

CONTRIBUTIONS OF NORBERT WIENER TO COMMUNICATION THEORY

BY WILLIAM L. ROOT

Information theory, communication theory, detection theory are terms now commonly used by engineers who are concerned with radio or telephone communications, electromagnetic and sonic measurements, seismic measurements, and indeed the transformation, storage and transmittal of data from any source by electronic or mechanical devices. They are terms also used by scientists concerned with the function of the human brain and nervous system, the behavior of small groups of people, the interplay between men and machines, and the behavior of whole societies. Coupled with the study of transmittal of information is the study of use of information in mechanisms, and at least one direction this latter study leads is toward the modern theory of control. Norbert Wiener was interested, as is well known, in this whole broad expanse of problems; he made this interest completely explicit in his invention of the word cybernetics, and his explication of its meaning in the book *Cybernetics* [138].* One of Professor Wiener's contributions to the scientific society was his recognition and insistence that communication and control problems which occur in very different contexts often have the same essentials; or, stated differently, that there was potentially a science of communication and control. However, it is beyond the scope of these remarks, and beyond my competence, to talk about his work from such a general point of view; so the comments here shall refer only to communication theory as it applies to electrical engineering technology.

In *I am a mathematician* [177] Wiener says "One interesting problem which we attacked together was that of the conditions restricting the Fourier transform of a function vanishing on the half line. This is a sound mathematical problem on its own merits, and Paley attacked it with vigor, but what helped me and did not help Paley was that it is essentially a problem of electrical engineering." The solution to this problem for functions of class L_2 is contained in Theorem XII of *Fourier transforms in the complex domain* [92], where it ostensibly plays the role of a key result in the theory of quasi-analytic functions. This theorem was explicitly pointed out to engineers, and its implica-

* The bold-faced numbers in brackets refer to numbered references in the Bibliography of Norbert Wiener. Bold-faced numbers in parentheses refer to the References at the end of this article.

tions to electric circuit synthesis discussed, by H. Wallman in the M.I.T. Radiation Laboratory Series fourteen years later (1). It is a theorem very well known to mathematical analysts, but it also became, many years after its statement, very well known to electrical engineers. It might almost be called the fundamental theorem of linear electric networks instead of the fundamental theorem of quasi-analytic functions, as Wiener dubbed it. Its significance to engineering is the following: a linear, time-invariant system in a known state at $t=0$ can be characterized by its response to an impulse at $t=0$, or by the Fourier transform of that impulse response. The Fourier transform of the impulse response is known as the transfer function of the system and is a valuable concept for the engineer, both because it is useful in calculations, and because it is useful in designing and interpreting experiments. Its complex value at a particular frequency represents the gain and phase shift of the system at that frequency. The Paley-Wiener theorem gives a necessary and sufficient condition that a given gain function be possible for a linear, time-invariant system that is *realizable*, that is, that operates only on the past. It is remarkable that Wiener knew at the time he and Paley worked on this theorem that it was electrical engineering, because few electrical engineers at that time would have recognized it as electrical engineering.

This one theorem and its history of application strikes me as typical of a large part of Wiener's contribution to the development of modern communication theory. Whether in all such instances he knew he was providing mathematical structure for communications theorists and engineers ten, twenty or thirty years hence, I do not know, but he evidently often hoped that he was. He did, of course, receive early orientation in problems of electrical engineering in discussions with Vannevar Bush in the 1920's, and these discussions must have been unusually sophisticated by the engineering standards of that time. His collaboration with Y. W. Lee somewhat later, and contacts with various electrical engineers from the time of World War II onward kept strong his ties with electrical engineering.

In a communication theory problem, man-made signals must always be of finite duration, and often their duration is limited by a fixed bound. An electrical signal of finite duration has finite energy. Hence, if one is concerned with an harmonic analysis of such signals in the communication system, the Paley-Wiener theorem (Theorem X, *op. cit.*) characterizing the class of functions

$$F(z) = \int_{-A}^A f(u)e^{iuz} du,$$

where $f(u)$ belongs to L_2 over $(-A, A)$ is basic. It also applies, of course, to the structural characterization of linear systems with *finite memory*, i.e., systems whose response to any input goes to zero within some interval τ after the stimulus is removed. The theorem has perhaps been less useful in application than the one cited first; I think this is so because the condition is a function-theoretic one instead of an integrability one, and therefore farther removed from the kind of constraints that appear "naturally," but this difficulty would seem to be inherent and unavoidable.

In the introduction to *Generalized harmonic analysis* [73] Wiener says, "The two theories of harmonic analysis embodied in the classical Fourier series development and the theory of Plancherel do not exhaust the possibilities of harmonic analysis. The Fourier series is restricted to the very special class of periodic functions, while the Plancherel theory is restricted to functions which are quadratically summable, and hence tend on the average to zero as their argument tends to infinity. Neither is adequate for the treatment of a ray of white light which is supposed to endure for an indefinite time."

The reference here is obviously to harmonic analysis in an L_2 context; with that limitation one may say that the Fourier transform yields an harmonic description of functions of finite energy, and the Fourier series yields an harmonic description of a very special class of functions of finite power (i.e., infinite total energy, but finite energy per unit interval). What was needed and what Wiener developed was an harmonic description of a wide class of functions of finite power, including the classes of periodic and almost periodic functions, but also including functions with a continuous or mixed "power spectrum."

The class of functions to which Wiener's generalized harmonic analysis applies is that of those measurable functions $f(t)$ for which

$$(1) \quad \phi(\tau) = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T f(\tau + t) \overline{f(t)} dt$$

exists for every τ . This restriction has interesting heuristic implications; for $\phi(0)$ to exist means obviously that the indefinitely continuing function $f(t)$ has associated with it an average power; for $\phi(\tau)$ to exist in general would indicate that, almost all of the time, $f(t)$ exhibits some kind of average regularity in its behavior. One would expect that many physical phenomena that are random and unpredictable in their course, but appear in an environment for which the grosser attributes are unchanging, would be appropriately modeled by such

functions. Experience has borne this out, of course. One very important example is electrical noise as generated by a resistor at constant temperature, or by a vacuum tube or transistor under constant operating conditions. Since electrical noise is one of the chief ingredients in any continuous-time electrical communication problem, Wiener's generalized harmonic analysis is a fundamental tool in the analysis of such problems. It has to a certain extent been replaced formally in the analysis of electrical noise by the introduction of the stationary stochastic process as a mathematical model for noise phenomena. The theorem of Bochner that a positive-definite function is the Fourier-Stieltjes transform of a bounded monotone function, and the theorem of Khintchin that autocorrelation functions of stationary processes are positive-definite functions then provides an harmonic analysis which fits pretty much the same problems as the Wiener generalized harmonic analysis. Since the Birkhoff ergodic theorem guarantees that for almost every sample path of a metrically transitive stationary stochastic process the limit $\phi(\tau)$ as given in (1) exists and equals the autocorrelation function of the process evaluated at τ , the Wiener theory and the stationary process theory appear to one interested in physical applications to be equivalent in a strong sense. To apply the Wiener theory one has to represent the physical phenomena by a function $f(t)$ such that $\phi(\tau)$ exists; to apply the stochastic process theory one has to represent the physical phenomena by a stationary, metrically-transitive process. The extra-mathematical arguments to justify these two models seem pretty much the same. The Wiener theory came first; and it retains interest, I feel, even in applications, precisely because it deals with a single function instead of an ensemble of functions.

Electrical noise is usually such that it not only admits being represented as a stationary stochastic process, but more specifically as a stationary Gaussian stochastic process. This is true because the noise (voltage or current) is generally a macroscopic manifestation of the result of a great deal of independent activity at a microscopic level. The basic theory of Brownian motion developed by Wiener is indispensable for a meaningful study of stationary Gaussian processes with finite variance; the spectral representation of Cramer is in terms of Brownian motion for the Gaussian case; the really workable models for practical calculations involving Gaussian noise are "filtered white noise" processes, i.e., processes of moving averages,

$$(2) \quad x(t) = \int_{-\infty}^{\infty} k(t-u) d\xi(u),$$

where $\xi(u)$ is the Brownian motion process. Modern "noise theory" and "detection theory" are applied mathematical disciplines resting chiefly (for their mathematical part) on parts of the theories of stochastic processes, harmonic analysis and statistical inference. For the first two of these Wiener's contributions are a primary source.

I have made this brief reference to some of Wiener's pure mathematics, because its influence on contemporary electrical communication engineering is perhaps greater than most mathematicians realize. I would like to distinguish between this *applicable* mathematics and a certain body of his work, mostly of later date, which seems to be more properly *applied mathematics* (or even theoretical engineering). It is a somewhat arbitrary distinction and has to do not only with subject matter but with style of writing, the presumed class of readers, and the emphasis or lack of emphasis on rigor.

The book *Extrapolation, interpolation, and smoothing of stationary time series* [144] published in 1949, but originally written as a wartime report to the National Defense Research Committee, is a borderline case, but it seems to me to belong to the second category of applied mathematics. It was written primarily for electrical engineers, even though most electrical engineers of the 1940's could not read it. This was perhaps not a bad thing; the results were sufficiently interesting to provide a stimulus to engineers to acquaint themselves with enough mathematics so that they could read it. In 1965 most theoretical communications and control engineers can read it, and in fact are fairly conversant with its contents. There is now probably not a graduate program in control and communications engineering in the country in which the material of this book is not offered. Wiener lost priority to A. N. Kolmogorov for the essential mathematical results of the prediction theory (see the footnote p. 59, op. cit., for Wiener's own comment on this), but his work was done independently of Kolmogorov's and it was addressed more explicitly to the filtering problems of electrical engineering.

One kind of problem to which the Kolmogorov-Wiener theory applies may be stated in general terms as follows: let $s(t)$, $-\infty < t < \infty$, be an intelligence-bearing signal, $n(t)$, $-\infty < t < \infty$, a noise which interferes with $s(t)$ so that an observer has available only $y(t) = s(t) + n(t)$. Suppose the observer wants a time-realizable, running estimate of $s(t)$; that is, for each t he wants $\hat{s}(t) = L(t)y_t$, where the function \hat{s} should be close to s by some criterion, where

$$\begin{aligned} y_t(u) &= y(u), & u \leq t \\ &= 0, & u > t, \end{aligned}$$

and where $L(t)$ is a functional defined on the class $\{y_t\}$ for a suitably large class of functions y . The problem is to find a set of functions $L(t)$, $-\infty < t < \infty$, which makes ξ close to s . This is the kind of problem that occurs for example in designing a receiver for radio transmissions; the receiver is an apparatus which implements the operations $L(t)$.

In the original Wiener solution to this basic "smoothing" problem it is stipulated that: (1) Both $s(t)$ and $n(t)$ be functions satisfying (1), so that each has an autocorrelation function, and further that the *cross-correlation function*

$$\phi_{sn}(\tau) = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T s(t + \tau)n(t)dt$$

exists for each τ . These correlation functions are to be known in advance.

(2) The operations $L(t)$ are to be linear operations,

$$\xi(t) = L(t)y_t = \int_0^{\infty} y(t - \tau)dK(\tau)$$

which are time-invariant.

(3) The criterion for the closeness of $\xi(t)$ and $s(t)$ is the mean-squared-error,

$$(3) \quad \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T \left| s(t) - \int_0^{\infty} y(t - \tau)dK(\tau) \right|^2 dt.$$

The solution for $K(\tau)$, or rather for its Fourier-Stieltjes transform, is given in Equation (3.28) *op. cit.* The solution is formally the same if condition (1) is replaced by the condition that $s(t)$, $n(t)$ and $(s(t), n(t))$ are stationary stochastic processes and the time-average auto and cross-correlation functions are replaced by the corresponding statistical averages, as, for example,

$$\lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T s(t + \tau)\overline{s(t)}dt \sim Es(t + \tau)s(t),$$

where E denotes mathematical expectation, and if the criterion (3) is replaced by

$$E \left| s(t) - \int_0^{\infty} y(t - \tau)dK(\tau) \right|^2$$

for each t . Further, if $s(t)$, $n(t)$ and $(s(t), n(t))$ are metrically transi-

tive, the solution is exactly the same for each pair of sample functions $s(t)$, $n(t)$ except for those in a set of probability zero.

Modern communication and signal detection theory began when it was realized that a meaningful way to treat such a problem was to treat it statistically, to ascribe to $s(t)$ and $n(t)$ certain average properties, rather than to specify them explicitly, and to use some kind of statistical or average criterion to judge the closeness of \hat{s} to s . There is, of course, a great variety of changes on such a problem; the perturbation may not be caused by simple additive noise, the structure of possible signals $s(t)$ may be more or less narrowly specified, the *a priori* statistical information available may vary widely in kind and quantity, the statistical criteria need not be mean-squared-error, etc. The statistical point of view had gained some acceptance during and shortly after the war, due largely to the work of a dozen or so people, some of whom were mathematicians and physicists who had been led to think of radio communication (and measurement) problems because of their wartime jobs. I will venture the opinion that there were three papers or monographs published in this country in the middle and late 1940's that were more than any others responsible for establishing statistical ratio communications (and radio measurement) theory; these are the papers on random noise by S. O. Rice (2), on information theory by C. Shannon (3), and the monograph on smoothing and predicting by Wiener [144].

The original theory of filtering and prediction was linear (although Wiener also wrote on nonlinear prediction, see [170], [196]). It fitted neatly onto Wiener's earlier work on generalized harmonic analysis and his work with E. Hopf on a class of linear integral equations [78]. But Wiener was also interested in nonlinear problems in engineering; he wanted to provide a rather general structure of theory to support problems in the synthesis and analysis of nonlinear electrical networks and other nonlinear systems. Apparently most of what he did in this area was done in the 1940's, but he published only one report, *Response of a Nonlinear Device to Noise*, M.I.T. Radiation Laboratory Report No. 129, 1942, and no papers in the scientific journals. The first really open publication was the set of lectures, *Nonlinear Problems in Random Theory* [191], in 1958. Some of this work was done in collaboration with Y. W. Lee, and initiated a considerable body of engineering research by Lee and his students in the years after 1950.

Wiener considered nonlinear systems whose output at a particular instant when driven by Brownian motion could be written as functionals of the Brownian motion of the form

$$(4) \quad y(\alpha) = K_0 + \sum_{n=1} \int \cdots \int K_n(\tau_1, \cdots, \tau_n) dx(\tau_1, \alpha) \cdots dx(\tau_n, \alpha)$$

where the kernels K_n are characteristic of the system and where $x(\tau, \alpha)$ is Brownian motion. If the output at a time t seconds later is represented by the same functional except with the arguments of the K_n translated by t , this expression gives the running output of a time-invariant nonlinear system. If two systems yield the same output $y(\alpha)$ when excited by $x(\tau, \alpha)$ for all α in a set of probability one, they are equivalent in a strong sense; they would certainly seem to be equivalent for all engineering purposes. An engineering analysis problem, then, is to find the K_n 's for a given system specified in some arbitrary way; a synthesis problem is, e.g., to specify an electrical network which will realize a given set of K_n 's. Wiener attacked these problems by first orthogonalizing the sequence of homogeneous polynomial integral functionals appearing in (4) into a sequence of non-homogeneous polynomial integral functionals $G_n(K_n, \alpha)$ with the property that $G_n(K_n, \alpha)$ is orthogonal to $G_m(H_m, \alpha)$, $m \neq n$, for any kernel H_m , with respect to Wiener measure. Then any functional of the form of (4) can be written

$$(5) \quad y(\alpha) = \sum_n G_n(K_n', \alpha)$$

for an appropriate choice of kernels K_n' , where now the summands are orthogonal in a statistical sense, i.e., with respect to Wiener measure.

This work of Wiener and Lee (see also (4)) together with the contributions of three or four others largely set the direction that has been followed since for most of the investigations of a general theory of nonlinear communication and control systems with stochastic inputs. Usually the computational complexities arising when one tries to apply such general theory to a specific problem are so great that, thus far, the practical results are limited. Ad hoc procedures are often more effective. There is hope, of course, that electronic computation will eventually overcome these difficulties.

Finally I should like to note that apart from the theorems he proved and the calculations he made, Wiener performed one other service for electrical communication engineering: he was an effective propagandist for the new ideas of the 1940's about the statistical nature of information and its transmission. Because of his prestige as a mathematician, but more than that, because of his reputation

amongst engineers as a first rank mathematician aware of electrical engineering problems, he was listened to when he made statements such as the following, from the magazine, *Electronics*, 1949 [145]: "There is no vagueness in the definition of energy, so that power engineering is a field in which the objectives have been fully understood for a long time. On the other hand, most books on communications say nothing about information, and the average communication engineer does not have a definite measure of information. He studies communication circuits as they are affected by sinusoidal inputs, but he does not discuss the relation between sinusoidal and information-carrying inputs. Only within the past few years have a few engineers begun asking what information is and using the concept of information as a basis for design."

"Because information depends, not merely on what is actually said, but on what might have been said, its measure is a property of a set of possible messages, or of what is called an *ensemble* in statistical mechanics. Such an ensemble is more than a set taken simply; it is a set to which is attributed a probability measure. We thus have a situation that is closely akin to that in statistical mechanics, more especially in the form which Gibbs gave it."

REFERENCES

1. G. R. Valley and H. Wallman, *Realizability of filters*, Appendix A, Vacuum tube amplifiers, MIT Radiation Laboratory Series, Vol. 18, McGraw-Hill, New York, 1948.
2. S. O. Rice, *Mathematical analysis of random noise*, Bell System Tech. J. 23 (1944), 282-332; 24 (1945), 46-156.
3. C. E. Shannon, *A mathematical theory of communications*, Bell System Tech. J. 27 (1948), 379-423; 27 (1948), 623-656.
4. Y. W. Lee, *Contributions of Norbert Wiener to linear theory and nonlinear theory in engineering* in Selected Papers of Norbert Wiener, S.I.A.M. and M.I.T. Press, Cambridge, 1964.

THE UNIVERSITY OF MICHIGAN