investigation, it was decided to study only a restricted class of all of the possible voice waves, "conversational" speech; that is, speech characterized by continuous speaking at a reasonably uniform volume level. This was accomplished by having the speakers read aloud. Two forms of reading material were used: a) the preface by Max Lerner to Adam Smith's "Wealth of Nations," and b) excerpts from the magazine "Time". No detectable differences were noted in the results obtained from these two sources. Most of the measurements were made on one male voice (WD). In order to provide an estimate of the effect of the speaker on this type of measurement, certain of the distributions were repeated for two other speakers, one female voice (KA) and one male voice (JC).

Approximately one working day per distribution was required for the taking of data. Speaking continuously for such a long interval of time rather severely taxes a speaker and thus causes the speech wave to correspond to a definitely non-stationary random process. Accordingly, it was decided to make all measurements on recorded speech. These recordings were made with a wide frequency range, low noise level, magnetic-tape recording system. No significant deterioration of the recorded waves was noticed throughout the numerous playbacks required by the data taking.

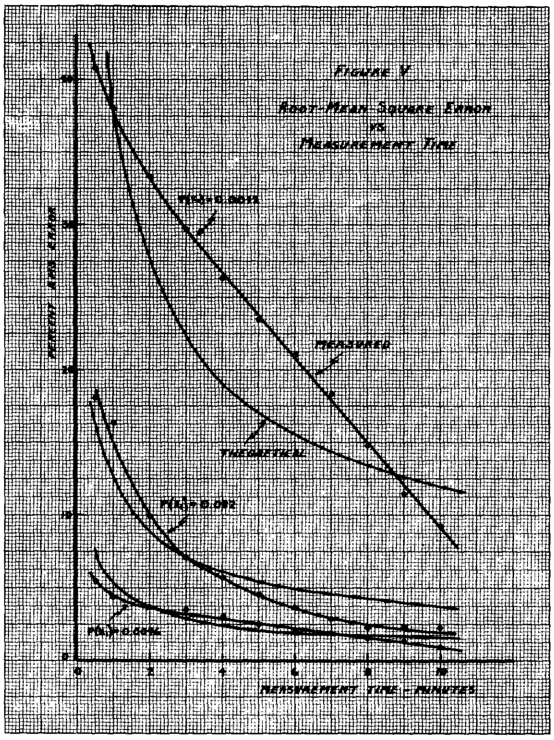
The physical environment of the speaker perturbs the reception of the speech wave chiefly in two ways: through the presence of reflections, and through the presence of external noise. The effect of the noise may be considerably reduced by appropriate acoustical treatment of the room in which the speaker is located. The effect of reflections may either be substantially eliminated by placing the speaker in a practically reflection-free room, or the reflection characteristics of the environment may be completely specified. In view of the difficulties involved in the latter choice, it was decided to make most of the measurements on speech waves recorded in the M.I.T. Acoustics Research Laboratory anechoic chamber. In order to provide an estimate of the effect of reflections upon this type of measurement, certain of the distributions were repeated on speech waves recorded in a rather live studio.

A review of the statistical accuracy study shows that the mean square expected error is inversely proportional to the length of the time interval over which measurements are taken. Thus an increase in probable accuracy is purchased only by the expenditure of measurement time, and a compromise between the two must be reached. From the practical aspect, this is an especially important point if, as was the case in this study, the equipment used permitted the taking of data for only one point at a time on a distribution curve.

In order to apply directly the theoretical error results discussed earlier, one must have an a priori knowledge of the autocorrelation function of the derived random variable corresponding to the measurement being made. Lacking such data, the probable error must be investigated experimentally. Accordingly, measurements were made of the probability that the instantaneous amplitude, $x(t)$, of the speech wave lies on a given amplitude interval $(x_1, x_1 + \Delta x_1)$. Relative frequencies of the event $(x_1 \leqslant x < x_1 + \Delta x_1)$ were obtained for measurement intervals varying from one-half minute to ten minutes with a sampling period of about 12 μsec. Thus it was assured that there would be at least two and one-half million samples per measurement period. These measurements were made for three different values of x_1, and the calculated values of root mean square error vs. measurement time were plotted to form Fig. V. Also plotted in Fig. V are matching curves having the theoretical inverse time variation for the mean square error. A fair fit is obtained.

Subject to the aforementioned conditions and criteria, measurements were made of the first probability density distribution of the speech wave instantaneous amplitude. Three voices, one female (KA) and two male (JC and WD), were studied. One of the male voices (WD) was recorded in both the anechoic chamber and the studio, but the other voices were studied only in the anechoic chamber. Each speaker read the same fifteen minute selection. The reading speeds for all three speakers were the same to within a few percent. The results of these measurements are plotted in Figs. VI-IX. Each distribution is plotted twice: once on linear graph paper to show better the character of the distribution for small values of amplitude, and once on semi-log graph paper to show better the character of the distribution for large values of amplitude. The independent variable in each case is the speech wave amplitude normalized with respect to its root mean square value.

The measurement time per point on these distributions was five minutes, or more, for large values (4 or higher) of the normalized amplitude and also for values of normalized amplitude in the interval $(-1, +1)$. The measurement time for the remaining points was three minutes. Reference to the plots of root mean square expected error



vs. measurement time, Fig. V, shows that the probable error is thus about five percent for the values of the probability density corresponding to amplitudes in the range (−2, +2). For larger values of amplitude, larger errors are to be expected. An estimation of the precision, or repeatability, of the data obtained may be made from the fact that the data for the coarsely spaced points and the data for the finely spaced points in the amplitude interval (−1, +1) were always taken on different days (usually on successive days).

A comparison of the three distributions measured for voices in an anechoic chamber, Figs. VI, VIII, and IX, shows a marked similarity between the forms of the first probability distribution densities for the three different voices. In each case, the distribution has the form of a large spike centered at zero amplitude, superimposed upon an exponential distribution. The heights of the peaks at zero amplitude are all nearly the same (1.62, 1.84, and 1.66) and the semi-logarithmic slopes for large values of the normalized amplitude are all similar (+0.43, −0.56; +0.47, −0.32; and +0.56, −0.56). A comparison of the slopes for large amplitudes shows that a comparatively small asymmetry was observed, and that the three possible types of asymmetry (positive, negative, and zero) were observed. This phenomenon cannot be ascribed to the measuring equipment as the equipment was identical for each set of measurements.

If we examine a photograph of the instantaneous form of the speech amplitude wave, Fig. X, we see that there is some justification for saying that the large values of amplitude correspond to voiced sounds, (e.g. vowels, semi-vowels, etc.), and that the small values of amplitude correspond to unvoiced sounds (e.g. fricatives) and system noise. Thus we might say that the spike portion of the distribution corresponds to unvoiced sounds and noise, while the exponential portion corresponds to the voiced sounds.

Let us now consider the distribution of a speech wave (WD) in a live studio, as shown by Fig. VII. The process governing the formation of the wave received by a listener now consists of the random process forming the original wave and, in addition, the mechanism of reflections. The reflected waves, at the listening point, are attenuated in magnitude and delayed in time as compared with the original wave. Thus, to the listener, the significant portions of the reflected waves are those portions corresponding to the large amplitude portions of the original waves, i.e. to the voiced sounds. Because of the greater lengths of paths travelled by the reflected waves, the reflections will

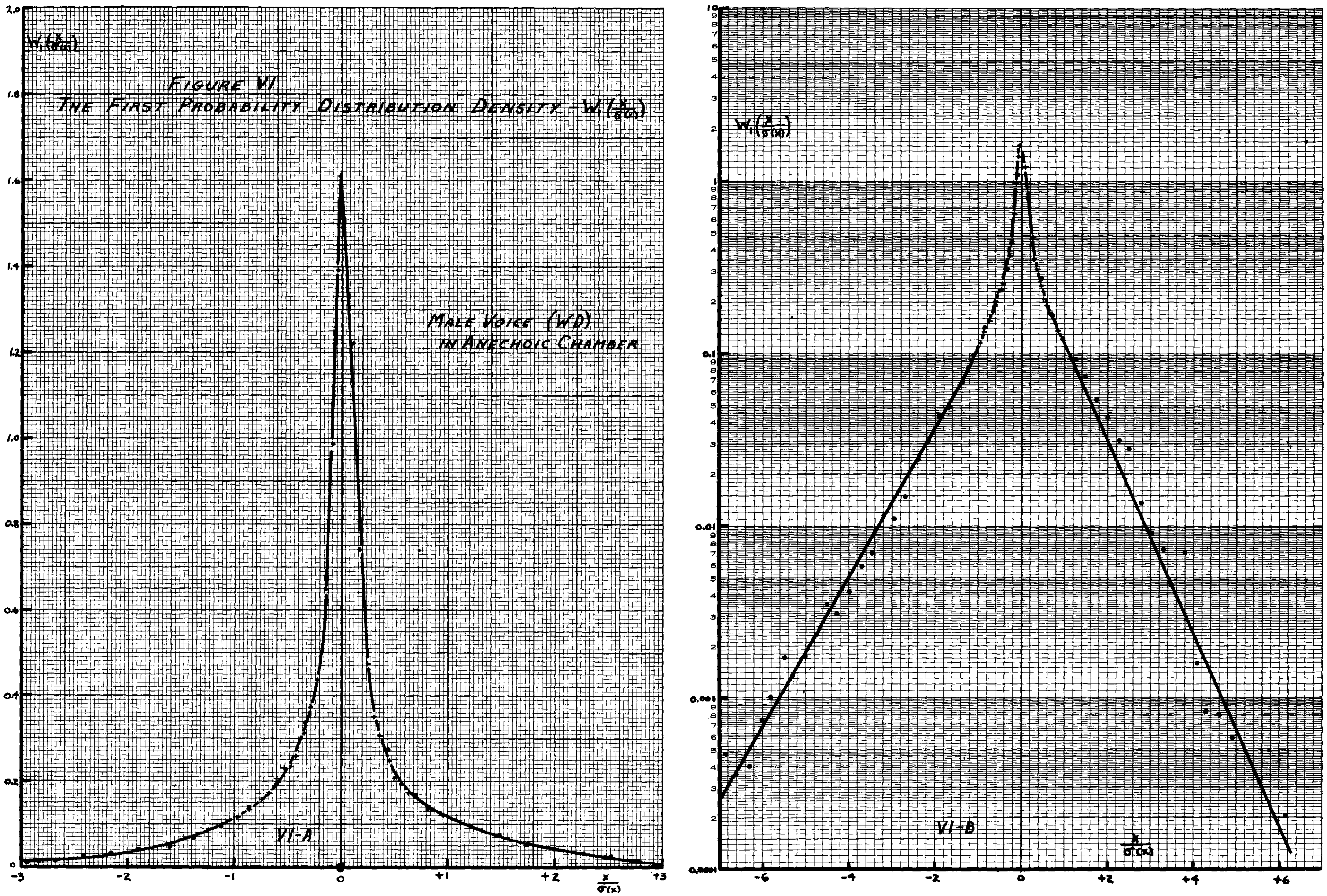


FIGURE VI
THE FIRST PROBABILITY DISTRIBUTION DENSITY - $W_1\left(\frac{x}{\sigma(x)}\right)$
$W_1\left(\frac{x}{\sigma(x)}\right)$
MALE VOICE (WD)
IN ANECHOIC CHAMBER
VI-A
VI-B
$\frac{x}{\sigma(x)}$

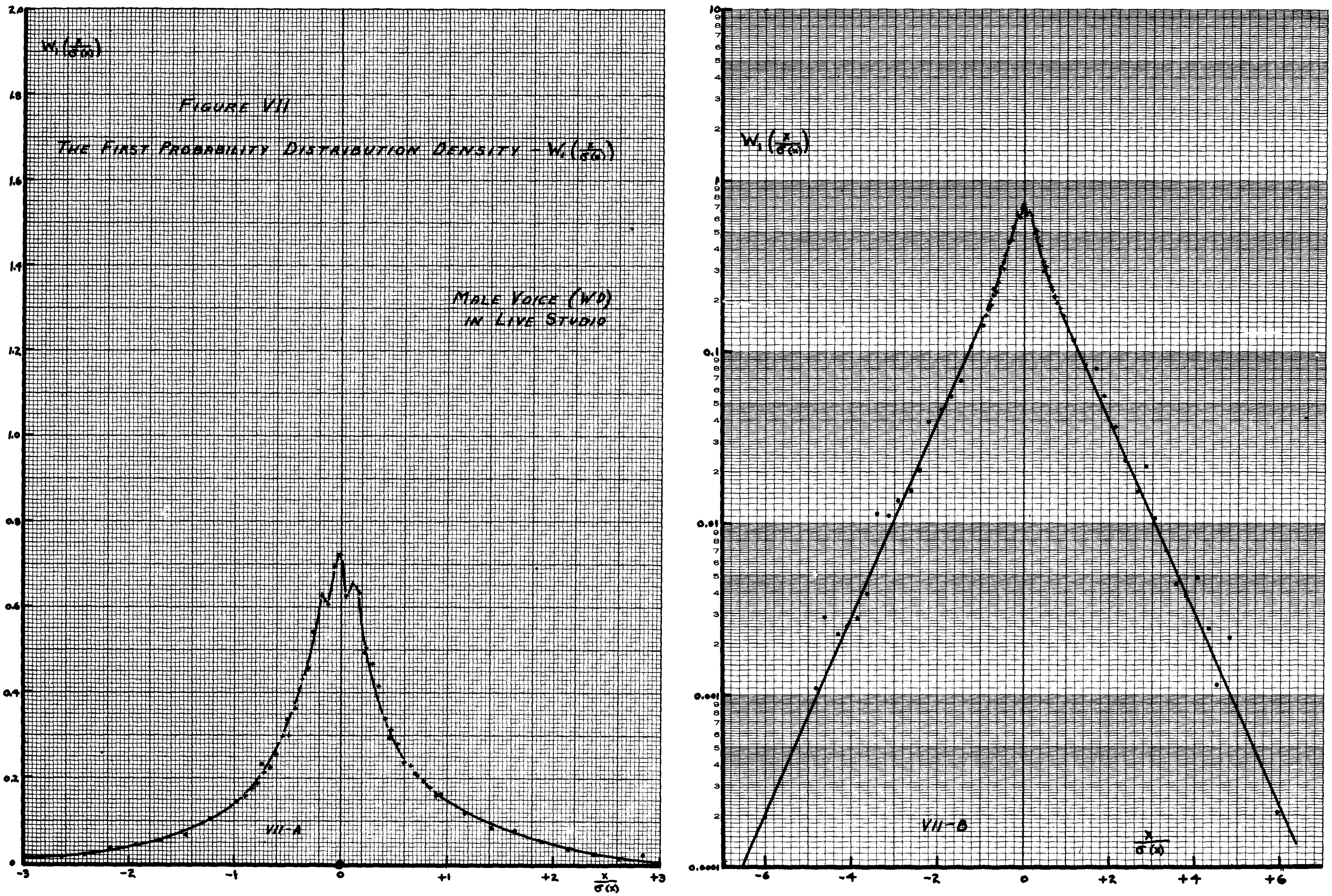


FIGURE VII
THE FIRST PROBABILITY DISTRIBUTION DENSITY – $W_1\left(\frac{x}{\sigma(x)}\right)$
MALE VOICE (WD)
IN LIVE STUDIO
$W_1\left(\frac{x}{\sigma(x)}\right)$
VII-A
VII-B
$\frac{x}{\sigma(x)}$

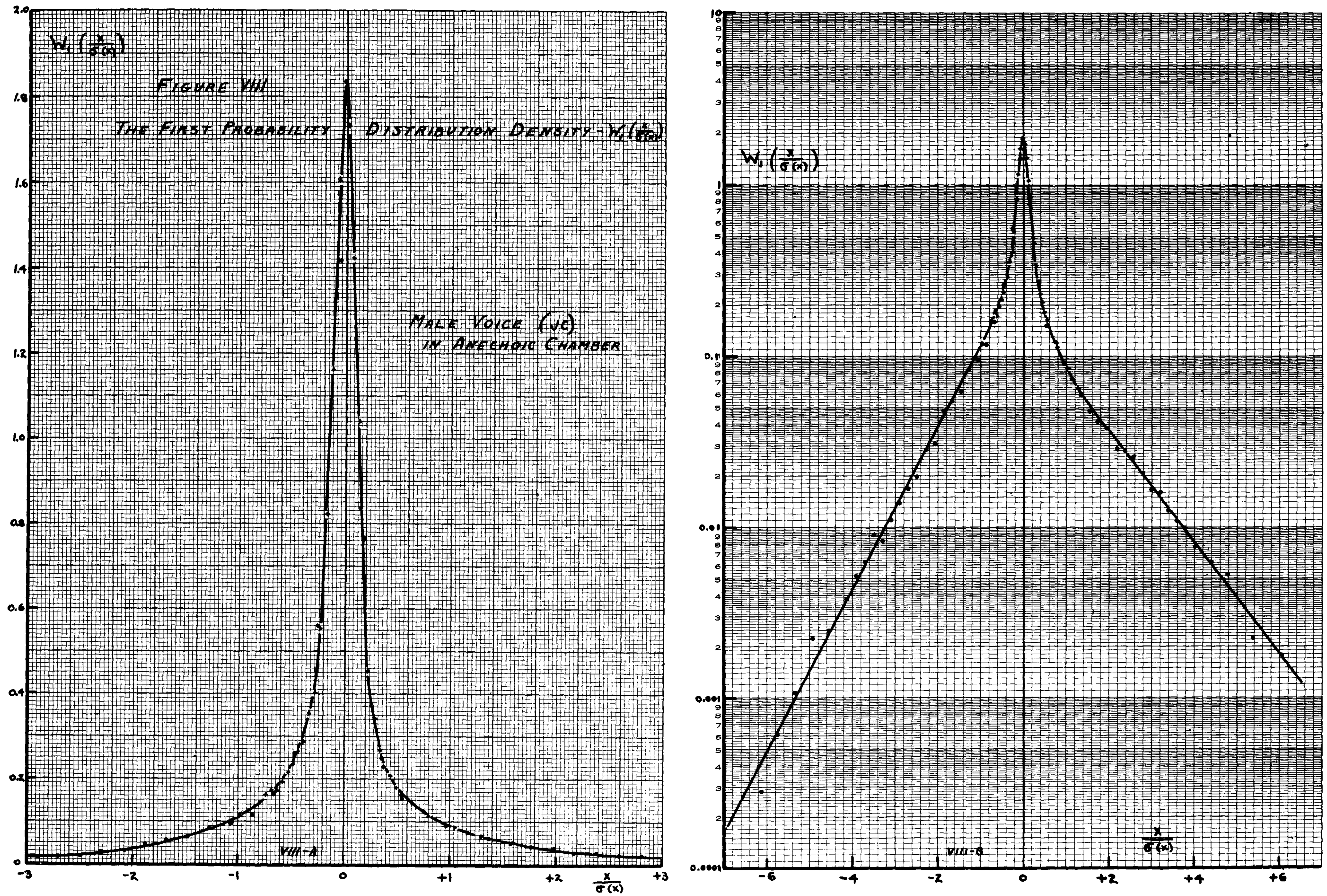
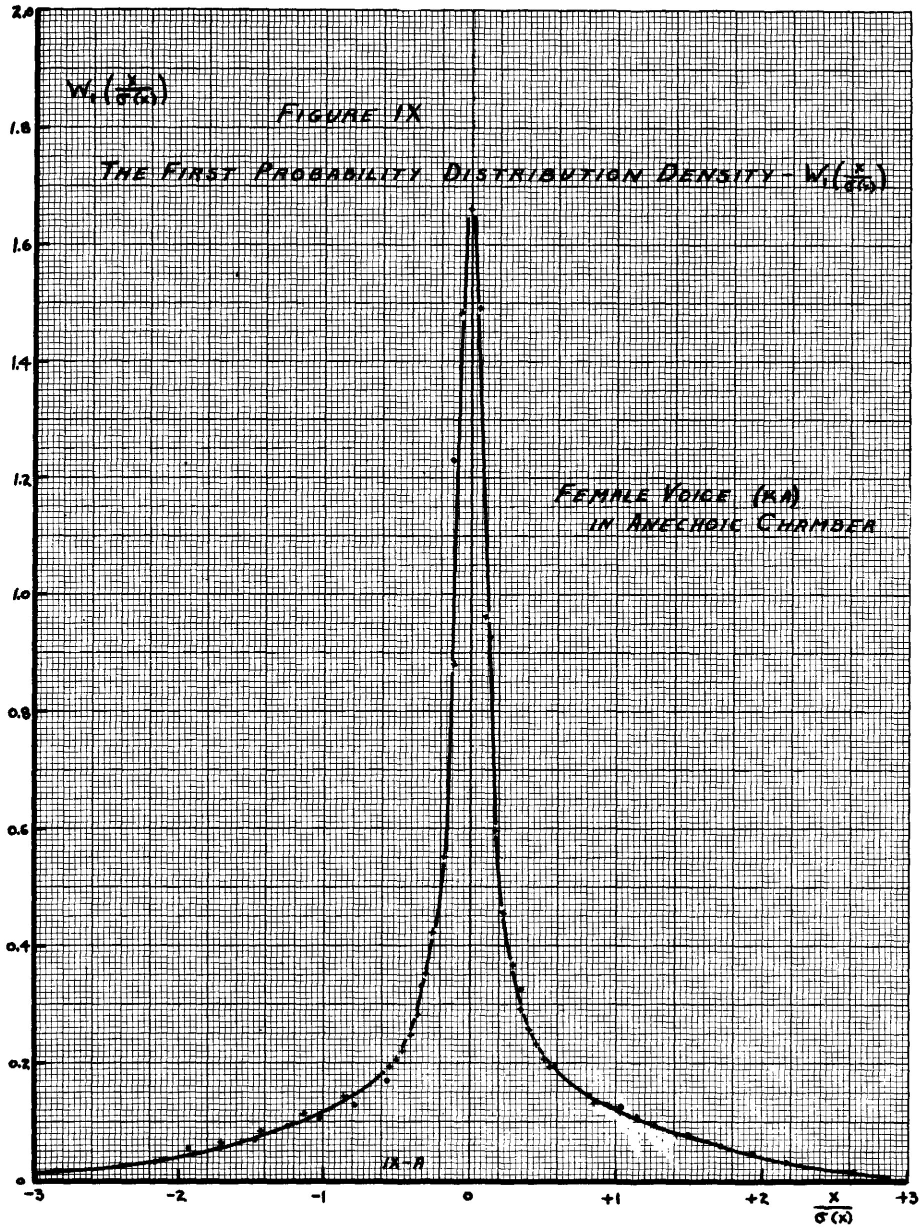
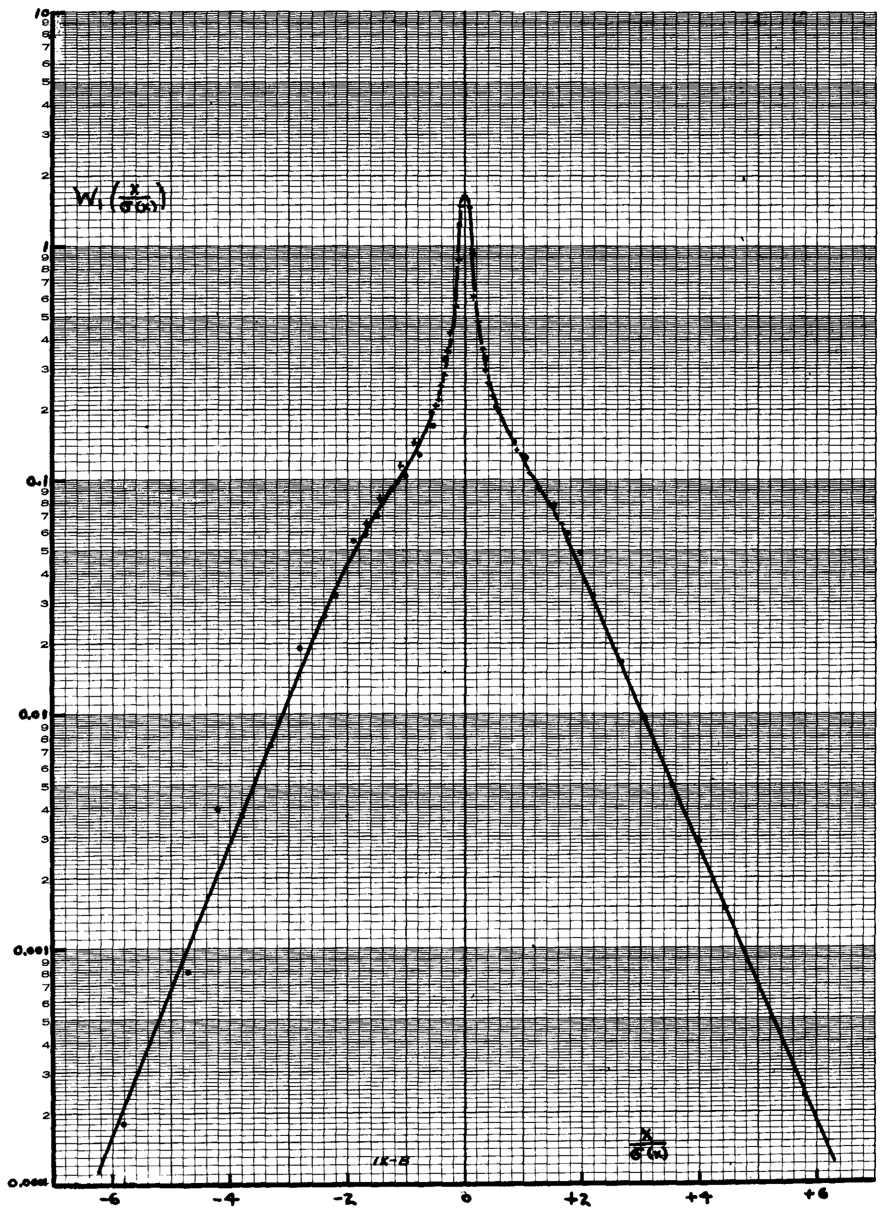


FIGURE VIII
THE FIRST PROBABILITY DISTRIBUTION DENSITY - $W_1\left(\frac{x}{\sigma(x)}\right)$
MALE VOICE (JC)
IN ANECHOIC CHAMBER
$W_1\left(\frac{x}{\sigma(x)}\right)$
$\frac{x}{\sigma(x)}$
VIII-A
VIII-B



$W_1\left(\frac{x}{\sigma(x)}\right)$
FIGURE IX
THE FIRST PROBABILITY DISTRIBUTION DENSITY - $W_1\left(\frac{x}{\sigma(x)}\right)$
FEMALE VOICE (KA)
IN ANECHOIC CHAMBER
IX-A
$\frac{x}{\sigma(x)}$

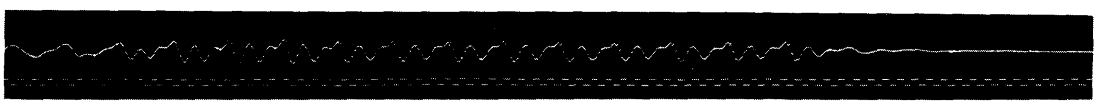


$W_1\left(\frac{x}{\sigma(x)}\right)$
IX-B
$\frac{x}{\sigma(x)}$

cause the received wave to persist after the original wave at the listener has ceased. Thus the reflections should appear mainly as a filling up of the "dead" time of the original wave with attenuated voiced sounds. This phenomenon is shown in Fig. X. Thus we should expect that the effect of reflections would be to decrease the probability occurrence of the very small sounds and to increase the probability of occurrence of the not quite so small sounds. In terms of the distribution, then, we should expect that the height of the spike would be reduced and that its width would be increased. In addition, the voiced sound portion of the distribution should not have its exponential character changed. A comparison of Figs. VI and VII verifies these conclusions. Thus we have seen that a change in environment may produce a significant difference in measured statistics. While it is relatively simple to explain, qualitatively, the differences in the forms of the distributions for the same voice in different environments, it would be exceedingly difficult to calculate the differences theoretically. Thus we feel that studies supposedly concerned with the properties of the speech wave itself should be made mainly in reflection-free environments.

The first probability distribution densities for the three voices studied in the anechoic chamber were seen to be quite similar. It was accordingly felt worthwhile to construct a model distribution having characteristics near the average of the measured distributions. Such a distribution was constructed graphically, and is shown in Fig. XI. A comparison of this distribution with the data of Dunn and White (14) is shown in Fig. XII.

"OF" IN LIVE STUDIO



"OF" IN ANECHOIC CHAMBER

Fig. X Speech wave instantaneous amplitude.
Male voice (WD) and 500 cps timing wave.

While distributions in graphical form are useful, it would be desirable to obtain simple analytic expressions for them. In our discussion of the measured distributions, it was pointed out that they consisted of two main parts: a spike, corresponding to the unvoiced sounds; and noise, superimposed upon an exponential distribution, corresponding to the voiced sounds. Both the unvoiced sounds and the system noise are essentially purely random in character. Accordingly it was decided to make an analytic approximation consisting of a normal distribution for the spike. Such an analytic approximation to the graphical model given by trial and error is

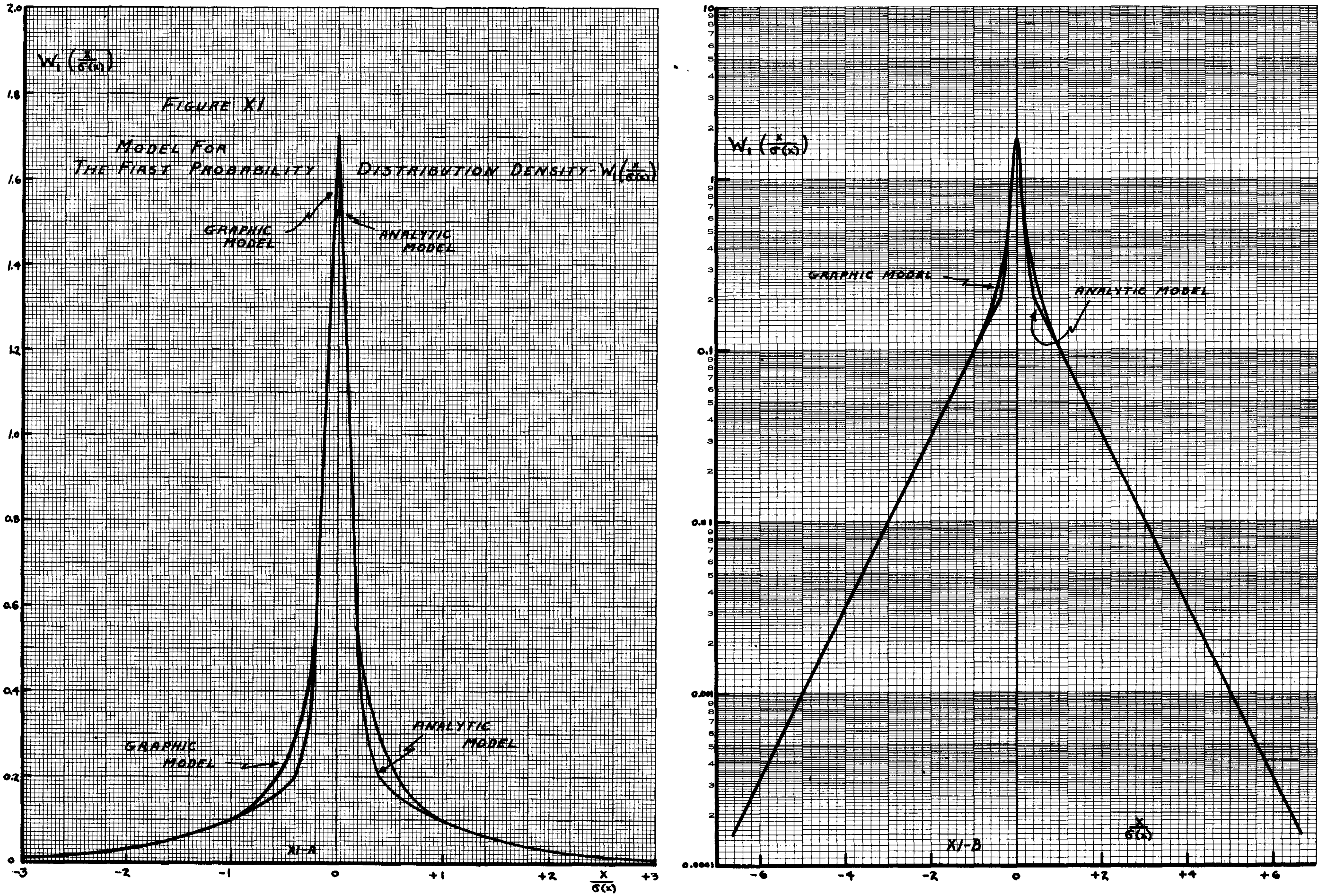


FIGURE XI
MODEL FOR
THE FIRST PROBABILITY DISTRIBUTION DENSITY - $W_1\left(\frac{x}{\sigma(x)}\right)$
$W_1\left(\frac{x}{\sigma(x)}\right)$
GRAPHIC MODEL
ANALYTIC MODEL
XI-A
$\frac{x}{\sigma(x)}$
XI-B

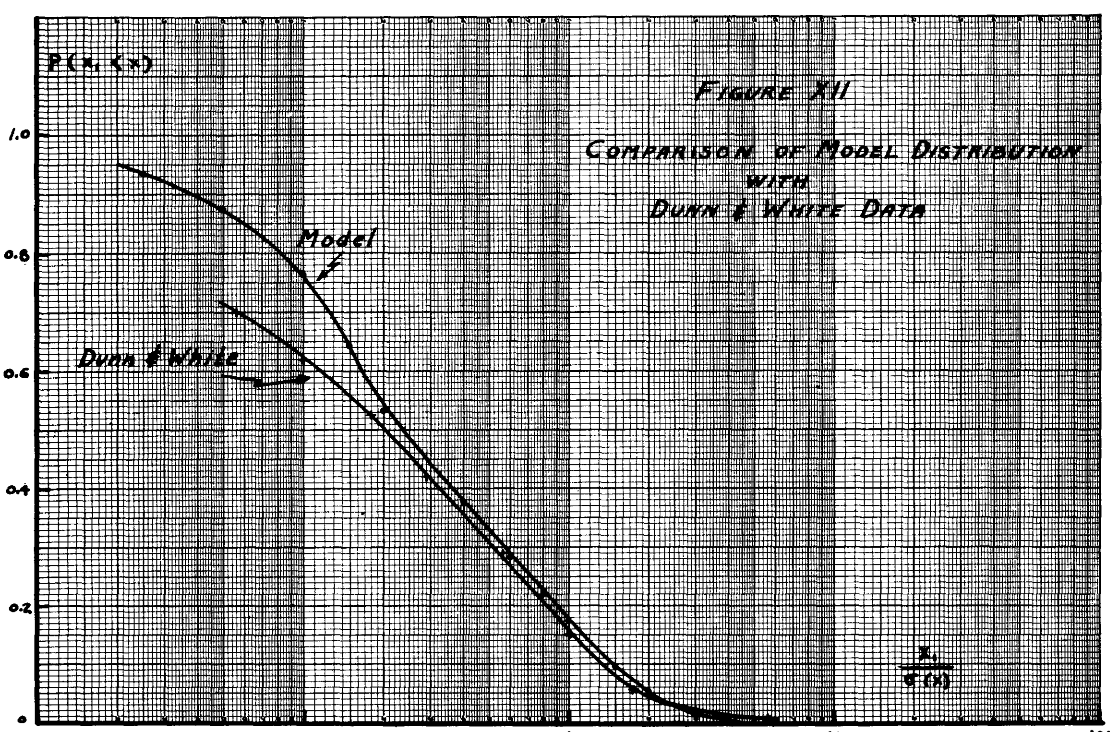
$$W_1(x) = \frac{0.36}{0.118\sqrt{2\pi}} \epsilon^{-\frac{1}{2}\left(\frac{x}{0.118}\right)^2} + 0.316\epsilon^{-1.15|x|} \quad . \tag{47}$$

This distribution is also plotted in Fig. XI. As may be seen, the analytic approximation is quite good except for the juncture region of the two component distributions. The lack of a complete fit may also be seen from the fact that the integral of the analytic approximation over the infinite amplitude interval $(-\infty, +\infty)$ is 0.91, rather than 1.00. It should be noted that the ratio of the root mean square values of the two components is about 10.4, or 20.3 db. This is an appreciably smaller ratio than the measured signal-to-noise ratio of the system under the same operating conditions (about 35 db). Thus the spike cannot be ascribed to the system noise alone, but must be due, in a considerable part, to the characteristics of the speech wave itself.

The exponential character of the first probability distribution density for large amplitudes has been observed by others, notably by Thierbach and Jacoby (13), and by Jacobsen (15). Thierbach and Jacoby obtained the expression

$$W_1(x) = \frac{1}{\sigma(x)} \epsilon^{-\frac{x}{\sigma(x)}} \tag{48}$$

where $\sigma(x)$ is the root mean square value of the voice wave instantaneous amplitude.



Thus Thierbach and Jacoby seem to have forgotten the fact that the speech wave has negative as well as positive amplitudes. Taking into account both negative and positive swings, one obtains, for a purely exponential distribution, the expression

$$W_1(x) = \frac{1}{\sqrt{2}\,\sigma(x)}\,\epsilon^{-\sqrt{2}\left|\frac{x}{\sigma(x)}\right|} \tag{49}$$

which is the expression given by Jacobsen. Both of these results (and in fact most other prior results) seem to have ignored the presence of the spike at the center of the distribution. An examination of the techniques used by previous investigators shows that, in most cases, these techniques lead to rather poor accuracy for the results corresponding to the very small values of the speech wave instantaneous amplitude. However, it should also be stated that in many practical cases the large amplitude region, i.e. the region of exponential variation, is the region of most importance.

C. Conditional Probability Distributions – Theory

We will now study the statistical relation between two points of the speech wave, displaced in time by a given amount. Therefore we will be concerned with the statistics of two random variables, namely the instantaneous amplitudes of the speech wave at two different instants of time.

Definitions

The basic statistics of two random variables are the joint probabilities. For the sake of simplicity, let us first consider the case of discrete random variables. In this case, a random variable $x(t)$ may assume only discrete values of x. The joint probability $P\left[x(t_1) = x_1;\ x(t_2) = x_2\right]$ is defined as the probability of the joint occurrence of the event $\left[x(t_1) = x_1\right]$ and the event $\left[x(t_2) = x_2\right]$. Again we postulate that we have a stationary random process, hence the location of the time origin need not be considered. For a stationary random process, then, the joint probability is given by the equation

$$P\left[x(t_1) = x_1;\ x(t_2) = x_2\right] = P\left[x(t) = x_1;\ x(t+\tau) = x_2\right] \tag{50}$$

where t may have any arbitrary value, and where τ is the time difference $(t_2 - t_1)$.

Let us consider the case of continuous random variables in which the random variable $x(t)$ may assume any value of x. Generally, in this case, the probability that $x(t)$ has a discrete value x_k is zero, and we will be concerned with the probability that the value of $x(t)$ falls in some interval $(x_1,\ x_1 + \Delta x_1)$. The discrete case, Eq. 50, may now be directly applied to the continuous case simply by replacing the discrete event $\left[x(t_k) = x_k\right]$ by the continuous event $\left[x_k \leqslant x(t_k) < x_k + \Delta x_k\right]$. Then, paralleling our definition of the first probability distribution density, we may define the second probability distribution density by the equation

$$W_2(x_1;\ x_2,\ \tau) \equiv \lim_{\substack{\Delta x_1 \to 0 \\ \Delta x_2 \to 0}} \frac{P\left[x_1 \leqslant x(t) < x_1 + \Delta x_1;\ x_2 \leqslant x(t+\tau) < x_2 + \Delta x_2\right]}{\Delta x_1\,\Delta x_2} \tag{51}$$

or by the equation

$$P\left[a \leq x(t) < b;\ c \leq x(t+\tau) < d\right] = \int_c^d \int_a^b W_2(x_1;\ x_2, \tau)\, dx_1\, dx_2 \quad . \tag{52}$$

The joint probability in Eq. 51 may not tend to zero as Δx_1 and Δx_2 tend to zero, hence in the most general case, $W_2(x_1;\ x_2, \tau)$ may consist in part or entirely of impulse functions.

The joint probability of occurrence of two events is related to the probability of occurrence of one of those events by means of the conditional probability. In the discrete case we have

$$P\left[x(t) = x_1;\ x(t+\tau) = x_2\right] = P\left[x(t) = x_1\right]\ P\left[x(t) = x_1 \middle|\ x(t+\tau) = x_2\right] \tag{53}$$

where $P\left[x(t) = x_1 \middle|\ x(t+\tau) = x_2\right]$ is the conditional probability that $x(t+\tau)$ equals x_2, subject to the hypothesis that $x(t)$ equalled x_1. If $P\left[x(t) = x_1\right]$ is nonzero, we may rewrite Eq. 53 as the defining equation for the conditional probability (Cramér, p. 157, ref. 7)

$$P\left[x(t) = x_1 \middle|\ x(t+\tau) = x_2\right] \equiv \frac{P\left[x(t) = x_1;\ x(t+\tau) = x_2\right]}{P\left[x(t) = x_1\right]} \quad . \tag{54}$$

Similarly, in the case of continuous random variables, we may define a conditional probability distribution density (Cramér, p. 269, ref. 7) by the equation

$$W_2(x_1 \,|\, x_2, \tau) \equiv \frac{W_2(x_1;\, x_2, \tau)}{W_2(x_1)} \tag{55}$$

which is identical in form with Eq. 54.

By their very definition, the conditional probabilities, and the conditional probability distribution density, express the statistical relationship between two points on the voice wave, spaced in time by a given amount. A statistical function having a similar character is the autocorrelation function (2) defined by

$$\phi_x(\tau) = \lim_{T\to\infty} \frac{1}{T} \int_0^T x(t)\, x(t+\tau)\, dt \quad . \tag{56}$$

Thus the autocorrelation function is the average product of the amplitudes of two points on the voice wave spaced τ seconds apart. By our assumption of the stationary statistical character of the voice wave, we may replace the time average in the definition of the autocorrelation function with the statistical average

$$\phi_x(\tau) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x(t)\, x(t+\tau)\, W_2(x_1;\ x_2, \tau)\, dx_1\, dx_2 \quad . \tag{57}$$

However, by Eq. 55, we may also express the autocorrelation function in terms of the conditional probability distribution density

$$\phi_x(\tau) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} x(t)\, x(t+\tau)\, W_2(x_1 \big| x_2, \tau)\, W_1(x_1)\, dx_1\, dx_2 \quad . \tag{58}$$

The autocorrelation function is thus related to the conditional probability distribution density through a definite integral. We might therefore say that the conditional probability distribution density gives more detailed information about the statistical structure of a random process than does the autocorrelation function. This may or may not be an advantage. For example, in the problem of spectrum determination, the autocorrelation function contains all of the necessary information.

Let us now consider the limiting values of the conditional probabilities as functions of τ. If $x(t)$ and $x(t+\tau)$ are statistically independent, the conditional probability becomes

$$\begin{aligned} P\left[x(t) = x_1 \big| x(t+\tau) = x_2\right] &= P\left[x(t+\tau) = x_2\right] \\ &= P\left[x(t) = x_2\right] \end{aligned} \tag{59}$$

by the definition of statistical independence. For this case, the joint probability is given by

$$P\left[x(t) = x_1;\ x(t+\tau) = x_2\right] = P\left[x(t) = x_1\right] P\left[x(t) = x_2\right] \quad . \tag{60}$$

That is, in the case of independent random variables, the joint probability is equal to the product of the independent probabilities.

Most generally, $x(t)$ and $x(t+\tau)$ are not independent random variables. However, if we are dealing with stationary random processes containing no periodicities, then as τ becomes larger and larger, the relation between $x(t)$ and $x(t+\tau)$ becomes more and more tenuous (6). Thus generally we have the relation

$$\lim_{\tau\to\infty} P\left[x(t) = x_1 \big| x(t+\tau) = x_2\right] = P\left[x(t) = x_2\right] \quad . \tag{61}$$

In the case of continuous random variables, the expression equivalent to Eq. 61 is

$$\lim_{\tau\to\infty} W_2(x_1 \big| x_2, \tau) = W_1(x_2) \tag{62}$$

where again we assume stationary random processes containing no periodicities.

On the other hand, as τ approaches zero, the conditional probability should attain a value corresponding to certainty if x_2 equals x_1, or improbability if x_2 does not equal x_1. Thus in the case of discrete random variables

$$\lim_{\tau\to 0} P\left[x(t) = x_1 \big| x(t+\tau) = x_2\right] = \begin{cases} 1 \text{ if } x_2 = x_1 \\ 0 \text{ if } x_2 \neq x_1 \end{cases} \quad . \tag{63}$$

The equivalent expression for the case of the continuous random variable is

$$\lim_{\tau \to 0} W_2(x_1 | x_2, \tau) = \delta(x_2 - x_1) \qquad . \tag{64}$$

In this expression, $\delta(x_2 - x_1)$ is the impulse function

$$\delta(x_2 - x_1) = \begin{cases} \delta(x_1) & \text{if } x_2 = x_1 \\ 0 & \text{if } x_2 \neq x_1 \end{cases} \tag{65}$$

such that

$$\int_{-\infty}^{+\infty} \delta(x)\, dx = 1 \qquad . \tag{66}$$

Method of measurement

Before solving the problem of how to measure, we must decide what to measure. The speech wave is a continuous random variable; therefore the pertinent statistic is the conditional probability distribution density, $W_2(x_1 | x_2, \tau)$.

Let us consider the problem of measurement of $W_2(x_1 | x_2, \tau)$. Our equipment can measure a probability directly, but can measure a probability density only indirectly through a limiting process. Thus we must establish a limiting relation between the density $W_2(x_1 | x_2, \tau)$ and some probability. For this purpose, let us introduce a conditional probability, $P(x_1 | x_2, \tau)$, defined by

$$P(x_1 | x_2, \tau) \mathrel{\overline{=}} P\left[x_1 \leqslant x(t) < x_1 + \Delta x_1 \middle| x_2 \leqslant x(t+\tau) < x_2 + \Delta x_2\right] \qquad . \tag{67}$$

From the definition of conditional probability we then obtain

$$P(x_1 | x_2, \tau) = \frac{P(x_1;\, x_2, \tau)}{P(x_1)} \tag{68}$$

where $P(x_1;\, x_2, \tau)$ and $P(x_1)$ are defined by equations analogous to Eq. 67. If Δx_1 and Δx_2 are sufficiently small, the definitions of $W_2(x_1;\, x_2, \tau)$ and $W_1(x_1)$ show us that

$$P(x_1 | x_2, \tau) \doteq \frac{W_2(x_1;\, x_2, \tau)\, \Delta x_1\, \Delta x_2}{W_1(x_1)\, \Delta x_1} \tag{69}$$

to a high degree of approximation. Therefore, from the definition of $W_2(x_1 | x_2, \tau)$, Eq. 55, we obtain

$$W_2(x_1 | x_2, \tau) \doteq \frac{P(x_1 | x_2, \tau)}{\Delta x_2} \tag{70}$$

an approximation that improves as Δx_1 and Δx_2 become smaller.

A complete study of $W_2(x_1 | x_2, \tau)$ would require the measurement of $P(x_1 | x_2, \tau)$ for sets of intervals Δx_1 and Δx_2 covering the entire range of variation of x_1 and x_2. Such an undertaking would be quite formidable. In view of the preliminary nature of our investigation, it was decided to study only a few "cross-section" distributions derived from the conditional probability distribution density, $W_2(x_1|x_2, \tau)$. One type of cross-section distribution chosen was $P(x_1|x_1, \tau)$, where x_1 is fixed and τ is varied. The other type chosen was $W_2(x_1|x, \cdot)$, where x_1 and τ are fixed, and x is varied. The combination of these distributions provides a picture of the three-dimensional solid representing $W_2(x_1|x_2, \tau)$ as a function of x_2 and τ, with x_1 fixed. From such a picture we may gain an insight into the statistical relation between two points, spaced in time, on the speech wave.

Certain properties of $P(x_1|x_1, \tau)$ and of $W_2(x_1|x, \tau)$ can be determined analytically. Under the assumption of a stationary random process having no periodicities, the limiting values of $P(x_1|x_1, \tau)$ may be obtained directly from Eqs. 61 and 63, and are given by

$$\lim_{\tau \to 0} P(x_1|x_1, \tau) = 1 \tag{71}$$

and

$$\lim_{\tau \to \infty} P(x_1|x_1, \tau) = P(x_1) \quad . \tag{72}$$

Similarly, under the assumption of a stationary random process having no periodicities, the limiting values of $W_2(x_1|x, \tau)$ are obtainable from Eqs. 62 and 64, and are given by

$$\lim_{\tau \to 0} W_2(x_1|x, \tau) = \delta(x - x_1) \tag{73}$$

and

$$\lim_{\tau \to \infty} W_2(x_1|x, \tau) = W_1(x) \quad . \tag{74}$$

As may be shown by a discussion paralleling that leading up to Eq. 70, the distribution $W_2(x_1|x, \tau)$ may be determined by measuring the probability $P(x_1|x_2, \tau)$. Thus our problem becomes that of measuring $P(x_1|x_1, \tau)$ and $P(x_1|x_2, \tau)$. If we can show that these probabilities are directly related to time averages, we may make extensions of the measurement techniques developed in the preceding section to cover the present problem, and we may apply the same statistical accuracy study. For this purpose let us introduce the derived random variable, $\xi(t)$, defined by the equation

$$\xi(t) = \begin{cases} 1 \text{ if } x_1 \leq x(t) < x_1 + \Delta x_1 \\ 0 \text{ otherwise} \quad . \end{cases} \tag{75}$$

The random variables $x(t)$, $\xi(t)$, and $\xi(t + \tau)$ are shown in Fig. XIII.

Let us consider the autocorrelation function of the derived random variable, $\xi(t)$. Since it may assume only discrete values, the second probability distribution density of $\xi(t)$ consists entirely of impulses, and the integral in Eq. 57 becomes a double summation

$$\phi_\xi(\tau) = \sum_{i=1}^{2} \sum_{k=1}^{2} \xi_i \xi_k \, P\Big[\xi(t) = \xi_i;\ \xi(t+\tau) = \xi_k\Big] \tag{76}$$

where ξ_i and ξ_k have the two possible values zero and one. Evaluation of the summation then gives

$$\phi_\xi(\tau) = P\Big[\xi(t) = 1;\ \xi(t+\tau) = 1\Big] \quad . \tag{77}$$

However, we may express the joint probability in terms of the conditional probability, hence

$$\phi_\xi(\tau) = P\Big[\xi(t) = 1 \Big| \xi(t+\tau) = 1\Big] \, P\Big[\xi(t) = 1\Big] \quad . \tag{78}$$

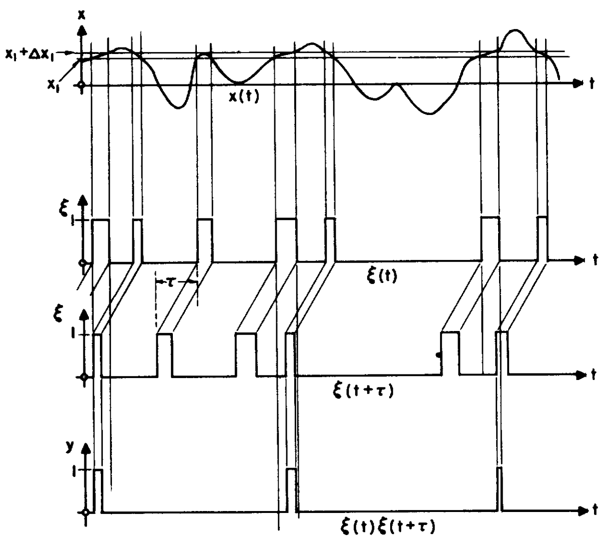


Fig. XIII Conditional probability derived random variables.

Then remembering the definition of the derived random variable, we finally obtain

$$\phi_\xi(\tau) = P(x_1 | x_1, \tau) \, P(x_1) \quad . \tag{79}$$

Thus the autocorrelation function of the derived random variable is equal to the conditional probability $P(x_1 | x_1, \tau)$ times the limiting value of $P(x_1 | x_1, \tau)$ as τ approaches infinity.

The basic definition of the autocorrelation function is in terms of a time average. If we then define the average over a finite interval, of duration T, of the product of $\xi(t)$ times $\xi(t+\tau)$ as

$$< \xi(t) \, \xi(t+\tau) >_T \;\overset{=}{=}\; \frac{1}{T} \int_0^T \xi(t) \, \xi(t+\tau) \, dt \tag{80}$$

the autocorrelation function of the derived random variable is given by

$$\phi_\xi(\tau) = \lim_{T\to\infty} < \xi(t) \, \xi(t+\tau) >_T \quad . \tag{81}$$

As was shown earlier the probability of occurrence of the event $\Big[\xi(t) = 1\Big]$ is equal to the time average of the derived random variable $\xi(t)$. That is

$$P(x_1) = < \xi(t) >_{av} = \lim_{T\to\infty} < \xi(t) >_T \tag{82}$$

where we have defined

$$< \xi(t) >_T = \frac{1}{T} \int_0^T \xi(t)\, dt \quad . \tag{83}$$

From Eq. 79, the conditional probability is given by

$$P(x_1 | x_1, \tau) = \frac{\phi_\xi(\tau)}{P(x_1)} \quad . \tag{84}$$

Then from Eqs. 81 and 82 we see that

$$P(x_1 | x_1, \tau) = \lim_{T \to \infty} \frac{< \xi(t)\, \xi(t + \tau) >_T}{< \xi(t) >_T} \quad . \tag{85}$$

Thus the problem of measuring the conditional probability $P(x_1 | x_1, \tau)$ reduces to that of measuring the two finite period time averages, and our problem is essentially similar to that discussed in Method of Measurement.

Let us now consider the problem of determining the desired time averages by a process of sampling and summation. Suppose that n samples are made of $\xi(t)$, and of the product $\xi(t)\,\xi(t + \tau)$, at a periodic rate throughout a measurement interval of duration T. Then by defining a quantity ν as the total number of occurrences of the event $\left[\xi(t) = 1\right]$ in the n samples in the interval T, we have

$$< \xi(t) >_T = \lim_{n \to \infty} \frac{\nu}{n} \tag{86}$$

as was proved earlier. In an entirely similar manner, we may define a quantity ν_p as the total number of occurrences of the event $\left[\xi(t)\, \xi(t + \tau) = 1\right]$, and hence of the joint event $\left[\xi(t) = 1; \xi(t + \tau) = 1\right]$, in the n samples in the interval T. We then have

$$< \xi(t)\, \xi(t + \tau) >_T = \lim_{n \to \infty} \frac{\nu_p}{n} \quad . \tag{87}$$

Hence finally from Eqs. 85-87 we obtain

$$P(x_1 | x_1, \tau) = \lim_{T \to \infty} \lim_{n \to \infty} \frac{\nu_p}{\nu} \quad . \tag{88}$$

Thus the desired conditional probability $P(x_1 | x_1, \tau)$ may be obtained as the limiting ratio of the number of occurrences of the joint event $\left[\xi(t) = 1;\ \xi(t + \tau) = 1\right]$ to the number of occurrences of the event $\left[\xi(t) = 1\right]$ as both the duration of the measurement interval and the number of samples taken in that measurement interval increase without limit.

A simplified block diagram of the equipment utilized to measure the conditional probability $P(x_1 | x_1, \tau)$ is shown by Fig. XIV. Up to the output of the level selector, the operation of this system is identical with the operation of the system used to measure $P(x_1)$, which has been described. Thus at the output of the level selector, there is a pulse each time the voice wave amplitude (at the instant of sampling) lies within the interval $(x_1, x_1 + \Delta x_1)$. The output of the level selector is applied to a delay unit and to a coincidence circuit. The delay unit generates an output pulse at a time (t) if a pulse was applied to its input at a previous time $(t - \tau)$. The output of the delay unit is then applied to the coincidence circuit and to counter No. 2. The coincidence circuit generates an output pulse whenever it receives a pulse from the delay unit simultaneously with a pulse from the level selector. Thus the output of the coincidence circuit corresponds to the joint occurrence of the events $\left[x_1 \leqslant x(t) < x_1 + \Delta x_1\right]$ and $\left[x_1 \leqslant x(t - \tau) < x_1 + \Delta x_1\right]$.

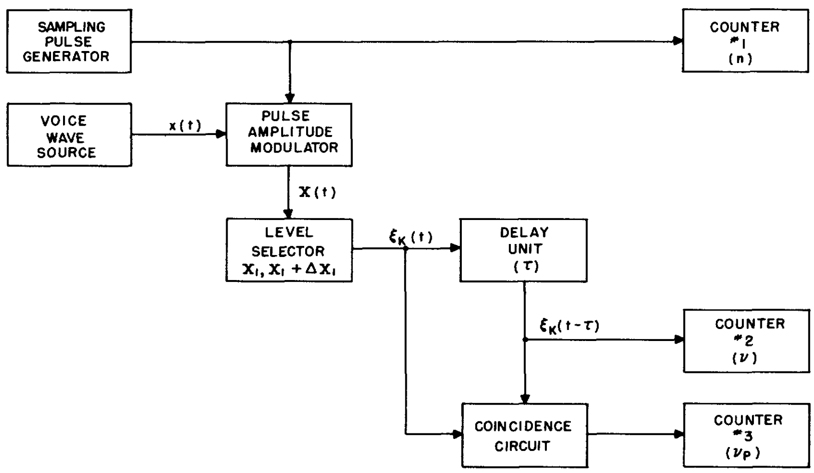


Fig. XIV Measurement of the conditional probability $P(x_1 \; x_1, \tau)$.

Because of the stationary character of x(t), the location of the time origin may be changed. Therefore

$$\lim_{n \to \infty} \left(\frac{\nu_p}{\nu} \right) = \frac{< \xi(t) \; \xi(t + \tau) >_T}{< \xi(t) >_T} \quad . \tag{89}$$

Thus the limit of the ratio of the reading of counter No. 3 to that of counter No. 2, as both the number of samples in a measurement interval and the duration of a measurement interval are increased without limit, is the desired conditional probability, $P(x_1 | x_1, \tau)$.

The conditional probability $P(x_1 | x_2, \tau)$ may be determined by a process of sampling and summation in a way analogous to that used in the determination of $P(x_1 | x_1, \tau)$. Let us define a quantity, ν_2, as the number of joint occurrences of the events $\left[x_1 \leqslant x(t) < x_1 + \Delta x_1\right]$ and $\left[x_2 \leqslant x(t + \tau) < x_2 + \Delta x_2\right]$ in n samples obtained during a measurement interval T. Then, a paraphrase of the derivation of Eq. 88 would show that

$$P(x_1 | x_2, \tau) = \lim_{T \to \infty} \lim_{n \to \infty} \frac{\nu_2}{\nu} \quad . \tag{90}$$

A simplified block diagram of the equipment utilized in the measurement of the conditional probability distribution density is shown by Fig. XV. A comparison of this figure with that concerning $P(x_1 | x_1, \tau)$, Fig. XIV, shows that the only difference is the

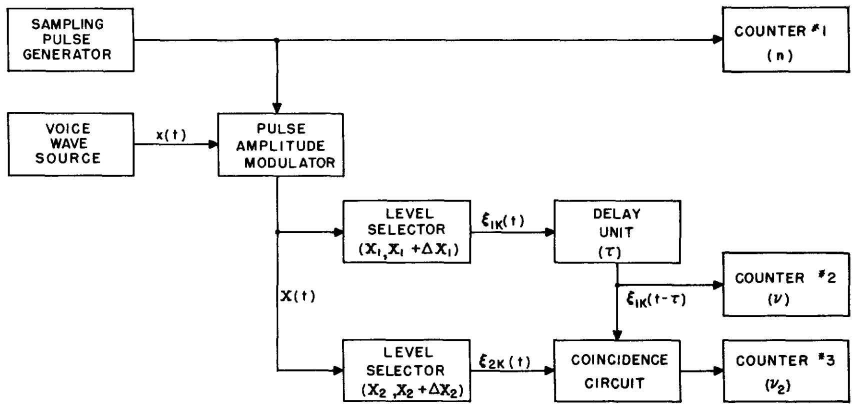


Fig. XV Measurement of the conditional probability distribution density, $W_2(x_1|x_1, \tau)$.

addition of a level selector to determine the occurrence of the event $\left[x_2 \leqslant x(t) < x_2 + \Delta x_2\right]$. The output of the coincidence circuit then corresponds to the joint occurrence of the events $\left[x_1 \leqslant x(t-\tau) < x_1 + \Delta x_1\right]$ and $\left[x_2 \leqslant x(t) < x_2 + \Delta x_2\right]$. Hence the ratio of counter No. 3 reading (ν_2) to counter No. 2 reading (ν), in the limit as both the number of samples in a measurement interval increase without limit, is the desired conditional probability $P(x_1|x_2, \tau)$.

D. Conditional Probability Distributions – Experimental Results

Measurements of the conditional probability $P(x_1|x_1, \tau)$ were made for the voices of three speakers: one female (KA) and two male (WD and JC). As in the study of the first probability distribution density, the voice waves were recorded with a magnetic tape recorder in an anechoic chamber and in a live studio. The measurement time per data point varied from a minimum of five minutes for small values of τ, to a maximum of fifteen minutes for large values of τ. The probable error varies from about five percent, for small values of τ, to about ten percent for large values of τ. The results of these measurements are plotted in Figs. XVII, XVIII, and XIX.

In the study of the first probability distribution density, $W_1(x)$, it was postulated that the exponential portion of that distribution was mainly representative of the voiced sounds, while the spiked portion was due mainly to the unvoiced sounds. References to the various plots of $W_1(x)$ thus show that the form of the conditional probability $P(x_1|x_1, \tau)$, for the values of x_1 greater than the root mean square, should be determined principally by the characteristics of the voiced sounds. On the other hand, as x_1 is decreased in magnitude, the form of $P(x_1|x_1, \tau)$ should be successively modified by the characteristics of the unvoiced sounds.

A study of the instantaneous waveform, Fig. XVI, of various voiced sounds shows that these sounds possess a common characteristic: a complete sound usually consists of a number of repetitions of a basic wave pattern. The amplitude of the wave envelope may change throughout a voiced sound, but the pattern is rather well perserved.
The fact that a basic pattern is repeated in time means that there will be an increase

Fig. XVI Instantaneous patterns of voiced sounds, "the U.S.".
Male voice (WD) in anechoic chamber and 500 cps timing wave.

in the conditional probability for values of τ about an average voiced sound pattern repetition period, τ_{vp}. This period is the reciprocal of the so-called pitch frequency. If the voice wave consisted entirely of a basic pattern repeated exactly periodically, then $P(x_1 \mid x_1, \tau)$ would reach the value unity for values of τ equal to integral multiples of τ_{vp}. By definition, random waves are not periodic; however it is possible for a random wave to be "almost periodic". An example might be a wave having a slowly varying, random change in τ_{vp}, or a similar change in envelope amplitude. Many voiced sounds have this type of behavior. Such a wave may have a well-defined average pattern period, τ_{vp}; in this case there will be peaks of $P(x_1 \mid x_1, \tau)$ for values of equal to $k\tau_{vp}$, where k is an integer. Such peaks will be less than unity, and must eventually decrease to the limiting value, $P(x_1)$, as τ increases without limit. On the other hand, if the wave is almost purely random, as is the case of fricatives, the concept of an average repetition period loses meaning and $P(x_1 \mid x_1, \tau)$ decreases more or less monotonically to its final value as τ is increased.

The effect on the conditional probability $P(x_1|x_1, \tau)$ of varying the magnitude of x_1 is shown by Fig. XVII. Referring to the plots of the first probability distribution density, $W_1(x)$, for the same voice, (WD), Fig. VI, it may be seen that the three values of x_1 occurring in Fig. XVII correspond to three different regions of the $W_1(x)$ distribution: the spike ($x_1 = 0.33 \times$ rms), the transition region ($x_1 = 0.65 \times$ rms), and the exponential region ($x_1 = 1.3 \times$ rms). Viewing Fig. XVII in the light of the preceding discussion, it would then seem that the increase in $P(x_1|x_1, \tau)$ around τ equal to 7.3 msec was due to the average voiced sound-pattern repetition period. In order to check this point, a photograph was made of a five and one-half second sample of the voice wave studied in Fig. XVII. In this sample, some thirty successive changes in pattern were noted. The pattern periods ranged from about 4 msec to 9 msec with an average of 7.6 msec, thus confirming our conclusions.

Plots of $P(x_1|x_1, \tau)$ for different speakers are shown in Fig. XVIII. The plot of the conditional probability for the female voice, (KA), shows a smaller value of τ_{vp}, and hence a higher pitch, than do the plots for the male voices, as is to be expected. The existence of a succession of peaks in the plot of $P(x_1|x_1, \tau)$ for the female voice, rather than a single peak, indicates that the voiced sounds of this speaker are more nearly periodic than are those of either of the male speakers.

The measurements for the plots in Figs. XVII and XVIII were made on voices recorded in an anechoic chamber. The effect of environment on this distribution is indicated by Fig. XIX. The two plots in Fig. XIX are for the same voice recorded first in an anechoic chamber and then in a live studio. The effect of reflections in the live studio appears in the form of humps in the one to five millisecond region.

Measurements of the conditional probability distribution density, $W_2(x_1 \mid x, \tau)$, were made on only one voice, (WD), and on that voice only in the anechoic chamber. The results of these measurements are shown in Fig. XX. It is of interest to relate these plots to the plot of the conditional probability $P(x_1|x_1, \tau)$ for the same voice and same

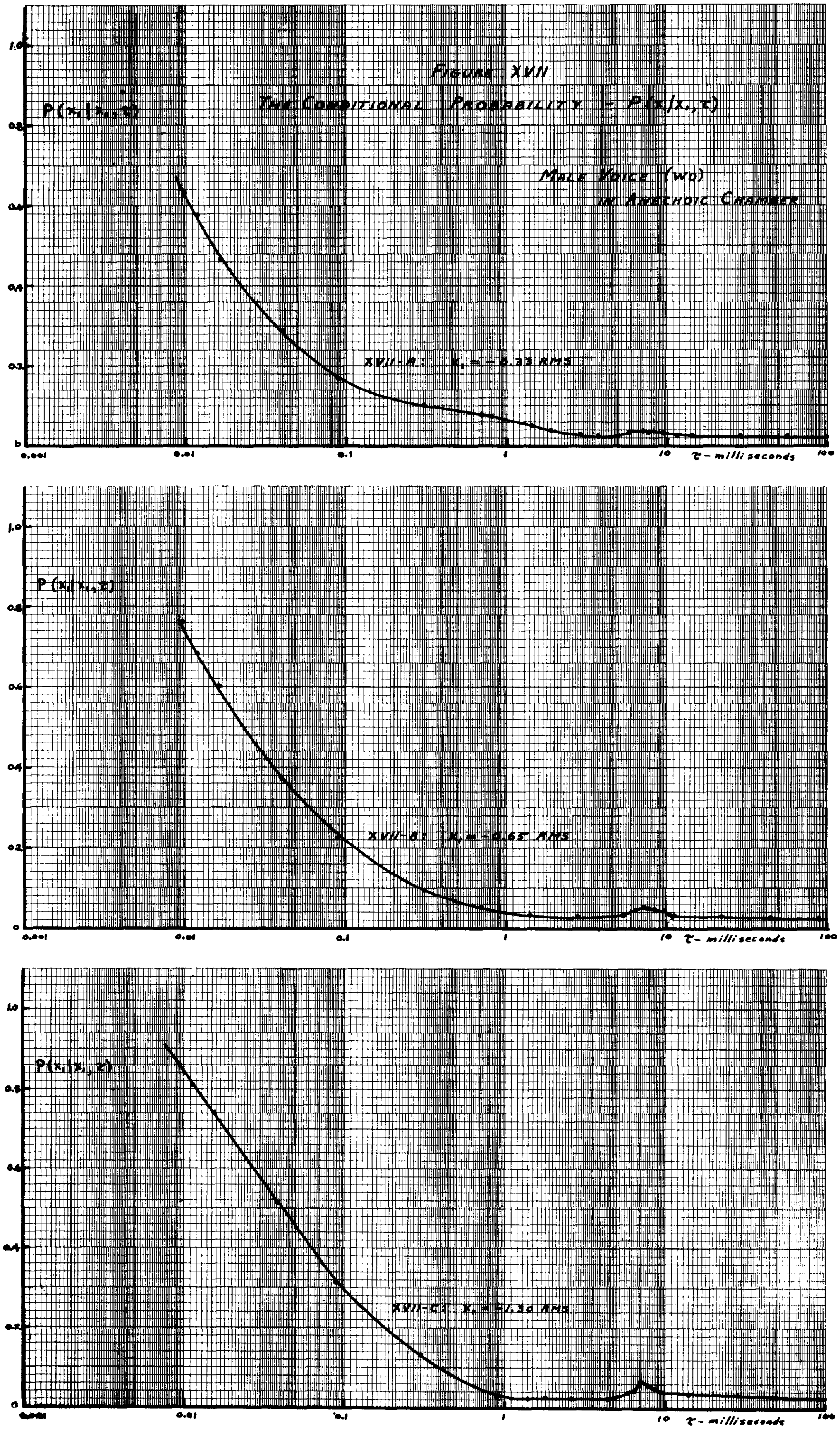


FIGURE XVII
THE CONDITIONAL PROBABILITY – P(x1|x1,τ)
MALE VOICE (WD)
IN ANECHOIC CHAMBER
P(x1|x1,τ)
XVII-A: x1 = −0.33 RMS
τ - milliseconds
P(x1|x1,τ)
XVII-B: x1 = −0.65 RMS
τ - milliseconds
P(x1|x1,τ)
XVII-C: x1 = −1.30 RMS
τ - milliseconds

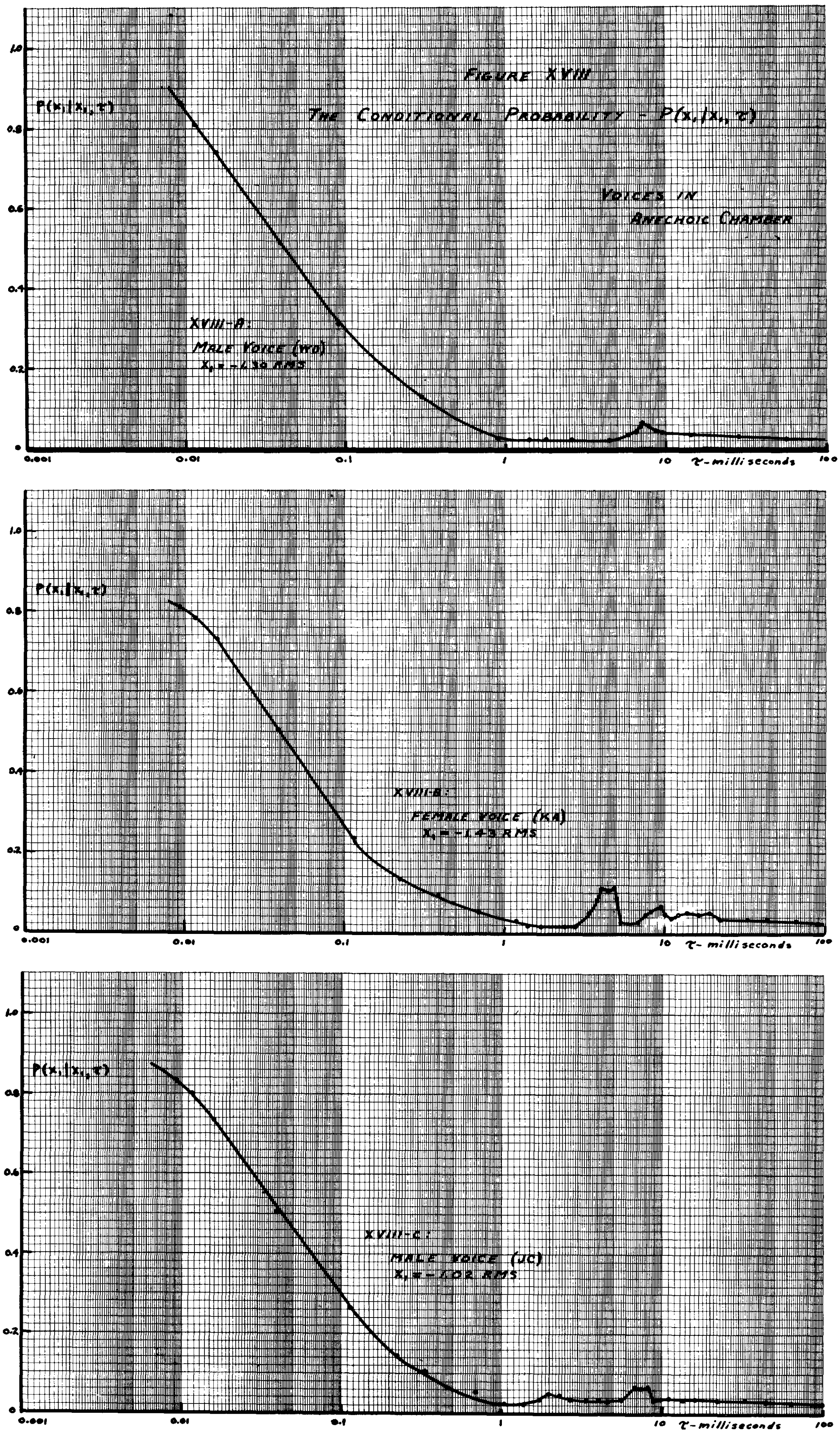


FIGURE XVIII
THE CONDITIONAL PROBABILITY - P(X1|X2, τ)
VOICES IN ANECHOIC CHAMBER
P(x1|x2, τ)
XVIII-A:
MALE VOICE (WD)
X2 = -1.30 RMS
τ - milliseconds
XVIII-B:
FEMALE VOICE (KA)
X2 = -1.43 RMS
XVIII-C:
MALE VOICE (JC)
X2 = -1.02 RMS

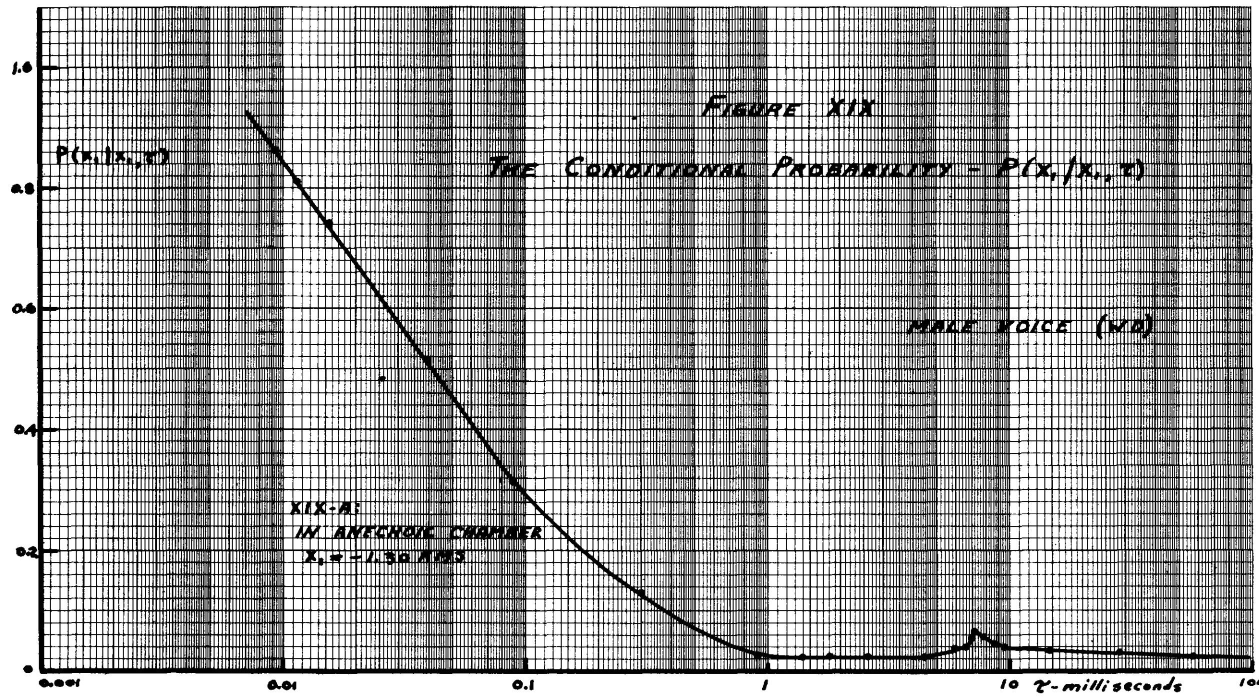
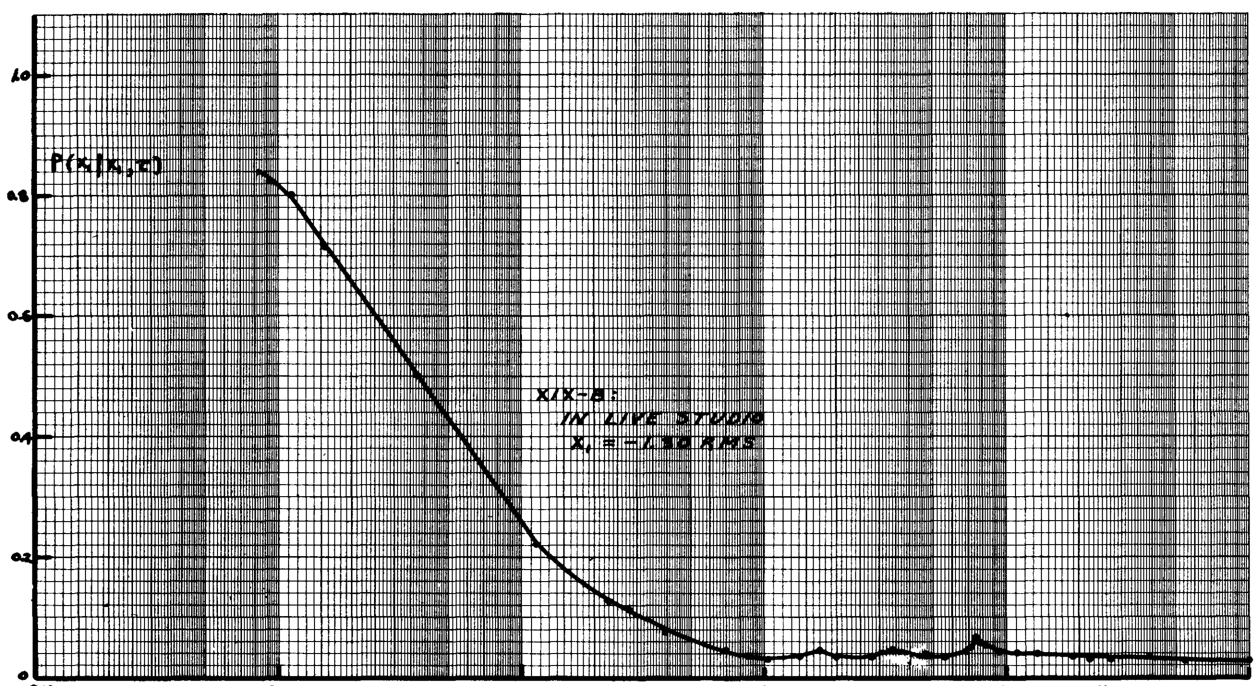


FIGURE XIX
THE CONDITIONAL PROBABILITY - $P(x_1|x_1, \tau)$
$P(x_1|x_1,\tau)$
MALE VOICE (WD)
XIX-A:
IN ANECHOIC CHAMBER
X_1 = -1.50 RMS
τ-milliseconds


$P(x_1|x_1,\tau)$
XIX-B:
IN LIVE STUDIO
X_1 = -1.50 RMS
τ-milliseconds

value of x_1 as given by Fig. XVII. For $\tau = 12\ \mu$sec, $P(x_1|x_1, \tau)$ has a high value and $W_2(x_1|x, \tau)$ has the form of an impulse function centered on x_1. (Note the change in scale for the W_2 axis between the plot for this value of τ and those for larger values of τ.) These results are consistent with the limiting relations for these functions as τ approaches zero, as shown by Eqs. 71 and 73. As τ increases, two points on the voice wave separated in time by τ become more independent statistically. Thus as τ increases, $P(x_1|x_1, \tau)$ decreases toward $P(x_1)$, and $W_2(x_1|x, \tau)$ changes from an impulse centered on x_1 to a less peaked distribution centered on zero amplitude. A further increase in τ to about 7.4 msec brings us into the region of an increase in correlation because of the almost periodic character of voiced sounds. Thus the value of $P(x_1|x_1, \tau)$ increases and the center of $W_2(x_1|x, \tau)$ is shifted back toward x_1. Finally as τ is increased beyond this value, we again approach statistical independence. Hence $P(x_1|x_1, \tau)$ approaches $P(x_1)$, and $W_2(x_1|x, \tau)$ tends toward a form similar to that of the first probability distribution density, $W_1(x)$. These results are thus consistent with the limiting relations for these functions (as τ increases without limit) expressed by Eqs. 72 and 74.

II ZERO-CROSSING DISTRIBUTIONS

A. The First Probability Distributions

Studies have been made by Licklider (9) (10) and others of the psychoacoustic effects of severe amplitude limiting, or clipping, on the speech wave. These studies show that clipping produces no serious reduction in word intelligibility. This fact indicates that, for certain types of communication systems, it may be desirable to transmit only the speech-wave zero-crossing periods. Accordingly, the various distributions of these periods are of interest.

Definitions

A clipped speech wave, or zero-crossing wave, is a rectangular wave whose value changes when the speech wave passes through zero. An example would be the wave, $Z(t)$, defined by the relations

$$Z(t) \equiv \begin{cases} 1 \text{ if } 0 \leq x(t) \\ 0 \text{ if } x(t) < 0 \end{cases} \tag{91}$$

where $x(t)$ is the voice wave itself. Thus, the zero-crossing wave changes its state whenever the original wave passes through zero, and the only characteristic of the original wave preserved by such a transformation is the interval between two successive zero crossings.

Let us now consider the distribution of the zero-crossing periods. Defining $T_o(t)$ as the duration of the zero-crossing period existing at time t, the probability $P(T_1)$, defined by the equation

$$P(T_1) \equiv P\left[T_1 \leqslant T_o(t) < T_1 + \Delta T_1\right] \tag{92}$$

is the probability that at a given instant of time, t, the duration of the corresponding zero-crossing period lies between the limits (T_1) and $(T_1 + \Delta T_1)$. It should be noted that this is not the same as defining $P(T_1)$ to be the limiting ratio of the number of periods of duration $(T_1, T_1 + \Delta T_1)$ to the total number of periods. The first probability distribution density, $W_1(T_o)$, of the zero-crossing periods may then be defined in terms of $P(T_1)$ by the equation

$$W_1(T_1) \equiv \lim_{\Delta T_1 \to o} \frac{P(T_1)}{\Delta T_1} \tag{93}$$

or, alternatively, we may define $W_1(T_o)$ by the equation

$$P\left[T_1 \leqslant T_o(t) < T_2\right] = \int_{T_1}^{T_2} W_1(T_o)\, dT \quad . \tag{94}$$

As before, the distribution density will consist of impulses if the duration of the zero-crossing periods may assume only discrete values.

Method of measurement

If the first probability distribution density, $W_1(T_o)$, of the zero-crossing periods is a reasonably slowly varying function of T_o, we may say that

$$W_1(T_1) \doteq \frac{P(T_1)}{T_1} \tag{95}$$

to a high degree of approximation. Thus our problem becomes essentially that of measuring the probability $P(T_1)$.

Let us see if we can now relate $P(T_1)$ to a time average in order to utilize the measurement techniques previously developed. For this purpose, let us define a derived random variable, $\xi(t)$, by the relations

$$\xi(t) \equiv \begin{cases} 1 \text{ if } T_1 \leqslant T_o(t) < T_1 + \Delta T_1 \\ 0 \text{ otherwise} \end{cases} \quad . \tag{96}$$

Then, by analogy with the development of sec. II-B, it can be shown that

$$P(T_1) = \bar{\xi} = \langle \xi(t) \rangle_{av} \quad . \tag{97}$$

Hence, the desired probability is equal to the average (statistical or time) of the derived random variable.

Apparently then, our problem is essentially the same as that considered in sec. II-B, and experimentally we need only construct the derived random variable and measure foldout

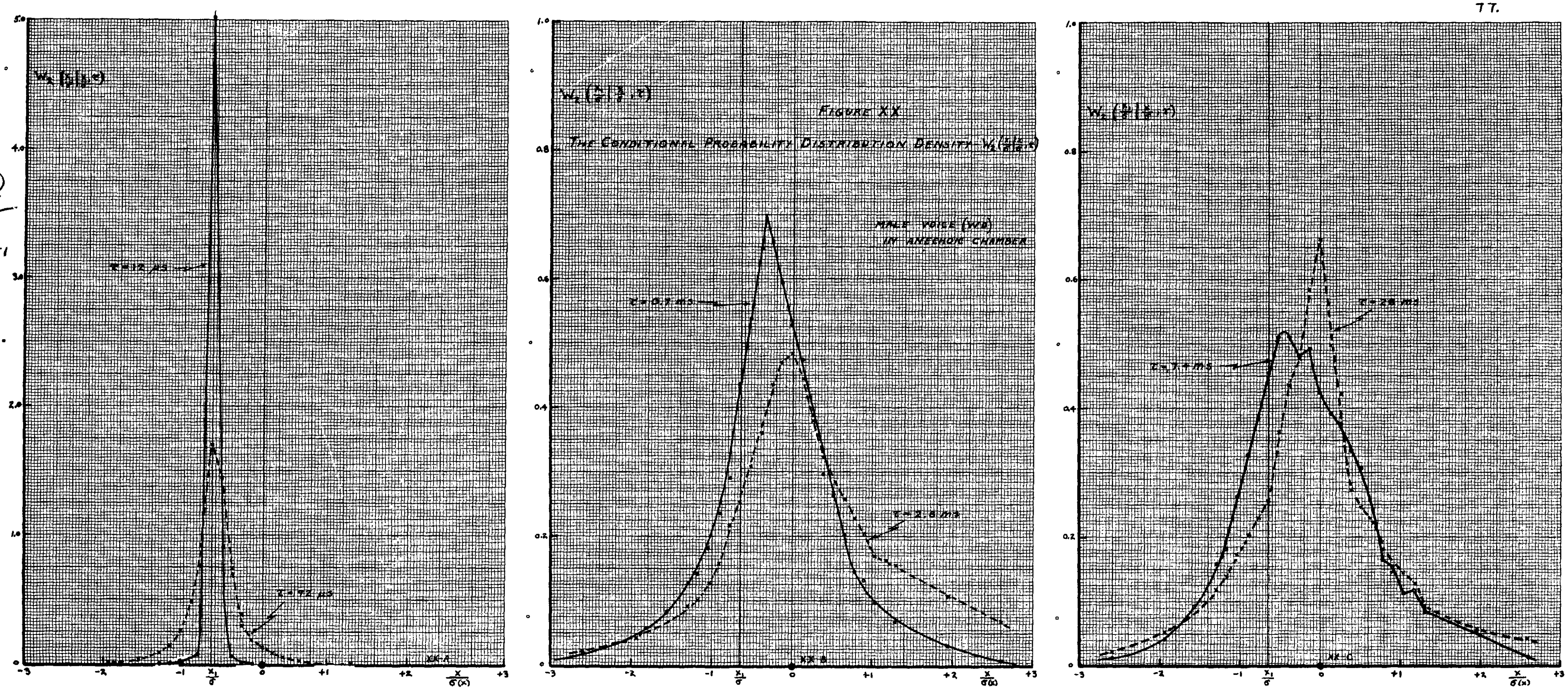


FIGURE XX

THE CONDITIONAL PROBABILITY DISTRIBUTION DENSITY

its time average. Practically, however, a difficulty arises in the generation of the derived random variable. The difficulty comes from the fact that $\xi(t)$ should equal one throughout a zero-crossing period whose duration falls within the prescribed limits. Our apparatus however cannot tell what the duration of a given zero-crossing period is until that period is completed!

One way out of this difficulty is indicated by the scheme of Fig. XXI. The voice wave, $x(t)$, is clipped to form the zero-crossing wave, $Z(t)$, which is applied to the period pulse generator. This unit generates an output pulse, $Z_k(t)$, whenever the voice wave crosses through zero, regardless of direction. The amplitude of the output pulse is proportional to the time interval between that pulse and the preceding pulse, and hence to the zero-crossing period. The level selector is then adjusted so that it generates an output pulse whenever the amplitude of the input pulse corresponds to a zero-crossing period whose duration falls in the interval $(T_1, T_1 + \Delta T_1)$. Counter No. 1 counts the total number, n, of period pulses, and counter No. 2 counts the number, ν, of period pulses corresponding to the event $\left[T_1 \leqslant T_0(t) < T_1 + \Delta T_1\right]$.

Suppose now that data are taken over a measurement period of duration T, and that n zero-crossing periods occur. The average of the derived random variable, $\xi(t)$, over the finite interval is then given by

$$< \xi(t) >_T = \frac{1}{T} \sum_{k=1}^{n} \xi_k T_k \tag{98}$$

where T_k is the duration of the kth zero-crossing period, and where

$$\xi_k = \begin{cases} 1 \text{ if } T_1 \leqslant T_k < T_1 + \Delta T_1 \\ 0 \text{ otherwise} \end{cases} \quad . \tag{99}$$

Now if ΔT_1 is small, T_k may be taken equal to T_1, and we then obtain

$$< \xi(t) >_T \doteq \frac{T_1}{T} \sum_{k=1}^{n} \xi_k = \frac{\nu T_1}{T} \quad . \tag{100}$$

As the probability $P(T_1)$ is equal to the infinite-interval time average of $\xi(t)$, we have

$$P(T_1) \doteq \lim_{T\to\infty} \frac{\nu T_1}{T} \quad . \tag{101}$$

Hence, the first probability distribution density, $W_1(T_0)$, of the zero-crossing period is given approximately by

$$W_1(T_1) \doteq \frac{\nu T_1}{T \Delta T_1} \tag{102}$$

and is given exactly by

$$W_1(T_1) = \lim_{T\to\infty} \lim_{\Delta T_1 \to 0} \frac{\nu T_1}{T \Delta T_1} \tag{103}$$

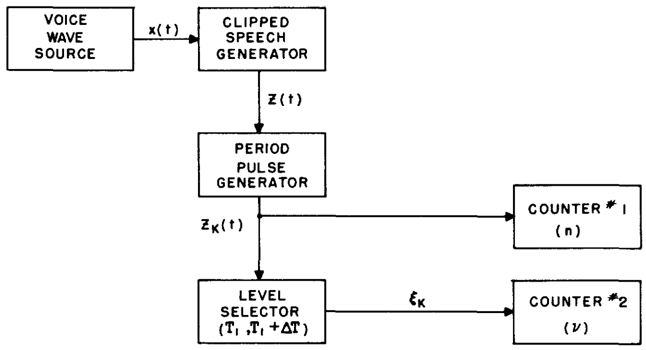


Fig. XXI Measurement of the first probability distribution density, $W_1(T_o)$.

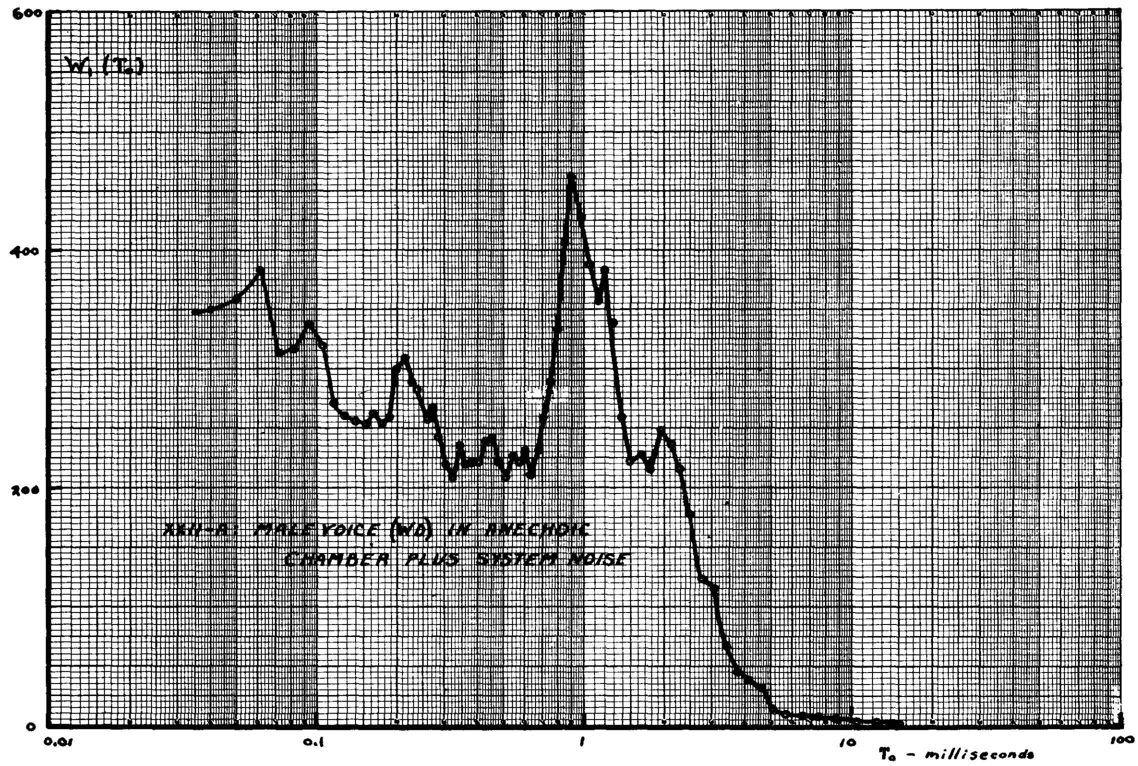
as the approximate Eq. 101 becomes an equality in the limit as ΔT_1 approaches zero.

An experimental determination of $W_1(T_o)$ may then be made by dividing the range of interest of T_o into a set of adjacent intervals, ΔT_k. In a measurement period T, the number, ν, of zero-crossing periods satisfying the event $\left[T_k \leqslant T_o(t) < T_k + \Delta T_k\right]$ may then be determined. From the calibration of the level selector, both T_k and ΔT_k are known, hence $W_1(T_o)$ may be calculated approximately from Eq. 102. The degree of approximation improves as ΔT_k is made smaller and as T is made larger.

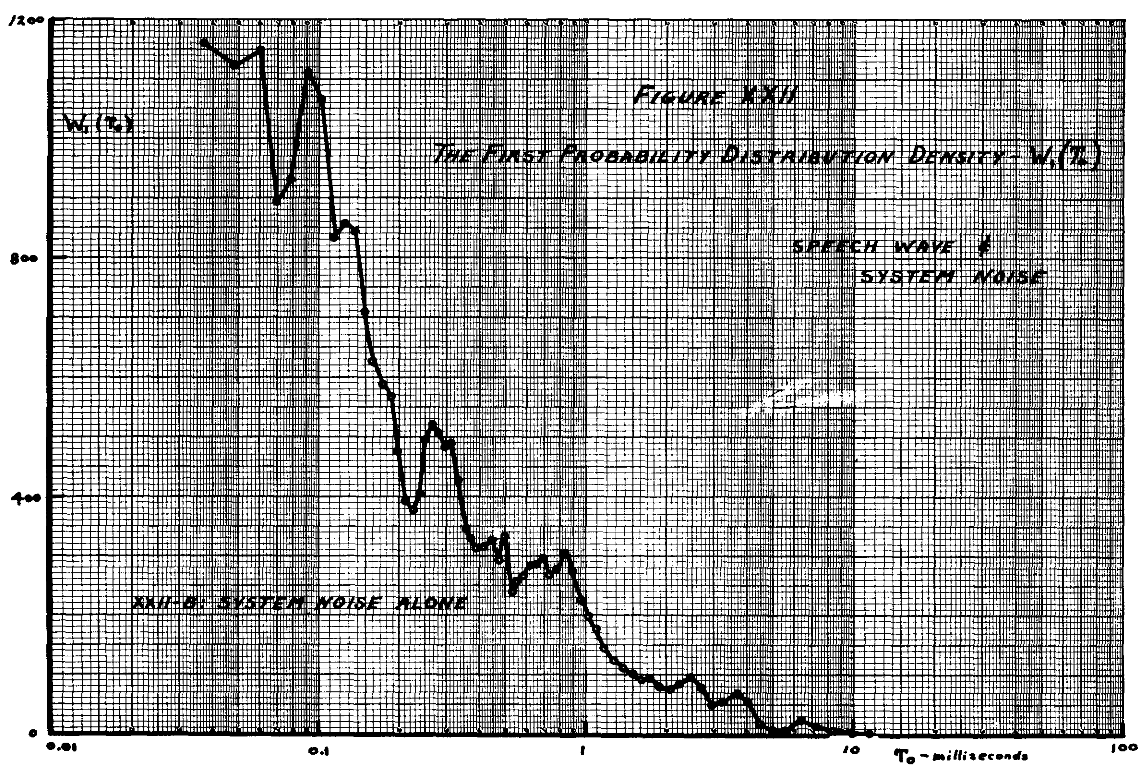
Experimental Results

Using the concepts and techniques discussed in the preceding sections, an attempt was made to measure the first probability distribution density, $W_1(T_o)$ of the zero-crossing periods of the speech wave. Immediately, however, a difficulty arose because of the unavoidable presence of system noise. In essence, the better the system is in transforming the speech wave into a rectangular wave, the smaller the noise amplitude must be in order not to produce a measurable effect. Thus a measured distribution will generally be due in part to the speech wave and in part to the system noise. For example, measurements of the first probability distribution density of a speech wave plus system noise, $W_1(T_o)_{S+N}$, and of system noise alone, $W_1(T_o)_N$, are plotted in Fig. XXII. A comparison of these plots shows a close correspondence of form for values of T_o less than a half millisecond. The question now arises: How much of the $W_1(T_o)_{S+N}$ distribution for T_o less than a half millisecond is due to the speech wave alone ? The answering of this question is not easy because of the basic similarity of the structures of the unvoiced sounds and of the system noise.

If the unvoiced sounds have significantly larger amplitudes than the amplitude of the system noise, then it should be possible to separate the noise from the speech wave effectively. Such a separation could be accomplished by the addition of a periodic bias signal whose amplitude is larger than that of the noise, but smaller than those of the unvoiced sounds. The bias signal will then override the noise and the randomly varying duration periods of the noise would be replaced by the essentially constant periods of the bias signal. A peak in the resulting $W_1(T_o)$ distribution will occur about the value of the bias period, but this peak will be isolated if the duration of the bias period is chosen so as to occur outside of the range of variation of the zero-crossing periods of the speech wave. There are then two choices: either the bias signal frequency must be higher than the highest frequency component of the speech wave, or it must be lower



600
400
200
0
$W_1(T_o)$
XXII-A: MALE VOICE (WD) IN ANECHOIC CHAMBER PLUS SYSTEM NOISE
0.01
0.1
1
10
100
T_o - milliseconds



1200
800
400
0
$W_1(T_o)$
FIGURE XXII
THE FIRST PROBABILITY DISTRIBUTION DENSITY - $W_1(T_o)$
SPEECH WAVE & SYSTEM NOISE
XXII-B: SYSTEM NOISE ALONE
0.01
0.1
1
10
100
T_o - milliseconds

than the lowest component. The choice of a supersonic bias signal is commonly made by investigators of the phychoacousitc aspects of clipped speech as the supersonic signal cannot be detected by the ear. However, reference to Fig. XXII shows that a very low bias signal frequency (i.e. very long bias period) is more suitable in our problem. Thus it was decided to use a 20-cps square-wave bias signal. The square waveform for the bias wave was chosen so that the transition of the bias wave through zero amplitude would be rapid. A plot is given in Fig. XXIII-A of the results of the measurement of of the $W_1(T_o)$ distribution for a speech wave plus a 20-cps square wave bias. The peak-to-peak amplitude of the bias signal was adjusted to be approximately twice that of the system noise.

A partial check on the validity of the experimental results obtained through the use of a bias signal was made as follows: Under the assumption that the unvoiced sounds are significantly larger in amplitude than the system noise, we may say approximately that we have either the speech wave or the noise present in the system at any given time. We are therefore dealing with mutually exclusive events whose probabilities are additive. Let us now define $P(T_1)_{S+N}$, $P(T_1)_S$, and $P(T_1)_N$ as the probabilities that the combined speech and noise wave, the speech wave alone, and the noise wave alone, respectively, have zero-crossing periods of duration $\left[T_1 \leqslant T_o < T_1 + \Delta T_1\right]$. Then

$$P(T_1)_{S+N} = dP(T_1)_N + (1 - d)\, P(T_1)_S \qquad (104)$$

where d is the fractional dead time (i.e. the fraction of time in which there is no speech in the combined wave). A study of a photograph of a five-second sample of the combined speech and noise wave indicated that the dead time was about 21.5 percent. This value for dead time, the period distribution of noise alone (Fig. XXII-B), and the period distribution of the combined speech and noise wave (Fig. XXII-A) were then substituted in Eq. 115, which was solved for $P(T_1)_S$. The resulting distribution is plotted as Fig. XXIII-B. A comparison of Figs. XXIII-A and XXIII-B shows that the experimentally obtained distribution is lower than the calculated distribution in the region below one half millisecond, and is higher in the region from one half to five milliseconds. These results indicate that the bias signal suppressed some of the unvoiced sounds as well as the noise, but that a fairly valid distribution was obtained experimentally.

An examination of photographs of a speech wave and its corresponding zero-crossing wave (Fig. XXIV) shows that for each voiced-sound pattern period there are several zero-crossing periods. Thus the peak in the first probability distribution density, $W_1(T_o)$, of the zero-crossing periods corresponds to the semi-periods of the speech wave within a voiced-sound pattern, rather than to the pattern period itself. If the pattern of the voiced sound is relatively simple (i.e. if the harmonic content within the voiced sound is not too high), the zero-crossing periods correspond to the fundamental frequency within that voiced sound. This statement corresponds to that made by the Northeastern University Visual Message Presentation group (ref. 17) that the short-time average

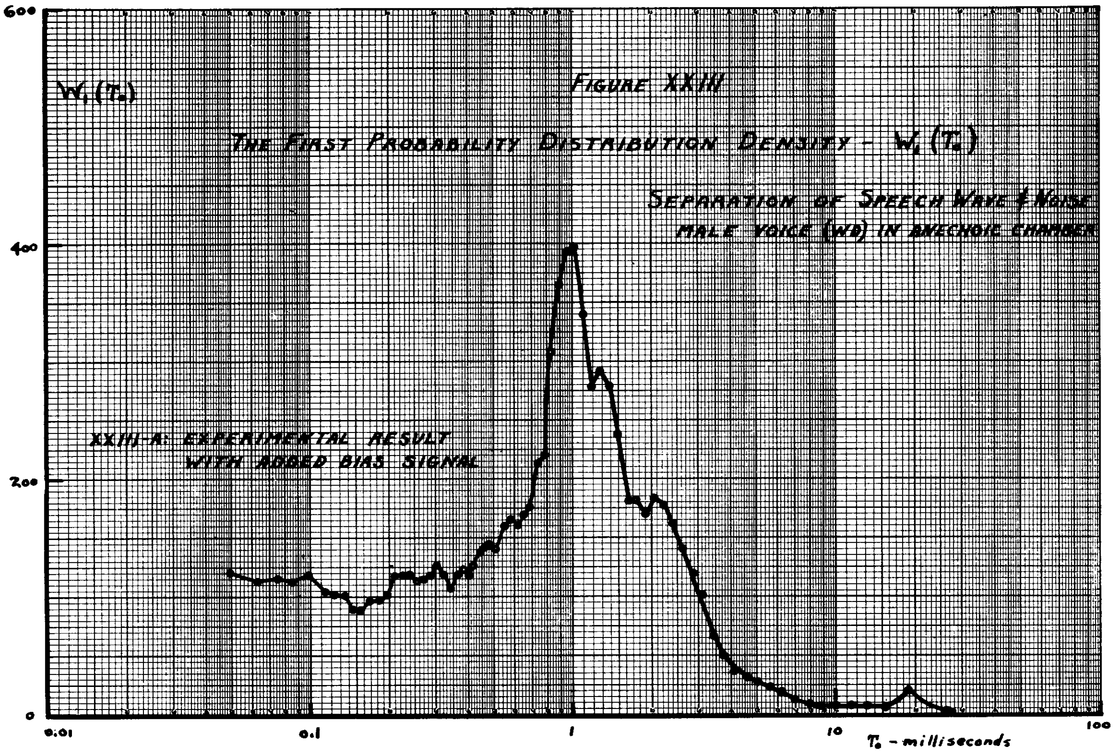
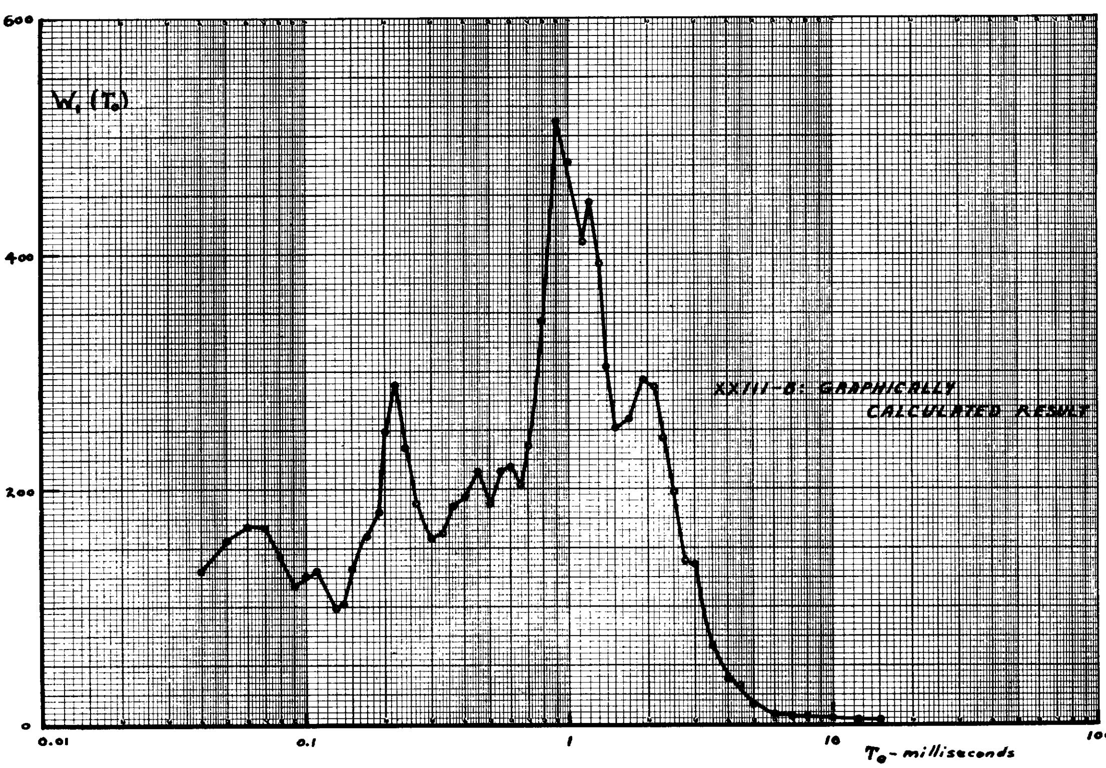


FIGURE XXIII
THE FIRST PROBABILITY DISTRIBUTION DENSITY - $W_1(T_0)$
SEPARATION OF SPEECH WAVE & NOISE
MALE VOICE (WB) IN ANECHOIC CHAMBER
$W_1(T_0)$
XXIII-A: EXPERIMENTAL RESULT
WITH ADDED BIAS SIGNAL
600
400
200
0
0.01
0.1
1
10
100
T_0 - milliseconds


$W_1(T_0)$
XXIII-B: GRAPHICALLY
CALCULATED RESULT
600
400
200
0
0.01
0.1
1
10
100
T_0 - milliseconds

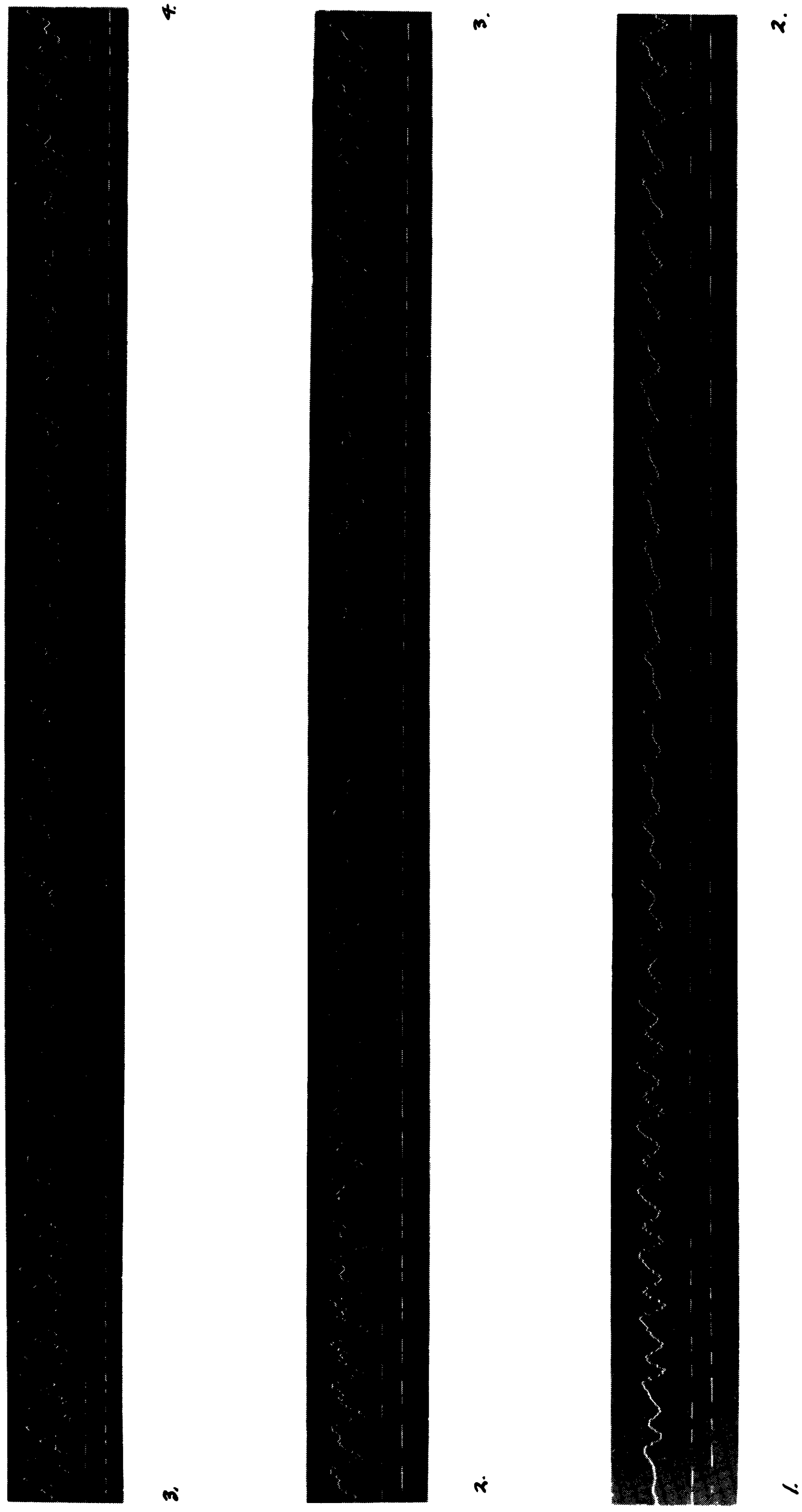
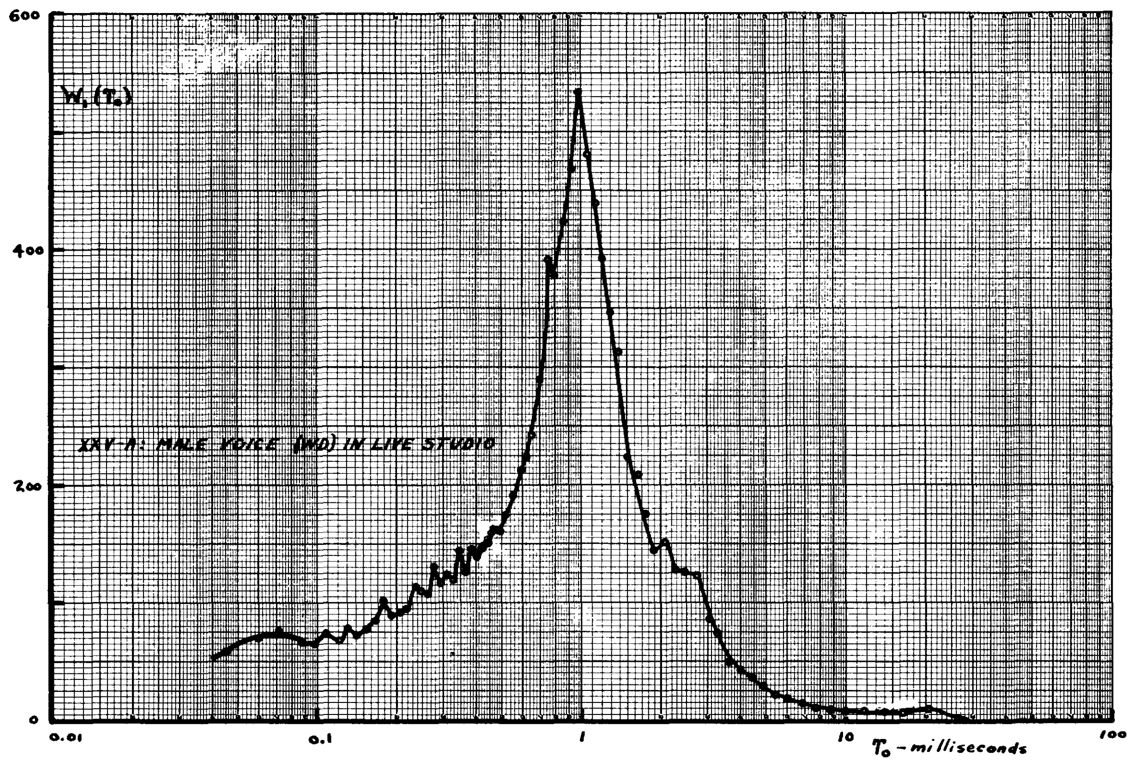
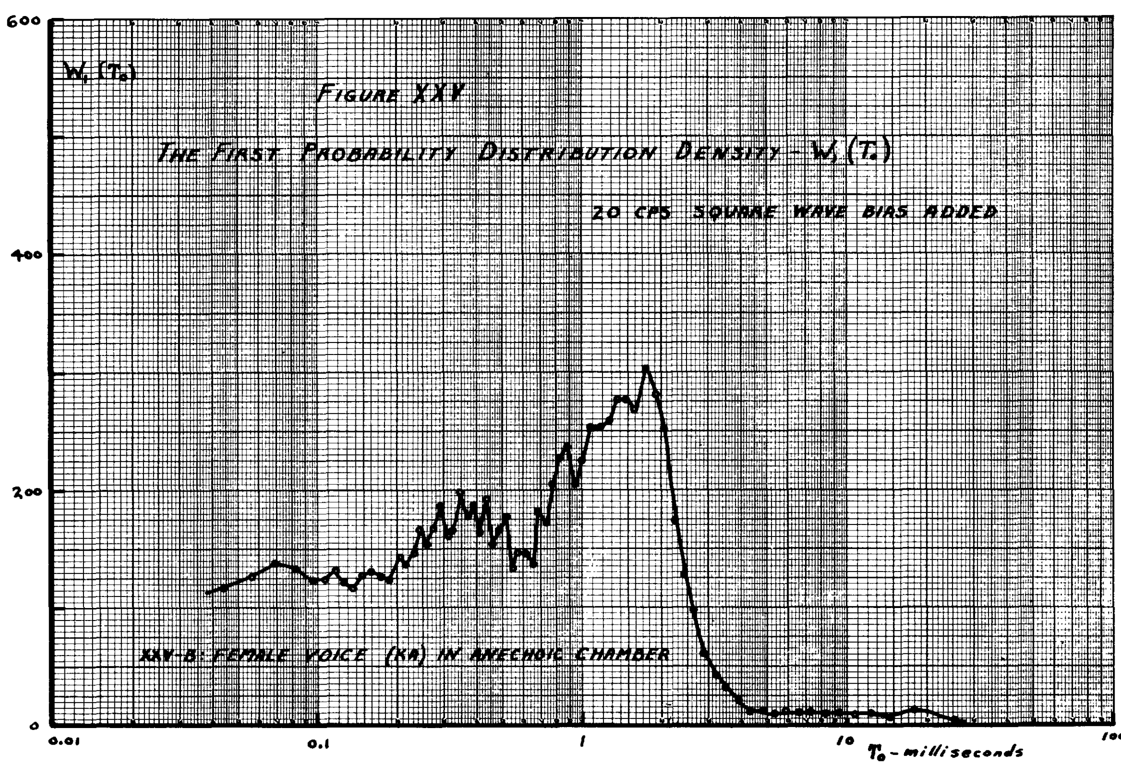


Fig. XXIV

Instantaneous amplitudes of the voice and zero-crossing waves. "The U.S." Male voice (WD) in anechoic chamber.


600
400
200
0
$W_1(T_0)$
XXV-A: MALE VOICE (WD) IN LIVE STUDIO
0.01
0.1
1
10
100
T_0 - milliseconds


600
400
200
0
$W_1(T_0)$
FIGURE XXV
THE FIRST PROBABILITY DISTRIBUTION DENSITY - $W_1(T_0)$
20 CPS SQUARE WAVE BIAS ADDED
XXV-B: FEMALE VOICE (KA) IN ANECHOIC CHAMBER
0.01
0.1
1
10
100
T_0 - milliseconds

density of speech wave zero-crossings, for a voiced sound, closely corresponds to the first formant of the sound spectrogram (i.e. plot of frequency spectrum vs. time).

The effect of environment, and speaker, on the first probability distribution density, $W_1(T_o)$, of the speech wave zero-crossing periods is shown in Fig. XXV. The data for these plots were obtained with the use of a 20-cps square-wave bias signal.

As previously discussed, one main effect of environment is the presence, or absence, of reflected waves. The significant portion of the reflected waves is that corresponding to the voiced sounds in the original wave. Thus the presence of reflections should show up as an increase in the probability of occurrence of zero-crossing periods of the voiced sounds, and consequently as a decrease in the probability of occurrence of the zero-crossing periods of the unvoiced sounds. A comparison of Figs. XXIIIA and XXVA verifies this conclusion.

B. Conditional Probability Distributions

Definitions and methods of measurement

Paralleling the definition of the conditional probability $P(x_1|x_1, \tau)$ for the speech wave instantaneous amplitude, as given in an earlier section, a conditional probability $P(T_1|T_1, m)$ for the voice wave zero-crossing periods may be defined by the equation

$$P(T_1|T_1, m) \equiv P\left[T_1 \leqslant T_o(k) < T_1 + \Delta T_1 \middle| T_1 \leqslant T_o(k+m) < T_1 + \Delta T_1\right] \quad . \quad (105)$$

That is, $P(T_1|T_1, m)$ is the probability that the (k + m)th zero-crossing period of the speech wave has a duration falling in the interval $(T_1, T_1 + \Delta T_1)$ conditional upon the fact that the kth zero-crossing period had a duration falling in the same interval.

A simplified block diagram of the equipment utilized in the measurement of the conditional probability $P(T_1|T_1, m)$ is given by Fig. XXVI. The operation of this equipment, up to the output of the level selector, is the same as that used for the measurement of the probability $P(T_1)$ described earlier. The delay unit is arranged to generate an output pulse "m" zero-crossing periods after the occurrence of a zero-crossing period whose duration falls in the interval $(T_1, T_1 + \Delta T_1)$. The outputs of the level selector and of the delay unit are applied to a coincidence circuit. The output of the coincidence circuit, and hence the reading of counter No. 3, hence corresponds to the joint occurrence of the event $\left[T_1 \leqslant T_o(k) < T_1 + \Delta T_1\right]$ and of the event $\left[T_1 \leqslant T_o(k-m) < T_1 + \Delta T_1\right]$. Similarly, the output of the delay unit, and hence the reading of counter No. 2, corresponds to the occurrence of the event $\left[T_1 \leqslant T_o(k-m) < T_1 + \Delta T_1\right]$. The ratio of the reading of counter No. 3, ν_p, to the reading of counter No. 2, ν, then gives the value of the conditional probability $P(T_1|T_1, m)$ in the limit as the duration of the measurement period increases indefinitely.

Another conditional distribution of interest is the autocorrelation function of the zero-crossing wave. As before, the zero-crossing wave Z(t), may be defined by the

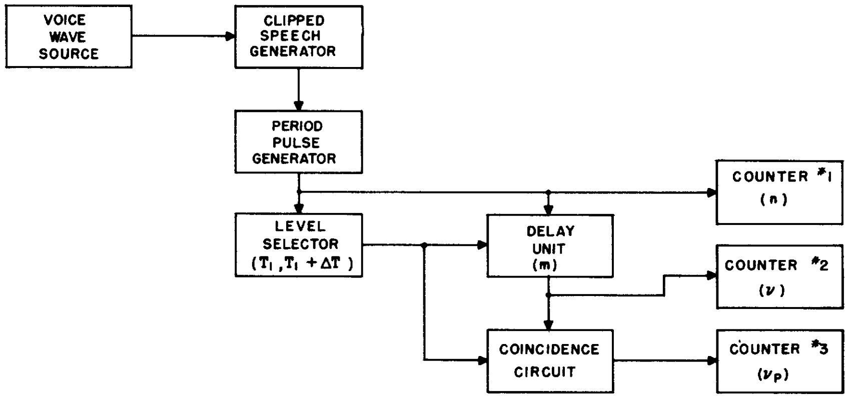


Fig. XXVI Measurement of the conditional probability $P(T_1 | T_1, m)$.

relations

$$Z(t) \equiv \begin{cases} 1 \text{ if } 0 \leqslant x(t) \\ 0 \text{ if } x(t) < 0 \end{cases} \tag{106}$$

where x(t) is the original voice wave. The autocorrelation function of the zero-crossing wave may be expressed by:

$$\phi_Z(t) = \overline{Z(t)\, Z(t+\tau)} \tag{107}$$

as we are dealing with a stationary random phenomenon. Paralleling the development of the autocorrelation function $\phi_\xi(\tau)$ given previously, it may be shown that

$$\phi_Z(\tau) = P\Big[Z(t) = 1 \Big| Z(t+\tau) = 1\Big] \; P\Big[Z(t) = 1\Big] \quad . \tag{108}$$

Then, remembering the definition of the zero-crossing wave, Eq. 106, we see that the autocorrelation function of the zero-crossing wave Z(t), may be expressed in terms of probabilities of the original voice wave, x(t)

$$\phi_Z(\tau) = P\Big[0 \leqslant x(t) \Big| 0 \leqslant x(t+\tau) \; P\Big[0 \leqslant x(t)\Big] \quad . \tag{109}$$

Thus the autocorrelation function of the zero-crossing wave may be determined by a measurement of the conditional probability $P\Big[0 \leqslant x(t) \Big| 0 \leqslant x(t+\tau)\Big]$ and of its limiting value as τ approaches infinity, $P\Big[0 \leqslant x(t)\Big]$.

Reference to the discussion of the measurement of the conditional probability $P(x_1\, x_1, \tau)$ given in an earlier section shows that the conditional probability $P\Big[0 < x(t) \Big| 0 < x(t+\tau)\Big]$ may be measured directly with the apparatus shown in block diagram form by Fig. XIV if the level selector is now adjusted so as to generate an output pulse whenever the event $\Big[0 \leqslant x(t)\Big]$ occurs, rather than whenever the event $\Big[x_1 \leqslant x(t) < x_1 + \Delta x_1\Big]$ occurs. The theoretical discussion of the measurement of the conditional probability $P(x_1 | x_1, \tau)$ applies directly to the present problem. Hence we have

$$P\Big[0 \leqslant x(t) \Big| 0 \leqslant x(t+\tau)\Big] = \lim_{T \to \infty} \; \lim_{n \to \infty} \frac{\nu_p}{\nu} \tag{110}$$

where ν_p is the number of joint occurrences of the event $[0 \leq x(t)]$ and of the event $[0 \leq x(t+\tau)]$, and where ν is the number of occurrences of the event $[0 \leq x(t)]$, both in a measurement of n samples throughout a measurement period of duration T. The autocorrelation function of the zero-crossing wave may thus be determined by measuring the ratio ν_p/ν for a number of values of the delay time, τ.

Experimental result

A practical difficulty arose in the measurement of the conditional probability $[P(T_1|T_1, m)]$ because of the design of the delay unit. As will be discussed later, the delay unit will permit only binary values of the number, m, of periods delay when used as shown in Fig. XXVI. That is, m may have only the values 1, 2, 4, 8, 16, 32, 64, and 128. Thus no data could be obtained as to the value of $P(T_1|T_1, m)$ for nonbinary values of m with the present equipment. Also an error is introduced into the measurement of $P(T_1|T_1, m)$ because of a "lost-pulse" effect discussed later. For these reasons, the experimental results are uncertain in their meaning and in their accuracy and no detailed discussion will be made of them. The only significant results obtained so far seem to be that the conditional probability $P(T_1|T_1, m)$ is generally smaller for m = 1 than it is for m = 2; and also that $P(T_1|T_1, m)$ does not appear to reach its final value until the number of periods of delay, m, has increased to about 64 or more. These results were obtained from studies made of a male voice (WD) in an anechoic chamber and in a live studio, and for a female voice (KA) in an anechoic chamber.

Measurements were made of the autocorrelation function, $\phi_Z(\tau)$, of the zero-crossing wave, Z(t), for two voices (one female, KA, and one male, WD) studied in an anechoic chamber. The results of these measurements are plotted on both linear and semi-logarithmic graph paper in Figs. XXVII and XXVIII. It is of interest to compare these plots with the plots of the conditional probability, $P(x_1|x_1, \tau)$, given earlier. Such a comparison shows that $\phi_Z(\tau)$ and $P(x_1|x_1, \tau)$ have similar characters. However, $\phi_Z(\tau)$ shows much greater variation than does $P(x_1|x_1, \tau)$ for values of τ greater than one millisecond.

III EQUIPMENT

A. General Requirements

One of the objectives of this study was the design and construction of equipment suitable for the determination of probabilities for random time functions. Although the detailed characteristics of the various probabilities could not be known beforehand, certain a priori equipment characteristics could be specified.

As was pointed out in sec. I, the decision was made at the start that it would be sufficient for the purposes of this investigation to design and construct apparatus suitable for the study of a specific random process, the voice wave. However, the resultant apparatus may also be used to study any random time function whose statistics

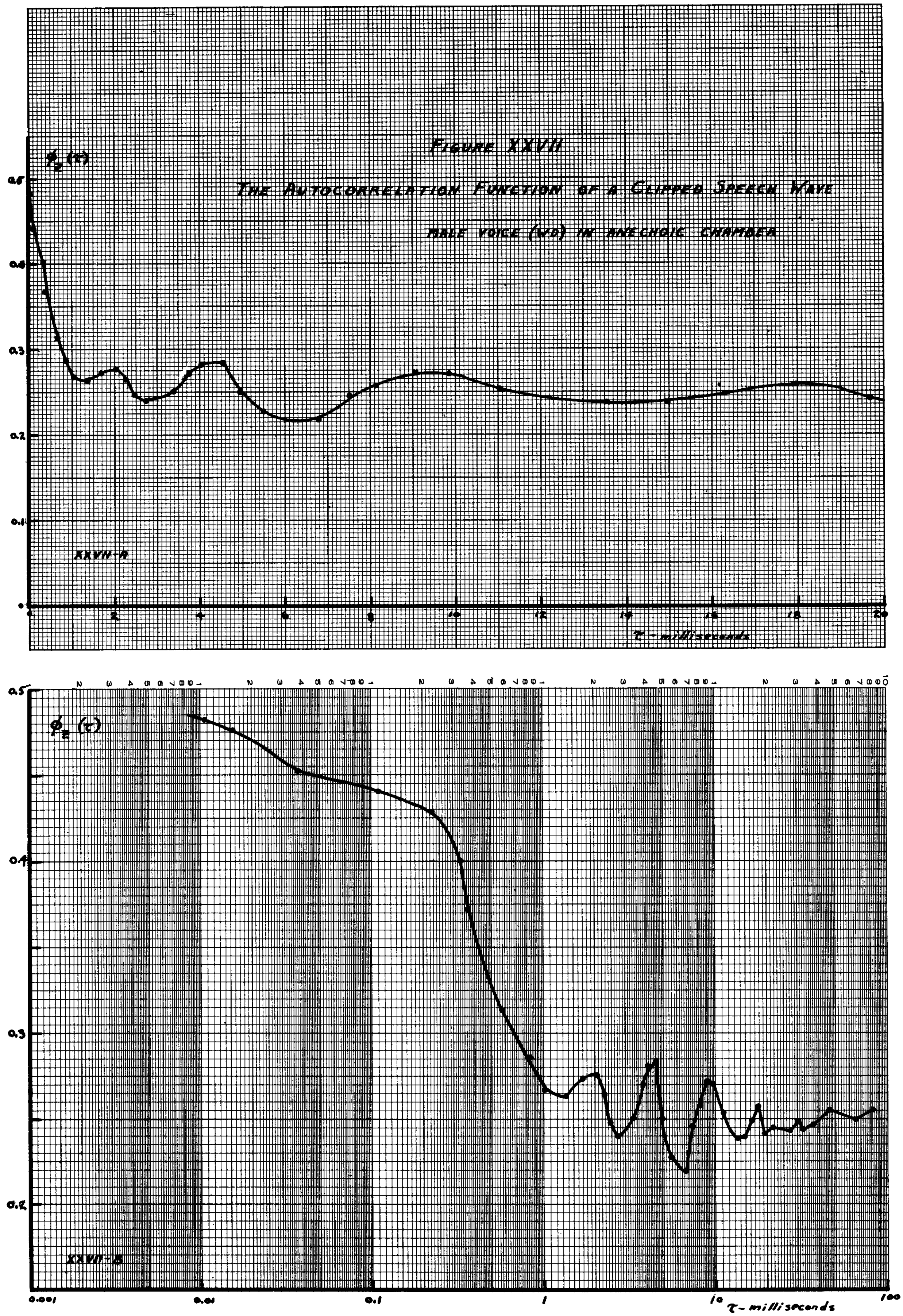


FIGURE XXVII
THE AUTOCORRELATION FUNCTION OF A CLIPPED SPEECH WAVE
MALE VOICE (WD) IN ANECHOIC CHAMBER
$\phi_z(\tau)$
XXVII-A
τ - milliseconds
$\phi_z(\tau)$
XXVII-B
τ - milliseconds

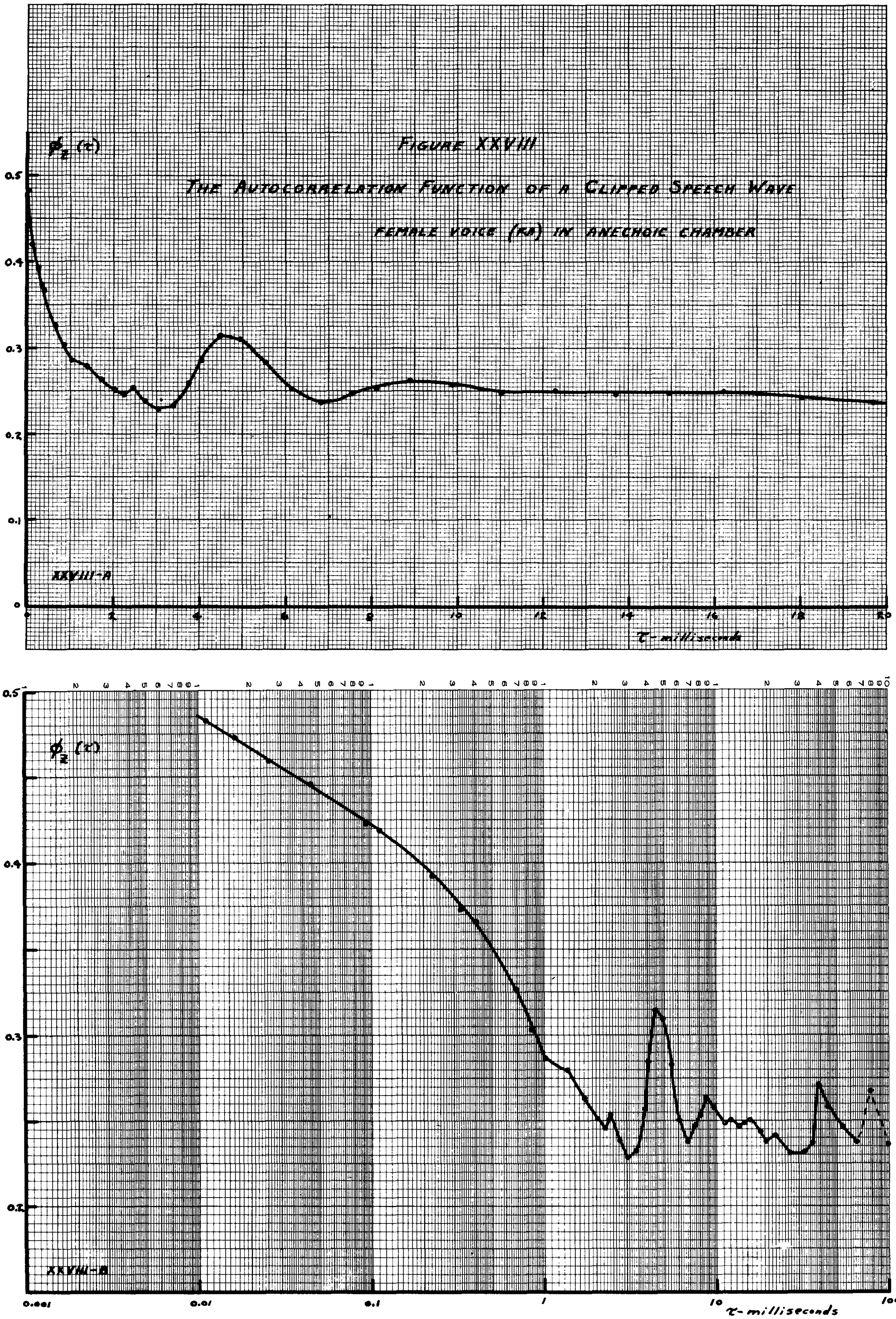


FIGURE XXVIII
THE AUTOCORRELATION FUNCTION OF A CLIPPED SPEECH WAVE
FEMALE VOICE (RA) IN ANECHOIC CHAMBER
XXVIII-A
XXVIII-B
τ - milliseconds

fall within the range of the voice statistics. The justification for the decision to tailor the equipment characteristics to the study of a specific random process is purely economical. For the same reason, it was decided to limit the apparatus to the making of one measurement at a time rather than a number of measurements simultaneously. The fact that the equipment is able to make only one measurement at a time requires the data-taking time to be about one working day per distribution. As discussed earlier, this requires practically that the measurements be made from a recorded voice wave.

A review of the Method of Measurement sections shows that two probabilities are basic to our study: the first probability $P\left[\xi_1(t) = 1\right]$, and the conditional probability $P\left[\xi_1(t) = 1 \middle| \xi_2(t+\tau) = 1\right]$, where ξ_1 and ξ_2 are appropriate random variables derived from the speech wave. For example, let us suppose that ξ_1 is defined by the equation

$$\xi_1(t) = \begin{cases} 1 \text{ if } x_1 \leqslant x(t) < x_1 + \Delta x_1 \\ 0 \text{ otherwise} \end{cases} \tag{111}$$

where x(t) is the speech wave. Then, as was shown in sec. I-A, the probability $P\left[\xi_1(t) = 1\right]$, when divided by Δx_1, gives the value of the first probability distribution density $W_1(x_1)$ for the speech wave instantaneous amplitude. Thus the basic requirements upon our equipment are that it be able to generate the appropriate derived random variables ξ_1 and ξ_2, and that it be able to measure the probabilities $P\left[\xi_1(t) = 1\right]$ and $P\left[\xi_1(t) = 1 \middle| \xi_2(t+\tau) = 1\right]$ related to those random variables.

An estimation of the required equipment accuracy may be made as follows: Our statistical accuracy discussion showed that a sampling error is to be expected, even with perfect measuring equipment, if the measurements are made over a finite interval of time. Assuming that the sampling errors due to the finite measurement time interval and the errors due to nonperfect equipment are statistically independent, the mean-square values of these errors are additive. A requirement may be established that the total root-mean-square error shall not exceed the sampling error by more than ten percent. Then assuming a minimum expected RMS sampling error of five percent, the above requirement may be met by requiring that the expected RMS equipment error should not exceed approximately 2.2 percent. The equipment accuracy must be maintained at least over the time required to measure a single distribution (i.e. at least over one working day).

B. Audio Equipment Characteristics

Since it was necessary to record the speech waves studied in this investigation, wide range audio equipment had to be used to preserve the voice waveform. Rather than design and construct the necessary audio equipment, it was decided to make use of available apparatus. The pertinent characteristics of the audio equipment used in this study are given as follows.

The microphone chosen was of the ribbon variety: the RCA velocity microphone, type 44-BX. The curve of frequency response to an incident plane sound wave given by the manufacturer (18) for this microphone is flat to within $\pm$ 2.5 db over the range of 40-12,000 cps when the microphone is used with the "music" connection. Tests made at the M.I.T. Acoustics Laboratory on various microphones of this type check these results reasonably well. It is necessary to have a substantially plane wave at the microphone in order to avoid the low-frequency boost phenomenon characteristic of a ribbon microphone when responding to spherical waves. For this purpose, the microphone was maintained at a distance of at least one meter from the speaker's lips. Under these conditions, less than 3 db boost is obtained for frequencies above about 55 cps.

The recorder used in this study was a magnetic tape recorder, the Magnecorder, type PT-6. Its measured overall frequency response was flat to within $\pm$ 2.5 db over the range of 20-15,500 cps. The maximum obtainable root-mean-square signal-to-noise ratio was about 47 db.

In order to study voice waves in a substantially reflection-free environment, certain of the speech waves were recorded with the above equipment in the anechoic chamber of the M.I.T. Acoustics Laboratory. This chamber was designed in accordance with the results given by Beranek and Sleeper (19). The measured lower cutoff frequency of this chamber is 70 cps. This cutoff frequency is defined as the frequency above which less than one percent of the energy incident upon the walls is reflected.

In order to study the effect of reflections on the voice wave, certain of the speech waves studied in this investigation were recorded in the experimental studio at the M.I.T. Acoustics Laboratory. This studio is rather "live". The reverberation time, as measured with 50-cycle bands of random noise, rises from about a half-second at 100 cps to about one second at 4000 cps (20). The form of this reverberation time characteristic is of relatively little significance to our studies as we were concerned only qualitatively with the effect of reflections. The estimated average value over the audio frequency range of the isolation of the studio from its surroundings is of the order of 45 db.

The characteristics of the various items of audio equipment as described above were considered adequate for the purposes of this investigation.

C. Amplitude-Distribution Equipment

Sections I and II discuss certain probability distributions, relating to the voice wave instantaneous amplitude, which were investigated in this study: the first probability distribution density $W_1(x)$, the conditional probability $P(x_1|x_1, \tau)$, and the conditional probability distribution density $W_2(x_1|x, \tau)$. These probability distributions can be measured by the system shown in block diagram form by Fig. XXIX. The amplitude distribution equipment is composed of five main units: a recorder, a pulse amplitude modulator, the level selectors, a conditional probability delay unit, and the counters. Of these units, the pulse amplitude modulator, the level selectors, and the conditional probability delay unit had to be designed and constructed specifically for this study. The recorder and the counters were chosen from available equipment.

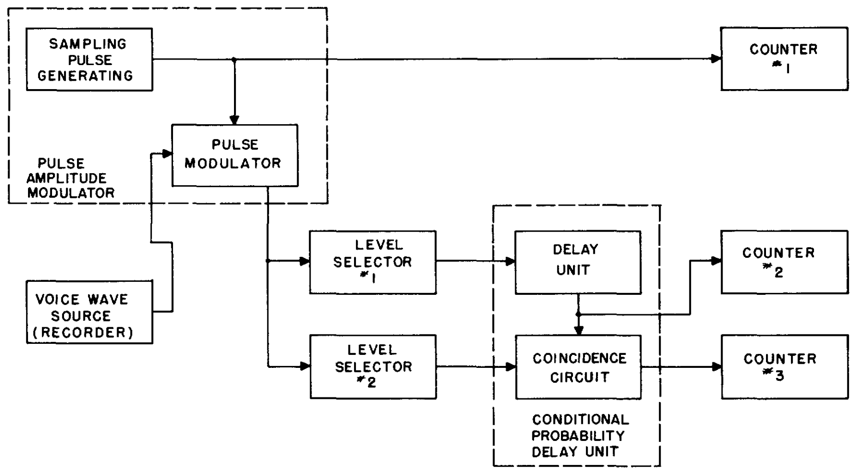


Fig. XXIX Amplitude distribution equipment.

Design requirements of the pulse amplitude modulator

The purpose of the pulse amplitude modulator is to convert the voice wave into a sequence of constant-duration repetitive pulses whose amplitudes are linearly related to the amplitudes of the voice wave at the instants of sampling. The requirements upon the sampling pulse repetition period may be obtained as follows: It was shown in the discussion of statistical accuracy that the greater the number of sampling pulses in a given measurement interval, the smaller is the expected mean square error. Thus, generally, the sampling pulse repetition period should be made as short as possible.

It should be emphasized, however, that there is no unique relation between the sampling rate and the highest frequency component of the wave being studied in the present problem of a measurement of a probability, as there is in the problem of pulse communication systems. That is, while there is a constant minimum usable sampling rate in a pulse communication system regardless of the total message length (4), the sampling rate for a probability measurement, with a specified accuracy, may be reduced as the total duration of the measurement interval is increased.

A second criterion for the sampling repetition period is imposed by the requirements of the conditional probability delay unit. As will be discussed later, the conditional probability delay unit requires that the sampling repetition periods be as long as one millisecond. Combining the above requirements, it was decided that the sampling repetition period should be variable from about 5 μsec to 1 msec. The choice of 5 μsec as a lower limit is a compromise choice resulting from the fact that, as the sampling repetition period becomes shorter, the equipment becomes more complex.

The maximum sampling pulse duration is determined by the method of operation of the level selector. The level selector generates an output pulse whenever the amplitude of its input pulse, $X(t)$, falls in a specified interval $(X_1, X_1 + \Delta X_1)$. Then the maximum sampling pulse duration should be such that the voice wave cannot cause the pulse instantaneous amplitude to change by more than ΔX_1, during the pulse. Let us assume that the voice wave has a maximum frequency, f_{max}. An estimate of the most severe case may then be made by assuming that the wave being studied is a sine wave, of frequency f_{max}, and that the sine wave amplitude equals the maximum amplitude that can be studied. That is, the input wave x(t) is given by

$$x(t) = \frac{N \Delta x_1 \sin \omega_{max} t}{2} \tag{112}$$

where N is the number of adjacent amplitude intervals of width Δx_1 that can be studied. The maximum slope of this wave is $(N \Delta x_1 \omega_{max})/2$. The requirement on pulse duration then becomes

$$\frac{N \Delta x_1 \omega_{max} \delta}{2} \leqslant \Delta x_1 \tag{113}$$

where δ is the sampling pulse duration. Solving for δ, we obtain the inequality

$$\delta \leqslant \frac{1}{\pi N f_{max}} \quad . \tag{114}$$

Choosing a sampling pulse duration of 0.5 µsec, we then see that f_{max} may be 6.4 kcps if N equals 100, or f_{max} may be 12.8 kcps if N equals 50. On this basis, it was decided that a one-half microsecond sampling pulse duration would be suitable for the purposes of this study.

The required amplitude range of variation of the amplitude modulated output pulse is determined by the characteristics of the level selectors. As discussed later, the amplitude interval of the level selectors is about one volt. If we then require that the system be able to investigate up to one hundred adjacent amplitude intervals, the required range of variation of the output pulses becomes 100 volts.

Design and operation details

The design of an equipment capable of meeting the above requirements is reasonably straight-forward and use may be made of well known pulse-circuit design techniques (21). The schematic circuit diagram of such an equipment is shown in Fig. XXX.

The operation of the pulse amplitude modulator is as follows: The sampling pulse repetition period is determined by the free-running repetition period of the plate-coupled multivibrator, V1. The repetition period of this multivibrator may be varied from 5 µsec to about 1.3 msec by changing the grid-circuit time constants. The waveform at the plate of V1A is inverted by the amplifier V2, and is peaked by a short-time constant RC coupling network so as to form pulses suitable for triggering the sampling pulse generator, V3. V3 is a cathode-coupled, "one-shot", multivibrator. It generates a constant-amplitude, fixed-duration output pulse each time it is triggered. The duration of the output pulse may be fixed at any value between 0.4 and 1.0 µsec by appropriately setting the pulse width control. The output pulses of the sampling pulse generator are applied, through the cathode-follower V4, to the suppressor grid of the Sampler, V6, and also to the

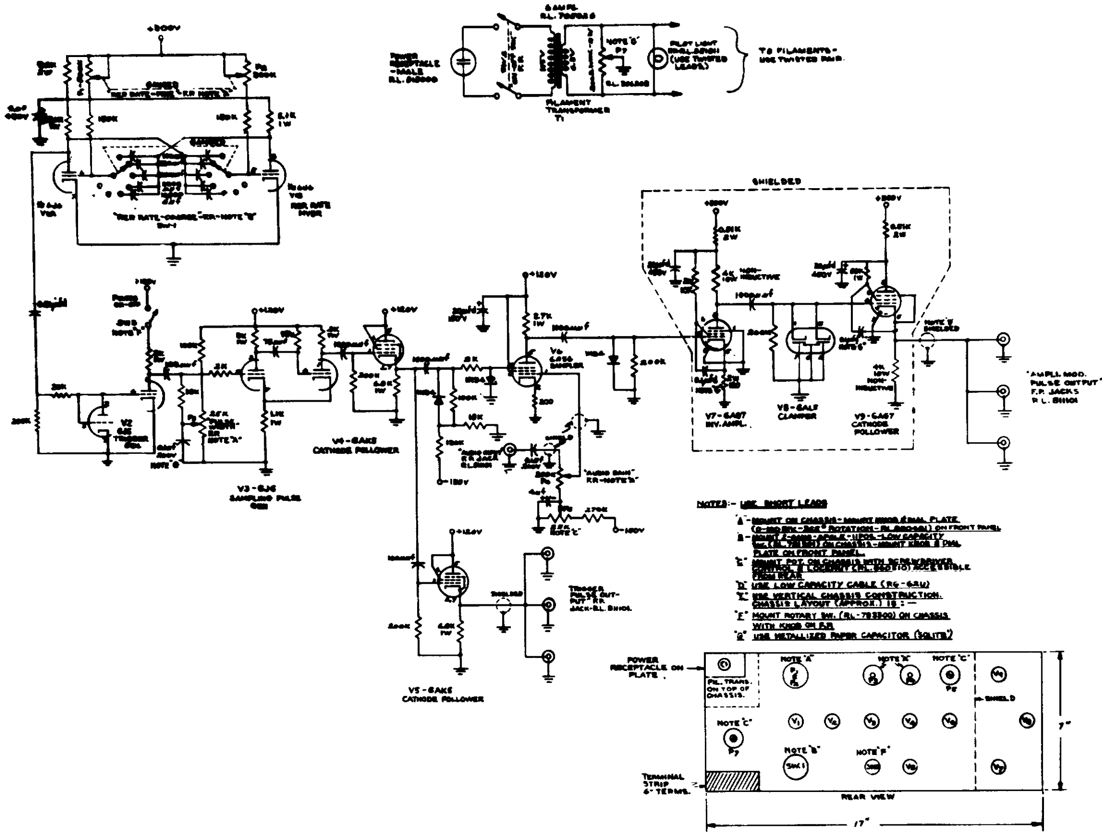


Fig. XXX Pulse amplitude modulator.

trigger pulse output jacks through the cathode follower, V5.

The voice wave is applied to the control grid of the Sampler, V6. The suppressor grid of this tube is normally biased beyond cutoff, hence no plate current may flow until the suppressor grid voltage is raised above cutoff by the sampling pulses. The sampling pulses are constant amplitude pulses, hence the amplitude of the sampler output pulses is determined by the amplitude of the voice wave during the sampling pulse. The sampler output pulses are amplified and inverted by V7, and are applied to the amplitude modulated pulse output jacks through the cathode follower V9. These output pulses are thus positive pulses whose amplitude may be varied linearly (from about + 20 volts to + 140 volts) by the control grid voltage of the sampler. The quiescent value of the control grid bias of the sampler is adjusted so that, with no voice wave applied, the amplitude of the output pulse lies in the center of its linear range of variation.

Design requirements of the level selector

The purpose of the level selector, or amplitude interval discriminator, is to generate an output pulse whenever the amplitude of its input pulse, X(t), falls within a specified amplitude interval, $(X_1, X_1 + \Delta X_1)$. The basic form of the level selector, shown by Fig. XXXI, consists of two amplitude discriminators and an anticoincidence circuit.

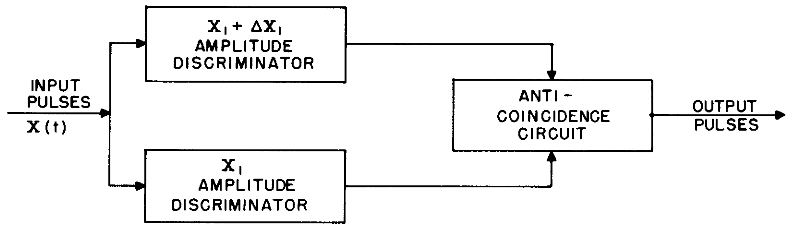


Fig. XXXI The level selector.

Ideally, an amplitude discriminator is a device whose output is identically zero for input amplitudes less than a specified constant value, and nonzero for an input amplitude greater than that discrimination amplitude. Thus if the event $[X(t) < X_1]$ occurs, neither amplitude discriminator generates an output, and no level selector output is obtained. If the event $[X_1 \leqslant X(t) < X_1 + \Delta X_1]$ occurs, the X_1 amplitude discriminator output is nonzero, while the $(X_1 + \Delta X_1)$ amplitude discriminator output is zero. Hence a level selector output is obtained. On the other hand, if the event $|X_1 + \Delta X_1 \leqslant X(t)|$ occurs, then both amplitude discriminators produce outputs which are mutually cancelled by the anticoincidence circuit, hence no level selector output is obtained.

Amplitude discriminators have previously been designed for voice wave studies (14, 15), for radar systems (21), and for nuclear physics studies (22). Also, level selectors have previously been designed for nuclear physics studies (22). Hence the basic design of a level selector is more or less straight-forward. However, the particular requirements placed upon the level selector by this study required a new design.

The design of a level selector is determined principally by the given characteristics of the input pulses, and by the specification of the amplitude interval width, ΔX_1. The principal characteristics of the input pulses that affect the operation of the level selector are the shape and the repetition period. Ideally, the input pulses are rectangular pulses whose height is determined by the amplitude of the voice wave at the time of sampling. However, these pulses pass through various circuits and waveform distortion invariably occurs. If the input pulse has sloping sides, any input pulse, whose peak amplitude is greater than the upper limit of the amplitude interval being observed, will take a short, but nonzero, interval of time to traverse the observation interval. The anticoincidence circuit must then be designed so that no level selector output is obtained from this sloping wavefront phenomenon.

The design of the level selector is affected by the repetition period of the input pulses, as the resolution time of the level selector must be shorter than the shortest input pulse repetition period used. That is, the level selector must be able to respond accurately to every input pulse, regardless of the time intervals between pulses.

The specification of the amplitude interval width, ΔX_1, is determined mainly by the desired equipment accuracy. Equipment measurement errors are introduced because of the fact that the discrimination amplitudes of physical devices cannot be maintained absolutely constant. In order to reduce the errors due to this phenomenon, the amplitude interval width should be made large as compared to the range of variation of the discrimination amplitude. However, the greater the width of the amplitude interval ΔX_1, the greater must be the range of amplitude variation of the input pulses, and some compromise must be made.

Design and operation details

A level selector capable of meeting the aforementioned design requirements is shown in schematic diagram form by Fig. XXXII. Idealized waveforms relating to this equipment are given in Fig. XXXIII for the two cases of greatest interest: an input pulse whose amplitude falls in the interval $(X_1 + \Delta X_1)$, and an input pulse whose amplitude lies above that interval. The detailed operation of this unit is discussed below.

The amplitude discriminators used in the level selector developed for this investigation are the lower amplitude discriminator composed of tubes V5, V7 and V9, and the upper amplitude discriminator composed of tubes V6, V8, and V10. Basically these amplitude discriminators are cathode-coupled, one-shot multivibrators, V7 and V8, modified by the addition of compensation tubes, V9 and V10, and by the addition of level set tubes, V5 and V6. The compensation and level set tubes are added to improve the resolution time and the constancy of the discrimination amplitude. With such an arrangement the resoltuion time is equal to the duration of the pulse generated by the upper level multivibrator, V8, plus about one microsecond, and the range of variation of discrimination amplitude is limited to about one-half volt. By setting the discrimination amplitude of the upper level multivibrator 10 volts higher than the discrimination amplitude of the lower level multivibrator, the range of variation of discrimination amplitude, at each end of the desired amplitude interval, is limited to about 5 percent of the amplitude interval width.

With a 10-volt amplitude interval required by the amplitude discriminators and a requirement of being able to study 100 adjacent intervals, the voice amplitude modulated pulses would have to have a range of amplitude variation of 1000 volts if applied directly to the amplitude discriminators discussed above. Such a difficulty is avoided by the insertion of a two-stage slicer, V2 and V3, between the amplitude discriminators and the input voice amplitude modulated pulses. A "slicer" is composed of a cathode follower driving a grounded-grid triode amplifier. If the slicer input voltage is any value below a certain voltage, E_1, the cathode follower is cutoff, and the output voltage is constant at a value dependent upon the design of the grounded-grid amplifier. If the slicer input voltage is any value above a certain other voltage, E_2 and E_1, then the output voltage of the cathode follower is sufficient to cutoff the grounded-grid amplifier. Hence the slicer output voltage is constant at the value equal to the plate supply voltage. If the slicer voltage is between the limits E_1 and E_2, the slicer acts as a linear amplifier. The advantages of a slicer over a conventional amplifier are the described limiting action and the high inherent stability due to the large amount of degenerative feedback on each tube. The two-stage slicer, V2 and V3, is adjusted so that the voltage gain

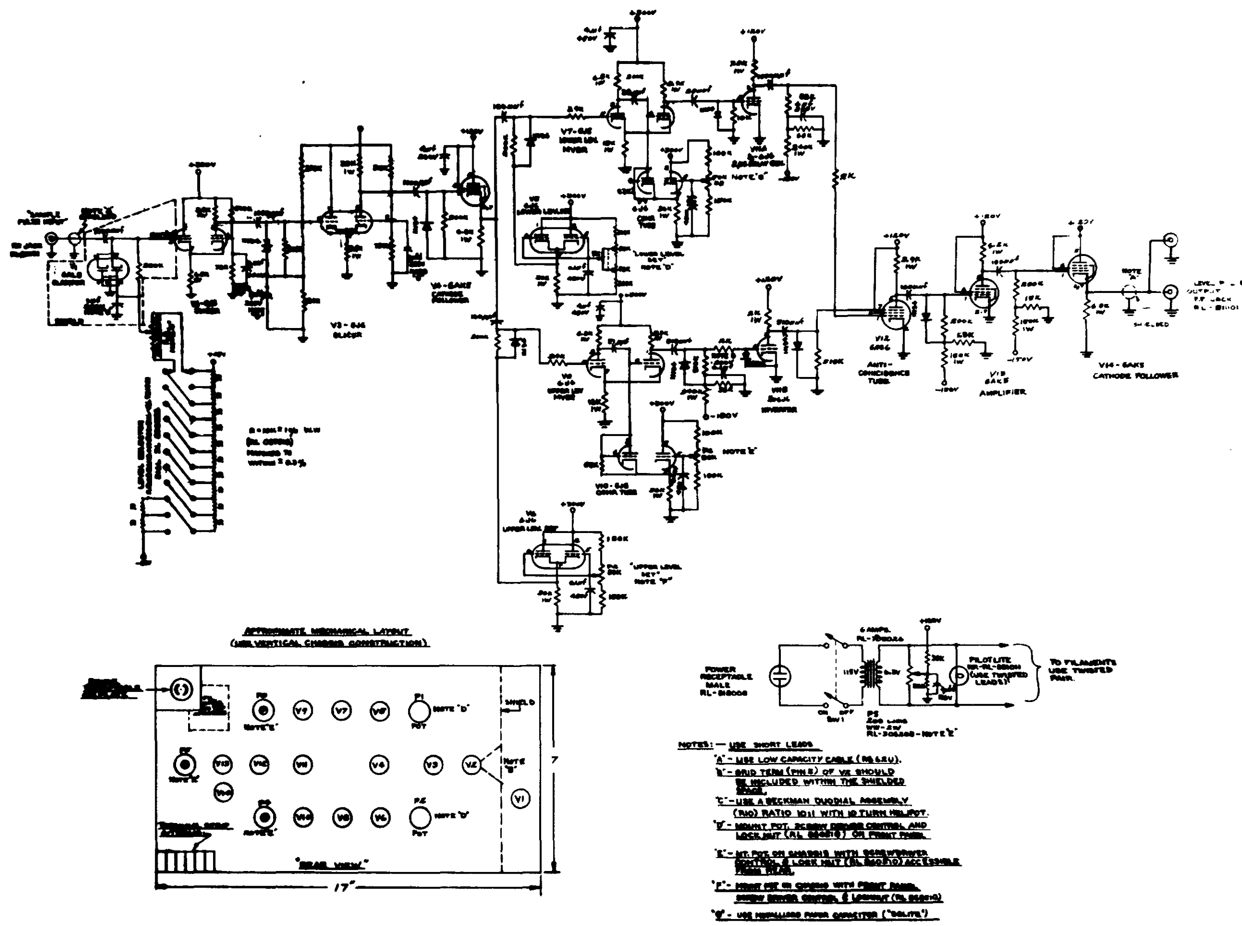


Fig. XXXII The level selector.

of the combination (including the cathode follower, V4) is 10, and the peak-to-peak amplitude swing at the output is about 30 volts. The amplitude discriminators are then adjusted so that the required 10-volt amplitude interval corresponds to the center portion of the region of the slicer output. Thus, the corresponding amplitude interval at the input to the slicer has a width of one volt. The required range of variation of amplitude of the input voice modulated pulses is hence 100 volts.

The amplitude discriminators are designed to give an indication of whether or not the positive peak of the pulse occurring at the input to the first slicer lies between two fixed voltage limits, set one volt apart. This fixed voltage interval may be adjusted to correspond to any arbitrarily located input pulse amplitude interval by establishing a suitable reference voltage level for the bases of the input pulses. This is accomplished by means of the clamper, V1, and the level selector potentiometer. Thus the voltage level of the base line of the pulses at the input to the slicer (E_{ref} in Fig. XXXIII-A) is established by the setting of the level selector potentiometer. Because of precision requirements, the level selector potentiometer was composed of wire-wound elements. It was also found necessary to supply this potentiometer from a set of batteries, as the

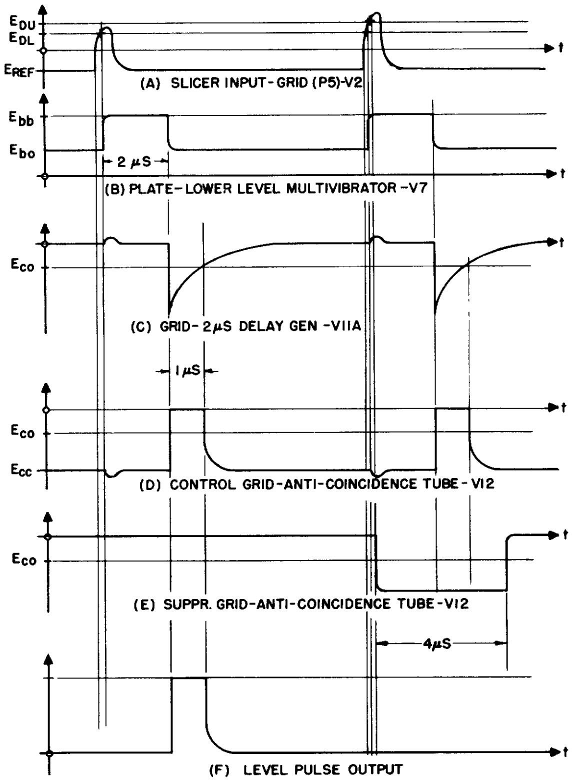


Fig. XXXIII Level selector idealized waveforms.

available regulated power supplies did not have adequate stability of output voltage over the period of a working day.

When the instantaneous amplitude of the slicer input pulse exceeds the value corresponding to the discrimination amplitude of the lower level multivibrator (E_{DL} in Fig. XXXIII-A), the lower level multivibrator generates a rectangular output pulse approximately 2 μsec long (Fig. XXXIII-B). The output of the lower level multivibrator is then peaked and clipped by the grid-coupling circuit of the two-microsecond delay generator V11A. The waveform at the grid of V11A (Fig. XXXIII-C) then consists of a small positive pulse occurring at the start of the lower level multivibrator output pulse, and a large negative pulse occurring at the end of the lower level multivibrator output pulse. The waveform at the plate of V11A thus consists mainly of a positive pulse (about one microsecond long) occurring at the end of the lower level multivibrator output pulse. The output of V11A is applied to the control grid of the anticoincidence tube, V12. The control grid of V12 is normally biased below cutoff, and is raised to about zero volts (Fig. XXXIII-D) by the output of V11A.

When the instantaneous amplitude of the slicer input pulse exceeds the value corresponding to the discrimination amplitude of the upper level multivibrator (E_{DU} in Fig. XXXIII-A), the upper level multivibrator generates a rectangular output pulse approximately 4 μsec long. This pulse is inverted by V11B and is applied to the suppressor grid of the anticoincidence tube, V12. The suppressor grid of V12 is normally biased to zero volts and if an upper level multivibrator output pulse occurs, the suppressor grid of V12 is driven below cutoff (Fig. XXXIII-E) for about 4 μsec.

We then have three cases of interest: (a) If the slicer input pulse amplitude is less than E_{DL}, then neither level multivibrator is triggered and the control grid of the anticoincidence tube V12 is maintained below cutoff. (b) If the slicer input pulse amplitude lies between E_{DL} and E_{DU}, both the control and suppressor grids of V12 are simultaneously above cutoff momentarily and a negative pulse appears at the output of V12. (c) If the slicer input pulse amplitude is greater than E_{DU}, the suppressor grid of V12 is driven below cutoff when its control grid is driven above cutoff and V12 is maintained in a cutoff condition. The length of the output pulse of the upper level multivibrator is made long enough to cancel the positive pulse on the control grid completely, regardless of the time required for the slicer input pulse to traverse the amplitude interval (E_{DL}, E_{DU}). Thus only if case (b) occurs, does a pulse occur at the output of V12. The output pulse of V12 is amplified and inverted by V12, and supplied to the level pulse output jack through the cathode follower, V14, as a positive, 25-volt pulse about 1 μsec long. (Fig. XXXIII-F).

Design requirements of the conditional probability delay unit

Reference to Fig. XXIX shows that the conditional probability delay unit is composed of two main parts: a delay unit and a coincidence circuit. The purpose of the conditional probability delay unit is to delay one of its two inputs by a time, τ, and to determine the coincidence, or lack of coincidence, between the delayed input and its other input.

The required range of variation of the delay time, τ, is determined by the characteristics of the random phenomenon being studied. From photographs of the instantaneous waveforms of various voiced sounds, it was determined that the time duration of these sounds was of the order of magnitude of 0.1 sec. On this basis it was estimated that significant variations in the various conditional probabilities might possibly be obtained for values of τ out to about 0.1 second. The experimental results given earlier confirm this estimate. The minimum required value of τ is theoretically zero, but it was felt that a practical lower limit would be between 5 μsec and 10 μsec.

The extent of the required range of variation of τ immediately makes it practically impossible to use a transmission line type of delay unit. These units have been built with some success for delays of the order of 1 msec, but become extremely unwieldy for delays of the order of 0.1 second. The longer delays may easily be obtained by using some such system as a moving magnetic type with displaced pickup heads. However this system becomes mechanically unwieldy for the very short values of τ. Rather than attempt to combine these methods, it was decided to use pulse counting techniques. The method chosen involves the formation of delay by counting a specified number of sampling pulse repetition periods. The magnitude of the delay time may then be varied by changing the duration of the sampling pulse repetition period and by changing the number of repetition periods counted. The system developed, and explained below, is believed to be new and to be one of the simplest inherently accurate systems capable of covering the desired range of variation of the time delay, τ.

Design and operation

A conditional probability delay unit capable of meeting the aforementioned requirements is shown in schematic diagram form by Fig. XXXIV. Idealized waveforms relating to this equipment are shown in Fig. XXXV. These waveforms are for the particular case of four periods delay.

In order to describe the operation of this unit, let us suppose that the system has been reset (i.e. there are no pulses in the counters). In this condition, the left half of

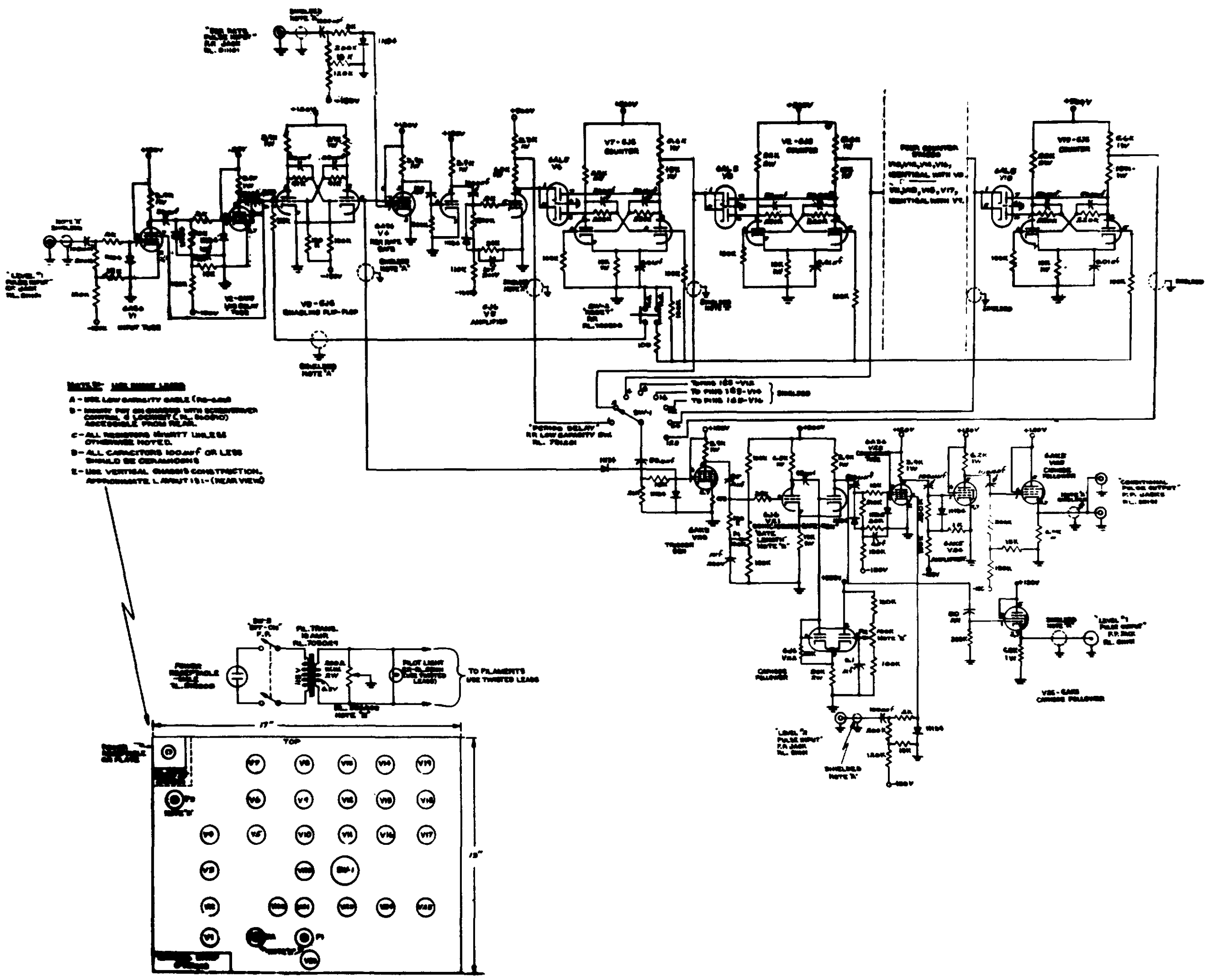


Fig. XXXIV Conditional probability delay unit.

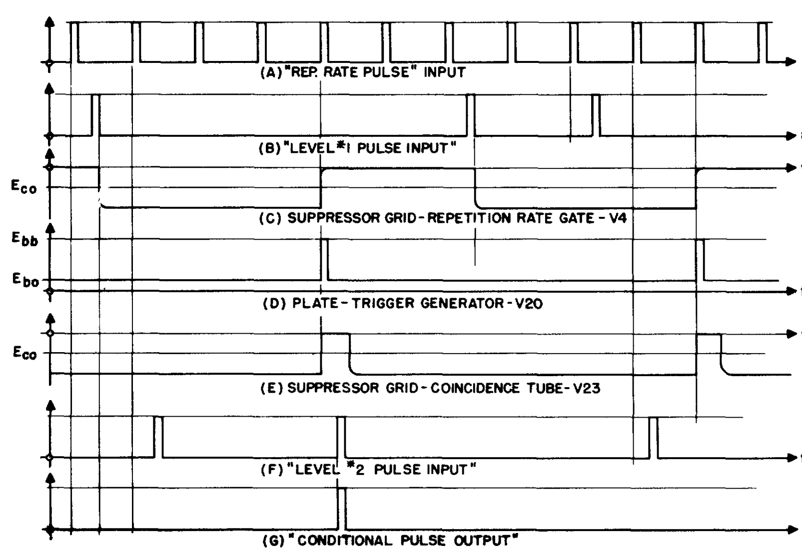


Fig. XXXV Conditional probability delay unit idealized waveform, 4 periods delay.

the enabling flip-flop, V3, is conducting, and the right half is cutoff. Thus the grid voltage of the left half of V3 is slightly above zero volts, while the grid voltage of the right half is below the cutoff voltage. Consequently, the suppressor grid voltage of the input tube, V1, is slightly above zero volts and this tube is able to conduct plate current. Also, the suppressor grid voltage of the repetition rate gate, V4, is below its cutoff value, and the plate current of this tube is cutoff. Let us suppose now that a pulse occurs at the output of level selector No. 1, and is shown by Fig. XXXV-B. This pulse will be amplified and inverted by input tube V1. The negative pulse at the plate of V1 will be peaked and clipped by the grid coupling circuit of the one-microsecond delay tube V2. The grid voltage at V2 will consist of a small negative pulse corresponding to the start of the input pulse, and a large positive pulse appearing at the end of the input pulse. Thus a large negative pulse will appear at the plate of V2 at a time corresponding to the end of the input pulse.

The negative pulse at the plate of V2 will cause the enabling flip-flop to flip, thus causing the suppressor grid of input tube V1 to be driven below cutoff, and causing the suppressor grid of the repetition rate gate, V4, to be driven upwards to about zero volts (Fig. XXXV-C). Thus the input tube will not respond further to level No. 1 pulses, but the repetition rate gate will now respond to the sampling pulses.

When the suppressor grid of the repetition rate gate V4 is above cutoff, the inverted sampling pulses will appear at the plate of V4. These negative pulses are amplified by the two-stage amplifier V5. The amplified negative pulses are then applied to a seven-stage binary counter, V7 through V19. The output waveform of the first counter, V7, is driven negative by the second sampling pulse after the occurrence of the level No. 1 pulse occurring at the grid of the input tube. Similarly, the output waveform of the second counter, V9, is driven negative by the fourth sampling pulse after the level No. 1 input pulse. Finally, the output waveform of the last counter, V19, is driven negative by the 128th sampling pulse after the occurrence of the level No. 1 input pulse. Depending upon the position of the periods delay switch, a negative pulse appears at the grid of the trigger tube, V20, corresponding to the first, second, fourth, . . . , or one hundred and twenty-eighth sampling pulse after the occurrence of the level No. 1 input pulse.

The negative pulse appearing at the grid of the triggering tube, is also applied to the right hand grid of the enabling flip-flop, V3, causing that tube to flop back to its original state, turning on the input tube, V1, and turning off the repetition rate gate, V4. In addition, the negative pulse at the grid of the triggering tube, V20, is inverted and amplified by that tube (Fig. XXXV-D) and used to trigger the coincidence gate generator, V21. The output of V21 is a positive pulse applied to the suppressor grid of the coincidence tube, V23 (Fig. XXXVII-E). This pulse also is applied, through the cathode follower, V26, to the level No. 1 pulse output jacks.

Normally, the suppressor grid voltage of the coincidence tube is maintained below the cutoff value. Hence, even though the level selector No. 2 output pulses (Fig. XXXV-F) are applied to the control grid of the coincidence tube, no output is obtained from the coincidence unless there is a simultaneous occurrence of a gate pulse at the suppressor grid and a level No. 2 pulse at the control grid. If such a coincidence occurs, a negative pulse appears at the plate of the coincidence tube, V23. This pulse is inverted and amplified by V24, and is applied to the conditional pulse output jacks, through the cathode follower, V25, as a positive rectangular pulse of about 25 volts amplitude and 1 μsec duration (Fig. XXXV-G).

The occurrence of a conditional pulse output at a time, t, thus corresponds to the joint occurrence of a pulse at the output of level selector No. 2 at the time t, and a pulse at the output of level selector No. 1 at the previous time, $t - \tau$. Thus, the ratio of the number of pulses appearing at the conditional pulse output jacks to the number of pulses appearing at the level No. 1 pulse output jacks, in a given measurement interval, gives a measure of the probability that a pulse occurs in the amplitude interval of level selector No. 2 conditional on the occurrence of a pulse in the amplitude interval of level selector No. 1 at a time τ previously. By varying the sampling repetition period from 5 μsec to 1 msec, and by varying the number of periods delay from one to one hundred and twenty eight, the delay time τ may be varied from 5 μsec to about 0.13 seconds.

The main disadvantage of the conditional probability delay unit discussed above is the so-called lost pulse effect. A review of the operation of this system discloses that once a level No. 1 pulse initiates the delay counter chain, no further level No. 1 pulses can affect the unit until after the delay interval, τ. Level No. 1 pulses that occur during the delay interval are the lost pulses. The effect of these lost pulses on the accuracy of measurement depends on whether or not the sampling wave and the wave being studied are statistically independent. If these two waves are independent statistically, as they are in the case of the amplitude distribution measurements (including the measurement of the autocorrelation function of the clipped speech wave), the lost pulse effect in essence converts the measurement process from the case of periodic sampling to the case of random sampling with a reduction in the effective number of samples. In order to maintain a given accuracy, measurements must then be made over a longer measurement interval. If, on the other hand, the two waves are not statistically independent, as is the case of the measurement of $P(T_1 | T_1, m)$ discussed previously, an undetermined error is introduced. The exact effects of this error have not been studied. In any case, the lost pulse phenomenon does not occur, of course, if only one period delay is used.

Overall characteristics

An overall view of the entire equipment used in this investigation is given in Figs. XXXVI and XXXVII. This equipment is contained in three six-foot relay racks. The left hand rack in Fig. XXXVI contains the zero-crossing distribution equipment, the center rack contains the amplitude-distribution equipment, and the right hand rack contains the counters.

The units of the amplitude distribution equipment contained in the center rack are,

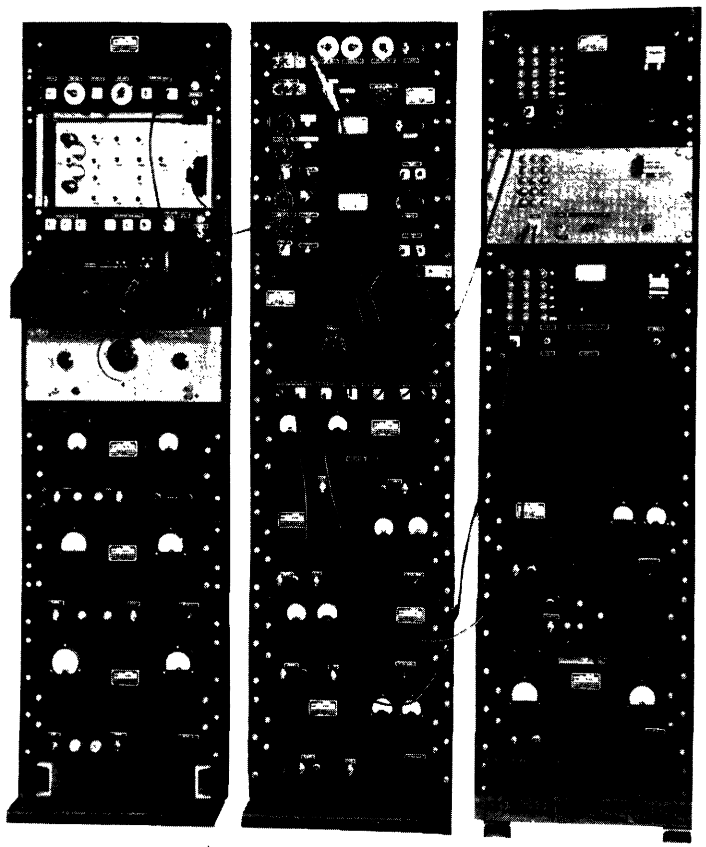


Fig. XXXVI

Probability measuring equipment, front view.

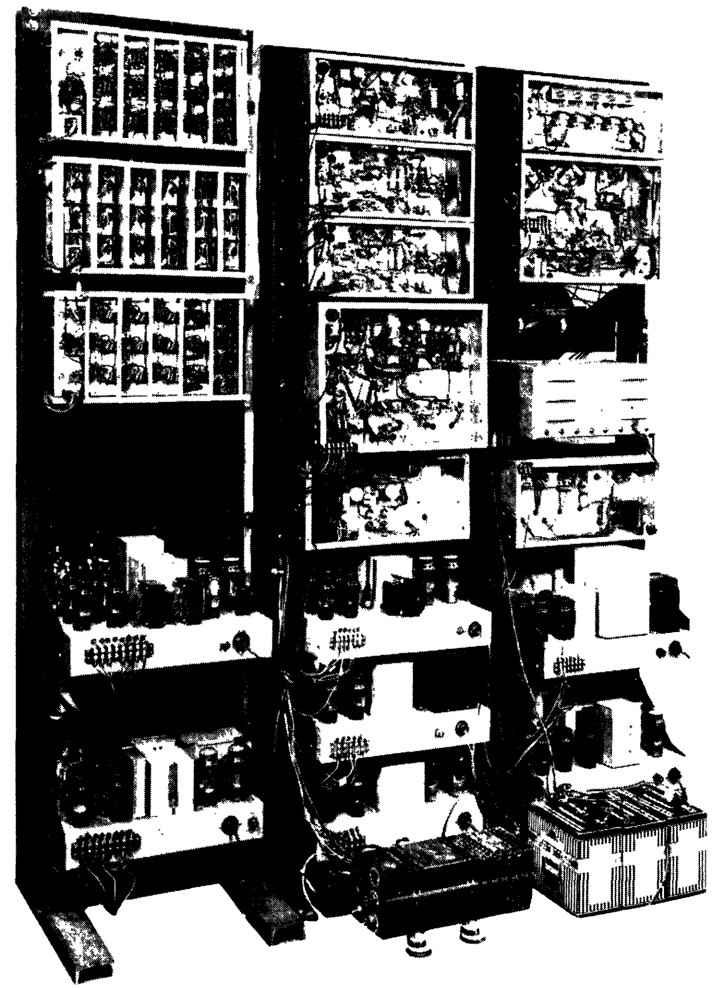


Fig. XXXVII

Probability measuring equipment, rear view.

from top to bottom: the pulse amplitude modulator, level selector No. 1, level selector No. 2, the conditional probability delay unit, a -100 volt regulated power supply, and two $+300$ volt regulated power supplies. The units contained in the counter rack are from top to bottom, three counters, a $+300$ volt regulated power supply, a filament transformer, and a $+300$ volt regulated power supply. The unit on the floor by the center rack is a voltage regulating transformer. In order to meet the stability requirements placed on this equipment, it was found necessary to use a separate filament transformer for each chassis of the amplitude-distribution equipment, and to supply the individual filament transformers from a voltage regulating transformer.

The overall long time stability of the amplitude distribution equipment was determined by making two types of measurement at specified times after the entire equipment was turned on. With no modulating signal applied to the pulse amplitude modulator, the height of the unmodulated output pulse of this unit was measured with each level selector. This test provides a measurement of the stability of the pulse amplitude modulator and of the level selector potentiometer circuits. A 1-kcps sine wave was then used to modulate the pulse amplitude modulator. The amplitude of this sine wave was adjusted until the range of variation of the amplitude modulated output pulses was about fifty volts. The relative frequency of occurrence of the center one volt amplitude interval was then measured. The first probability distribution density of the instantaneous amplitude for a sine wave is quite flat at its center, hence slight shifts in the location of the interval being studied will produce practically no changes in the measured relative frequency. However, any change in the width of the amplitude interval produces an almost equal change in the relative frequency. Thus this test mainly provides a measurement of the stability of the complete level selector.

The results of the above tests are plotted in Fig. XXXVIII. It may be seen from this figure that the accuracy requirements placed upon the amplitude distribution equipment are met after the system has been in operation for a period of about six hours. This period corresponds roughly to the time required for the system to reach a temperature equilibrium. Once warmed up, the system will operate satisfactorily unless the temperature rating of some component is exceeded, or some other failure occurs. With forced air cooling, this equipment has been operated satisfactorily for continuous periods up to three weeks (about five hundred hours). It should be mentioned at this time, that the auxiliary equipment may also require a warmup time. For example, significant tape speed changes may occur during the first hour of operation of the magnetic tape recorder.

D. Zero-Crossing Distribution Equipment

Introduction

A review of sec. II-B would show that the autocorrelation function of the clipped speech wave could be measured with the amplitude-distribution equipment discussed

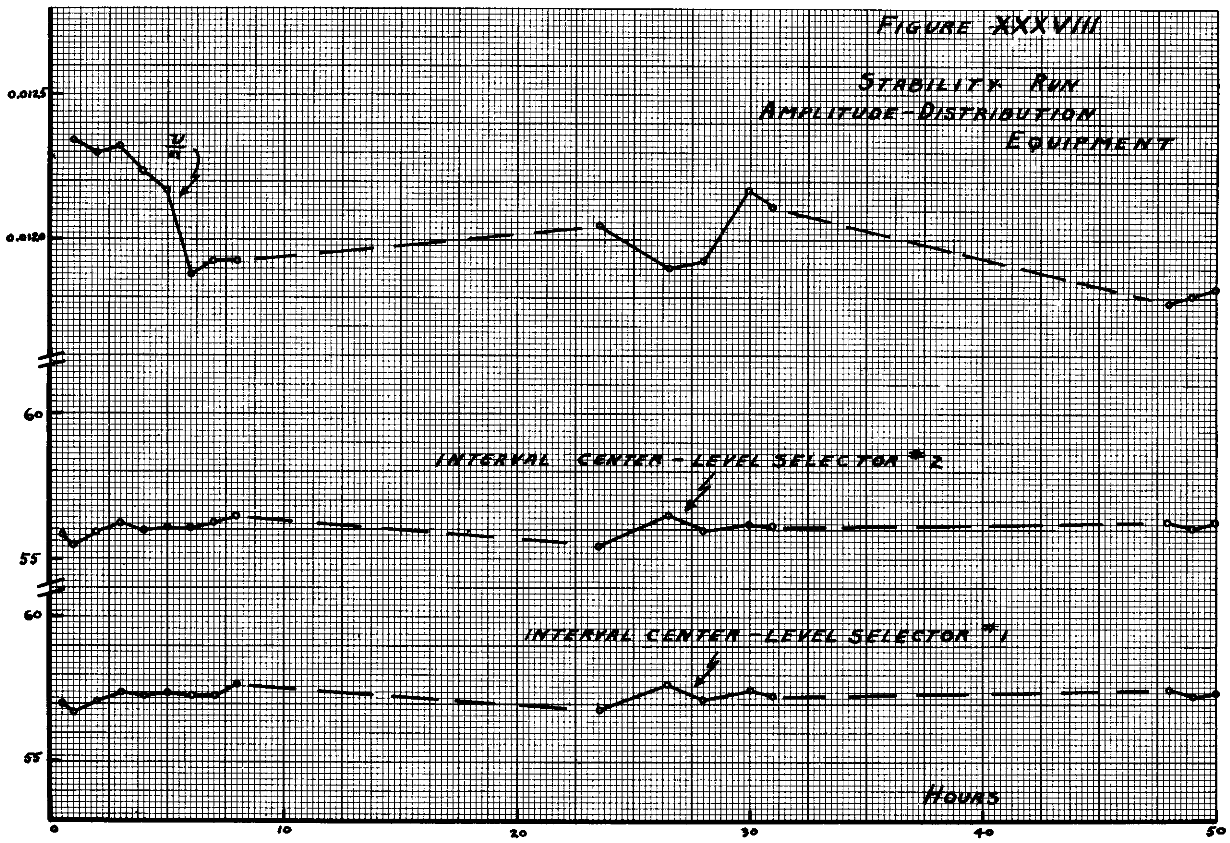


FIGURE XXXVIII
STABILITY RUN
AMPLITUDE-DISTRIBUTION
EQUIPMENT

earlier. The only change that would have to be made would be to disable the upper level multivibrator so that the level selectors would indicate whether or not the amplitudes of their input pulses were above a specified amplitude (corresponding to zero amplitude of the voice wave).

A review of secs. II-A and II-B would show that the first probability distribution density $W_1(T_0)$ and the conditional probability $P(T_1|T_1, m)$ could be measured with the system shown in block diagram form by Fig. XXXIX. Examination of this figure shows that the only items of equipment not previously discussed are the 20-cps square-wave bias generator, the clipped wave generator, and the zero-crossing pulse generator. An available square wave generator (Measurements Corporation – Model 71) was used, but the clipped-wave generator and the zero-crossing pulse generator had to be designed and constructed specifically for this investigation.

The clipped wave generator

The purpose of the clipped wave generator is to convert the voice wave into a zero-crossing wave. That is, the output of the clipped wave generator should be constant at one value when the voice wave is negative, and constant at a different value when the voice wave is positive. For satisfactory operation of the succeeding apparatus, the output pulses of the clipped wave generator should have a peak to peak amplitude of at at least twenty volts.

The design of an equipment capable of meeting the above requirements is straight forward, and use may be made of well known pulse circuit design techniques. The schematic diagram of such an equipment is shown by Fig. XL.

The design and operation of the clipped wave generator is as follows: The voice wave is applied to one grid of the dual-triode mixer, or adding tube, V1. The square wave bias signal is applied to the other grid of the mixer tube, V1. The plate waveform of V1 thus is the sum of the voice and bias waves. The output of V1 is then amplified and clipped by the five-stage slicer, V2 through V6. Each stage of the slicer is composed of a cathode follower driving a grounded grid amplifier, and is of the same form as that of the slicers used in the level selector. This form of clipper was chosen because of its high inherent stability. The output of the last slicer stage, V6, is supplied to the clipped wave output jacks through cathode follower, V7.

The measured output of the unit described above is a thirty volt, peak-to-peak, essentially rectangular wave. The measured rise time from ten percent to ninety percent amplitude is about 2 μsec, and there is about a five percent droop of the horizontal portions of the rectangular wave for a twenty cycle per second sinusoidal input wave. The clipping level of this unit (i.e. the largest signal input which will not be clipped) is about 1.1 mv peak to peak. Thus, if the applied voice wave has a peak to peak amplitude of 1.1 volts, a sixty db clipping ratio is obtained. With no signals applied, and a six hundred ohm shunt across the audio input terminals (as usually connected for cable termination), the effective noise voltage referred to the input terminals is about 0.3 mv peak to peak. These characteristics are entirely adequate for the purposes of this investigation.

Design requirements of the zero-crossing pulse generator

The purpose of the zero-crossing pulse generator is to generate sets of pulses which, when supplied to the amplitude-distribution equipment, will enable the measurement of the various zero-crossing period distributions. Two sets of pulses are required: the zero-crossing pulses which are constant amplitude pulses occurring at the times of the zero crossings, and the period pulses which are varying amplitude pulses occurring at the times of the zero crossings. The amplitude of any given period pulse should be a function only of the duration of the zero-crossing period occurring between that pulse and the preceding period pulse. This restriction eliminates the possibility of using previously designed equipments such as frequency meters or the like. These devices have invariably been designed so as to involve an averaging effect over several zero-crossing periods.

The main design criterion to be met is that the equipment be able to respond to the entire zero-crossing period range of the speech wave. As shown by the various curves given earlier, the required ratio of maximum to minimum period may well be of the order of one thousand to one. In order to cover easily such a range, it was decided that

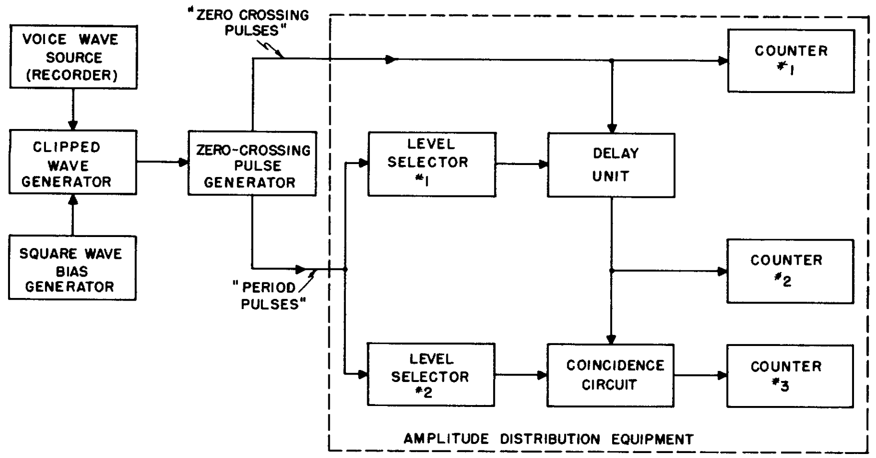


Fig. XXXIX Zero-crossing distribution equipment.

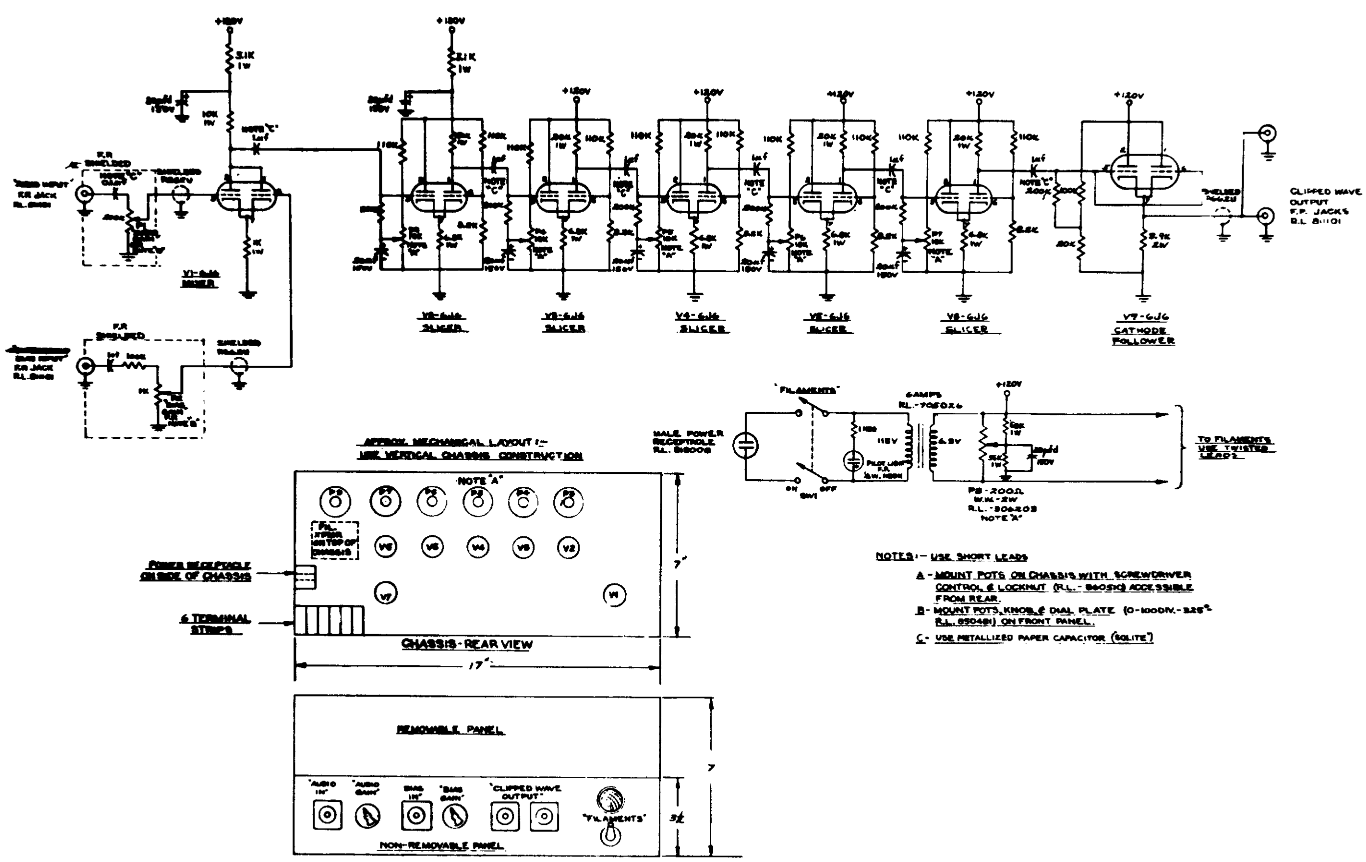
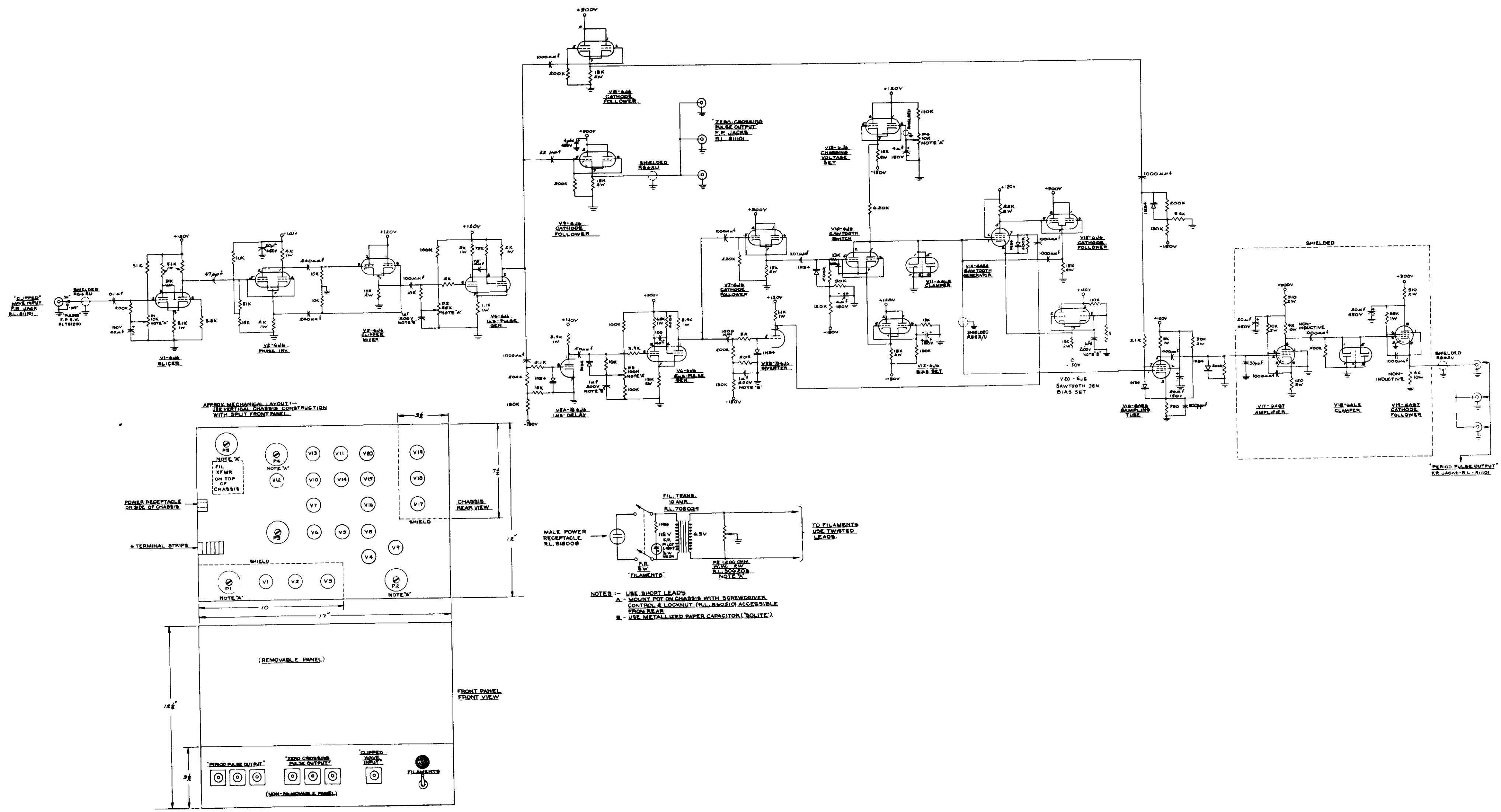


Fig. XL The clipped-wave generator.


"CLIPPED" WAVE INPUT F.P. JACK R.L. 811101
SHIELDED RG62U
V1-6J6 SLICER
V2-6J6 PHASE INV.
V3-6J6 CLIPPER MIXER
V5-6J6 1st PULSE GEN.
V4A-1/2 6J6 1st DELAY
V6-6J6 2nd PULSE GEN.
V7-6J6 CATHODE FOLLOWER
V8-6J6 CATHODE FOLLOWER
V9-6J6 CATHODE FOLLOWER
V4B-1/2 6J6 INVERTER
"ZERO-CROSSING PULSE OUTPUT" F.P. JACKS R.L. 811101
SHIELDED RG62U
V10-6J6 SAWTOOTH SWITCH
V11-6AL5 CLAMPER
V12-6J6 BIAS SET
V13-6J6 CHARGING VOLTAGE SET
V14-6AG5 SAWTOOTH GENERATOR
V15-6J6 CATHODE FOLLOWER
SHIELDED RG63/U
V20-6J6 SAWTOOTH GEN BIAS SET
V16-6AG5 SAMPLING TUBE
SHIELDED
V17-6AG7 AMPLIFIER
V18-6AL5 CLAMPER
V19-6AG7 CATHODE FOLLOWER
NON-INDUCTIVE
SHIELDED RG62U
"PERIOD PULSE OUTPUT" F.P. JACKS R.L. 811101
APPROX. MECHANICAL LAYOUT — USE VERTICAL CHASSIS CONSTRUCTION WITH SPLIT FRONT PANEL
POWER RECEPTACLE ON SIDE OF CHASSIS
6 TERMINAL STRIPS
FIL XFMR ON TOP OF CHASSIS
SHIELD
NOTE "A"
CHASSIS REAR VIEW
(REMOVABLE PANEL)
FRONT PANEL FRONT VIEW
"PERIOD PULSE OUTPUT"
"ZERO CROSSING PULSE OUTPUT"
"CLIPPED WAVE INPUT"
FILAMENTS
(NON-REMOVABLE PANEL)
MALE POWER RECEPTACLE R.L. 818008
FIL. TRANS. 10 AMP. R.L. 708029
115V
F.P. PILOT LIGHT 1/4 W NEON
6.3V
F.P. SW "FILAMENTS"
P6-200 OHM W.W. 2W R.L. 306203 NOTE "A"
TO FILAMENTS USE TWISTED LEADS
NOTES :- USE SHORT LEADS
A - MOUNT POT ON CHASSIS WITH SCREWDRIVER CONTROL & LOCKNUT (R.L. 860510) ACCESSIBLE FROM REAR
B - USE METALLIZED PAPER CAPACITOR ("SOLITE").

the relation between period pulse amplitude and zero-crossing period should be of the form

$$E = a\left(1 - \epsilon^{-\frac{T_1}{b}}\right) \tag{115}$$

where E is the amplitude of the period pulse, T_1 is the duration of the zero-crossing period, and where a and b are appropriately chosen constants. The main advantage of this particular functional relationship between E and T_1 is the ease with which it may be obtained practically.

Design and operation details

A zero-crossing pulse generator capable of meeting the aforementioned design requirements is shown in schematic diagram form by Fig. XLI. Idealized waveforms relating to this equipment are given in Fig. XLII. The detailed operation of this unit is discussed below.

The clipped speech wave input (Fig. XLII-A) is applied to a slicer, V1. This slicer differs from the previously discussed slicers in that positive feedback is used to decrease the rise and fall time of the resultant output wave. The output wave of V1 is then peaked and applied to the polarity inverter, V2. The input grid waveform of V2 then consists of a positive pulse whenever the clipped speech wave jumps positively, and a negative pulse whenever the clipped speech wave jumps negatively. The cathode output waveform from V2 has the same form as its grid waveform. The cathode waveform of V2 is further peaked and applied to the right hand grid of the common-cathode impedance clipper mixer, V3 (Fig. XLII-B). The plate output waveform of V2 is the polarity inverted counterpart of the cathode waveform. The plate waveform of V2 is further peaked and applied to the left hand grid of the clipper-mixer V3 (Fig. XLII-C). The negative peaks of the grid waveforms of V3 drive the respective halves of that tube beyond cutoff and thus do not appear at the cathode. Therefore, the waveform at the cathode of V3 consists of a positive pulse each time the clipped speech waveform changes state regardless of the direction of change. These pulses are used to trigger the 1-μsec pulse generator, V4. The plate waveform of V4 (Fig. XLII-D) thus consists of a positive, 1-μsec, rectangular pulse each time the clipped speech wave changes state. These pulses are supplied to the zero-crossing pulse output jacks through the cathode follower, V9.

The positive pulses appearing at the plate of the 1-μsec pulse generator are also applied to the grid of the 1-μsec delay tube, V5A. The plate waveform of V5A consists of negative, 1-μsec pulses. These pulses are peaked and the negative peak is clipped by the output coupling circuit of V5A. The resultant waveform thus consists of a positive pulse appearing at the end of the 1-μsec input pulse, and is used to trigger the 5-μsec pulse generator, V6. The plate waveform of V6 thus consists of a 5-μsec rectangular pulse (Fig. XLII-E) starting at the end of the 1-μsec zero-crossing pulse. This 5-μsec pulse is then used to operate the zero-crossing period sawtooth generator composed of tubes V5B, V7, V10 through V15, and V20.

The zero-crossing period sawtooth generator consists basically of a feedback-type sawtooth generator, V10, V14, and V15. The remaining tubes, V5B, V7, V11, V12, V13, and V20 are added to adapt this basic sawtooth generator to fit the particular requirements of our problem.

For convenience let us assume that a zero-crossing period has just ended. At the end of the 1-μsec zero-crossing pulse, the 5-μsec pulse output of V6 is inverted and applied to the suppressor grid of the sawtooth generator, V14, cutting off the plate current of that tube. A positive 5-μsec pulse is simultaneously applied to the normally below cutoff grid of the sawtooth switch tube, V10. The cathode of V10 is maintained at about −14 volts. The positive pulse on the grid of the switch tube, V10 then causes plate current to flow through that tube. The plate current of V10 flows from ground through the clamping diode, V11, hence, as the conducting impedance of the diode is very low, the plate voltage of the switch tube, V10, tends to be driven to about zero volts. The combination of cutting off the plate current of V14, and clamping of the plate voltage of V10, causes the 1000-μμfarad sawtooth generating condenser (connected from the control grid of V14 to the cathode of V15) to be rapidly charged. The charging time constant is considerably smaller than 5 μsec. Thus, by the end of the 5-μsec pulse, the control grid voltage of V14 has reached a constant value of about zero volts.

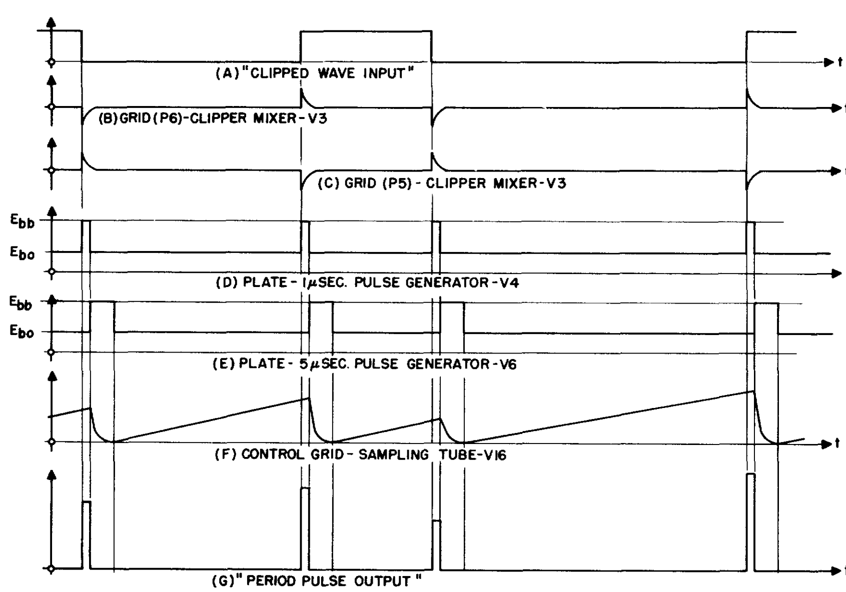


Fig. XLII Zero-crossing pulse generator idealized waveforms.

At the end of the 5-μsec pulse, the switch tube, V10, is returned to its normally cutoff condition, and the suppressor grid of the sawtooth generator, V14, is returned to a value permitting the plate current of that tube to flow. Thus the 1000-μμfarad sawtooth condenser starts to discharge with a time constant equal to 1000-μμfarad times the 620 kilohms of the discharge resistance (connected from the control grid of V14 to the cathode of the charging voltage set tube, V13) times the effective voltage gain of the sawtooth generator, V14. This process continues until the end of the next occurring 1-μsec zero-crossing pulse. The control grid waveform of V14 is thus an exponentially rising voltage. The maximum value to which this grid voltage may rise is set by the charging voltage set tube, V13, (and is about + 6 to + 8 volts). At the end of the next occurring zero-crossing pulse, another 5-μsec pulse is generated by V6, and the process repeats itself.

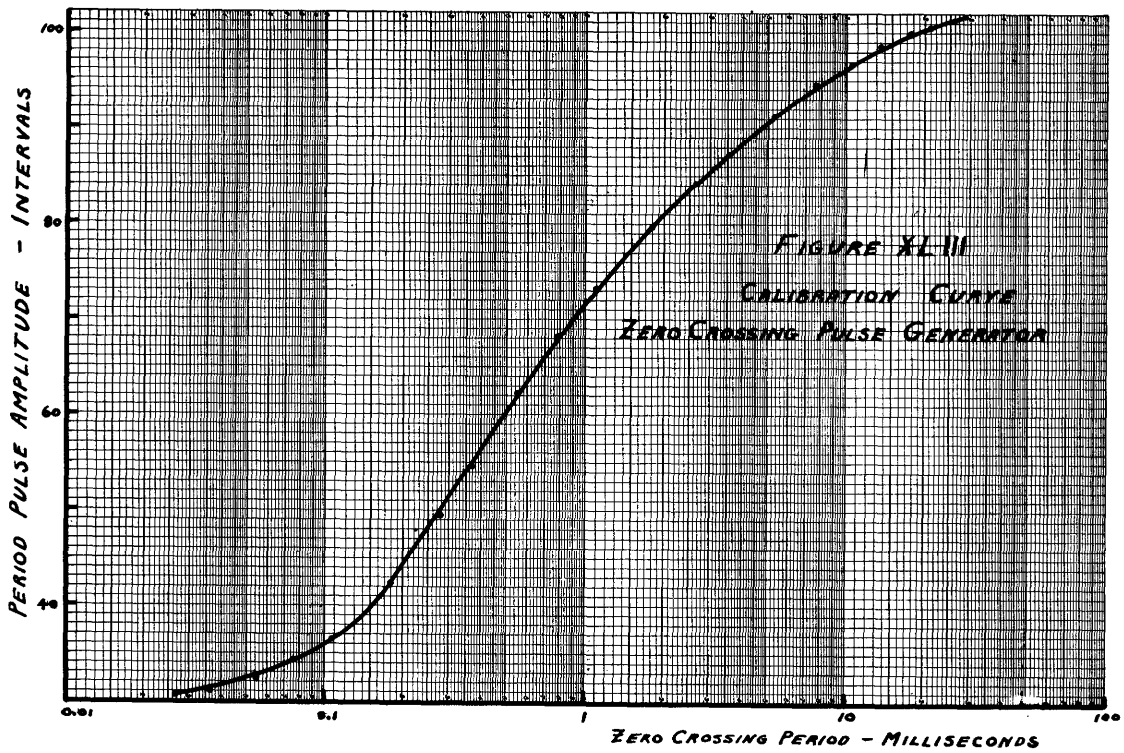
The sawtooth waveform appearing at the control grid of the sawtooth generator, V14, is simultaneously applied to the control grid of the sampling tube, V16 (Fig. XLII-F). The suppressor grid of the sampling tube is maintained below cutoff, thus cutting off the flow of plate current, except during the 1-μsec zero-crossing pulse. During the zero-crossing pulse, the suppressor grid of the sampling tube is driven positively into the conduction region. Thus at this time, a 1-μsec negative pulse appears at the plate of the sampling tube, V16. The amplitude of this pulse is determined by the amplitude of the control grid voltage of V16, and hence by the peak amplitude of the sawtooth waveform generated by the zero-crossing period sawtooth generator.

The negative pulse appearing at the plate of the sampling tube is inverted and amplified by V17. The resulting varying amplitude positive pulse is supplied to the period pulse output jacks through the cathode follower, V19 (Fig. XLII-G). This output pulse is thus a 1-μsec rectangular pulse occurring when the clipped speech wave changes state, and whose amplitude is a function only of the duration of the zero-crossing period occurring between that pulse and the preceding pulse. We have thus finally obtained the desired period pulse output.

Overall characteristics

Views of the zero-crossing distribution equipment are given in Figs. XXXVI (front view and XXXVII (rear view) given earlier. The units of the zero-crossing distribution equipment contained in the left hand rack of Fig. XXXVI are, from top to bottom: the clipped wave generator, the zero-crossing pulse generator, a vacuum tube voltmeter (Ballantine Type 300) for monitoring and testing purposes, a 20-cps to 20-kcps sine wave oscillator (Hewlett-Packard Model HP200 BR) for calibration and testing purposes, a negative 150-volt regulated power supply, a positive 120-volt regulated power supply, and a positive 300-volt regulated power supply.

The calibration of the zero-crossing pulse generator may be obtained by connecting the sine wave oscillator to the input of the clipped wave generator. The output of the clipped wave generator is in turn connected to the input of the zero-crossing pulse generator, Finally the period pulse output of the zero-crossing pulse generator is supplied to one of the level selectors of the amplitude distribution equipment. The amplitude of the period pulses may then be measured with the level selector, and may be plotted as a function of the duration of the zero-crossing period of the sine wave input to form a calibration curve. A typical calibration curve obtained by this method is given by Fig. XLIII.



Conclusions and Suggestions for Further Work

There were two main objectives for this investigation: (1) the design and construction of apparatus which could measure certain probability distributions of random time functions, and (2) the application of this apparatus to a preliminary study of the speech wave.

The problem of measuring a probability of a random time function was shown to reduce generally to the problem of measuring a time average. The time average may be measured by a process of sampling and summation, or by continuous integration. We used a sampling method as its choice simplified the design of the conditional probability delay unit. However, if suitable continuous delay systems become available, a simplification of the apparatus may be possible.

The apparatus developed during this study is flexible, reasonably accurate, and relatively simple. This apparatus can be used in the study of any random process whose statistics lie in the range of the voice wave statistics. These characteristics, however, were purchased at the expense of making tedious the taking of data. For example: for each measurement, the operator must record the pertinent apparatus settings, start the tape-recorder, start the measurement interval, wait five minutes or so, stop the measurement interval, stop the tape-recorder, read the counters, and change the apparatus settings for the next measurement. Later, the probabilities must be computed from the counter readings. The total time required per point is the measurement interval plus about three minutes. If intensive experimental studies of probabilities are to be made in the future, it is highly recommended that time and effort be devoted to the development of apparatus which will take a number of data simultaneously, and which will be automatic in operation.

The measurements made during this study were of the stationary, or long-time, speech statistics. It was shown that, in order to obtain a reasonable accuracy, the required measurement interval had to be of the order of three minutes, or more. These measurements were made mainly for substantially undistorted speech. The results obtained can, for the most part, be readily explained, and are of interest chiefly in that they indicate possible directions for future studies. For example, it would now be of interest to study the effect on the probability distributions of passing speech through various systems, such as band-pass filters, pre-emphasis circuits.

A further field of interest would be the study of the short-time speech statistics. This would include the study of the statistics of the individual sounds. Some measurements of the short-time autocorrelation functions have been made by Stevens (23), and others, but no published results are known for the short-time probabilities. Studies of this type might provide useful alternatives to the representation of speech sounds by spectrograms.

Acknowledgement

I wish to express my appreciation to Professor R. M. Fano for the inspiration, encouragement, and advice received from him during this study; and to Professors L. L. Beranek, Y. W. Lee, and J. C. R. Licklider for their continued support and encouragement. It would not be practicable to mention all of the others who have contributed to this investigation, however, I would like to acknowledge the assistance of my "voices": Kay Aborjaily and John Costas.

References

1. N. Wiener: Cybernetics, The Technology Press, John Wiley, N.Y. 1948

2. N. Wiener: The Extrapolation, Interpolation, and Smoothing of Stationary Time Series, The Technology Press, John Wiley, N.Y. 1949

3. C. E. Shannon: A Mathematical Theory of Communications, BSTJ, 27, I, 379-423, July, 1948; 27, II, 623-656, Oct. 1948

4. C. E. Shannon: Communications in the Presence of Noise, Proc. I.R.E. 37, 1, 10-21, Jan. 1949

5. S. Chandrasekhar: Stochastic Problems in Physics and Astronomy, Rev. Mod. Phys. 15, 1, 1-89, Jan. 1943

6. M. C. Wang, G. E. Uhlenbeck: On the Theory of the Brownian Motion, II, Rev. Mod. Phys. 17, 2 and 3, 323-342, April-June, 1945

7. H. Cramer: Mathematical Methods of Statistics, Princeton Univ. Press, 1946

8. T. P. Cheatham, Jr.: Experimental Determination of Correlation Functions and their Application in Communication Theory, Technical Report No. 122, Research Laboratory of Electronics, M.I.T. unpublished

9. J. C. R. Licklider, I. Pollack: Effects of Differentiation, Integration, and Infinite Peak Clipping upon the Intelligibility of Speech, J. Acous. Soc. Am. 20, 1, 42-51, Jan. 1948

10. J. C. R. Licklider, D. Bindra, I. Pollack: The Intelligibility of Rectangular Speech Waves, Am. J. Psychology, 61, 1, 1-20, Jan. 1948

11. L. J. Sivian: Speech Power and Its Measurement, BSTJ, 8, 4, 646-661, Oct. 1929

12. H. J. von Braunmühl: Über die Intensitätsverhältnisse von natürlichen Klangbildern mit besonderer Berücksichtigung der Rundfunksendung, Zeits. für Techn. Physik, 14, 11, 507-512, 1933

13. D. Thierbach, H. Jacoby: Über die Verteilung der Sprechspannungen bei der Ubertragung Zahlreicher trägerfrequenter Gespräche, Zeits. für Techn. Physik, 17, 12, 553-557, 1936

14. H. K. Dunn, S. D. White: Statistical Measurements on Conversational Speech, J. Acous. Soc. Am. 11, 278-288, Jan. 1940

15. B. B. Jacobsen: The Effect of Nonlinear Distortion in Multi-Channel Amplifiers, Elec. Communication, 19, 1, 29-54, July 1940

16. G. Sacerdote: Misure Statistiche di Intensita Vocali, Atti del Congresso Internazionale Della Radio, Rome, 389-401, 1947

17. G. E. Pihl, S. H. Chang, et al: Visual Message Presentation, Northeastern Univ. Quarterly Progress Report, Contract AF 19(122)-7, No. 1, July 25, 1949. Also No. 2, Oct. 25, 1949

18. Radio Corp. of America: Instructions for Velocity Microphone, Type 44-BX, RCA Victor Div, Camden, N.J.

19. L. L. Beranek, H. P. Sleeper, Jr.: The Design and Construction of Anechoic Sound Chambers, J. Acous. Soc. Am. 18, 1, 140-150, July, 1946

20. R. W. Roop: Studio Research, M.I.T. Acoustics Lab. Quarterly Progress Report, 7-9, April-July, 1948

21. B. Chance, et al: Waveforms, 19, M.I.T. Radiation Lab. Series, McGraw-Hill, N.Y. 1949

22. W. C. Elmore and M. Sands: Electronics-Experimental Techniques, Div. V, I, National Nuclear Energy Series, McGraw-Hill, N.Y. 1949

23. K. N. Stevens, D. L. Bowler: Investigation of Autocorrelation Functions as Representations for Speech Sounds, M.I.T. Acoustics Lab. Quarterly Progress Report, 8-11, Oct.-Dec. 1949

24. H. E. Singleton: A Digital Electronic Correlator, Technical Report No. 152, M.I.T. Research Laboratory of Electronics, Feb. 21, 1950

25. J. P. Costas: Periodic Sampling of Stationary Time Series, Technical Report No. 156, M.I.T. Research Laboratory of Electronics, May 16, 1950