Reprinted from *JAES*, vol. 19, no. 9, pp. 734–743, 1971 October.

Determination of Loudspeaker Signal Arrival Times*
Part I

RICHARD C. HEYSER

Jet Propulsion Laboratory, California Institute of Technology, Pasadena, Calif.

Prediction has been made that the effect of imperfect loudspeaker frequency response is equivalent to an ensemble of otherwise perfect loudspeakers spread out behind the real position of the speaker creating a spatial smearing of the original sound source. Analysis and experimental evidence are presented of a coherent communication investigation made for verification of the phenomenon.

INTRODUCTION: It is certainly no exaggeration to say that a meaningful characterization of the sound field due to a real loudspeaker in an actual room ranks among the more difficult problems of electroacoustics. Somehow the arsenal of analytical tools and instrumentation never seems sufficient to win the battle of real-world performance evaluation; at least not to the degree of representing a universally accepted decisive victory. In an attempt to provide another tool for such measurement this author presented in a previous paper a method of analysis which departed from traditional steady state [1]. It was shown that an in-place measurement could be made of the frequency response of that sound which possessed a fixed time delay between loudspeaker excitation and acoustic perception. By this means one could isolate, within known physical limitations, the direct sound, early arrivals, and late arrivals and characterize the associated spectral behavior. In a subsequent paper [2] it was demonstrated how one could obtain not only the universally recognized amplitude spectrum of such sound but also the phase spectrum. It was shown that if one made a measurement on an actual loudspeaker he could legitimately ask "how well does this speaker's direct response recreate the original sound field recorded by the microphone?" By going to first principles a proof was given [3] that a loudspeaker and indeed any transfer medium characterized as absorptive and dispersive possessed what this author called time-delay distortion. The acoustic pressure wave did not effectively emerge from the transducer immediately upon excitation. Instead it emerged with a definite time delay that was not only a function of frequency but was a multiple-valued function of frequency. As far as the sonic effect perceived by a listener is concerned, this distortion is identical in form to what one would have, had the actual loudspeaker been replaced by an ensemble of otherwise perfect loudspeakers which occupied the space behind the position of the actual loudspeaker. Furthermore. each of the speakers in the ensemble had a position that varied in space in a frequency-dependent manner. The sonic image, if one could speak of such, is smeared in space behind the physical loudspeaker.

This present paper is a continuation of analysis and experimentation on this phenomenon of time-delay distortion. The particular emphasis will be on determining how many milliseconds it takes before a sound pressure wave in effect emerges from that position in space occupied by the loudspeaker.

* Presented April 30, 1971, at the 40th Convention of the Audio Engineering Society, Los Angeles.

APPROACH TO THE PROBLEM

The subject matter of this paper deals with a class of measurement and performance evaluation which constitutes a radical departure from the methods normally utilized in electroacoustics. Several of the concepts presented are original. It would be conventional to begin this paper by expressing the proper integral equations, and thus promptly discouraging many audio engineers from reading further. Much of the criticism raised against papers that are "too technical" is entirely just in the sense that common language statements are compressed into compact equations not familiar to most of us. The major audience sought for the results of this paper are those engineers who design and work with loudspeakers. However, because the principles to be discussed are equally valuable for advanced concepts of signal handling, it is necessary to give at least a minimal mathematical treatment. For this reason this paper is divided into three parts. The first part begins with a heuristic discussion of the concepts of time, frequency, and energy as they will be utilized in this paper without the usual ponderous mathematics. Then these concepts are developed into defining equations for loudspeaker measurement, and hardware is designed around these relations. The second part is a presentation of experimental data obtained on actual loudspeakers tested with the hardware. The third part is an Appendix and is an analytical development of system energy principles which form the basis for this paper and its measurement of a loudspeaker by means of a remote air path measurement. The hope is that some of the mystery may be stripped from the purely mathematical approach for the benefit of those less inclined toward equations, and possibly provide a few conceptual surprises for those accustomed only to rigorous mathematics.

TIME AND FREQUENCY

The sound which we are interested in characterizing is the result of a restoration to equilibrium conditions of the air about us following a disturbance of that equilibrium by an event. An event may be a discharge of a cannon, bowing of a violin, or an entire movement of a symphony. A fundamental contribution to analysis initiated by Fourier [4] was that one could describe an event in either of two ways. The coordinates of the two descriptions, called the domains of description, are dimensionally reciprocal in order that each may stand alone in the ability to describe an event. For the events of interest in this paper one description involves the time-dependent pressure and velocity characteristics of an air medium expressed in the coordinates of time, seconds. The other description of the same event is expressed in the coordinates of reciprocal time, hertz. Because these two functional descriptions relate to the same event, it is possible to transform one such description into the other. This is done mathematically by an integral transformation called a Fourier transform. It is unfortunate that the very elegance of the mathematics tends to obscure the fundamental assertion that any valid mathematical description of an event automatically implies a second equally valid description.

It has become conventional to choose the way in which we describe an event such that the mathematics is most readily manipulated. A regrettable consequence of this is that the ponderous mathematical structure buttressing a particular choice of description may convince some that there is no other valid mathematical choice available. In fact there may be many types of representation the validity of which is not diminished by an apparent lack of pedigree. A conventional mathematical structure is represented by the assumption that a time-domain characterization is a scalar quantity while the equivalent reciprocal time-domain representation is a vector. Furthermore, because all values of a coordinate in a given domain must be considered in order to transform descriptions to the other domain, it is mathematically convenient to talk of a particular description which concentrates completely at a given coordinate and is null elsewhere. This particular mathematical entity, which by nature is not a function, is given the name impulse. It is so defined that the Fourier transform equivalent has equal magnitude at all values of that transform coordinate. A very special property of the impulse and its transform equivalent is that, under conditions in which superposition of solutions applies, any arbitrary functional description may be mathematically analyzed as an ordered progression of impulses which assume the value of the function at the coordinate chosen. In dealing with systems which transfer energy from one form to another, such as loudspeakers or electrical networks, it is therefore mathematically straightforward to speak of the response of that system to a single applied impulse. We know that in so doing we have a description which may be mathematically manipulated to give us the behavior of that system to any arbitrary signal, whether square wave or a Caruso recording.

In speaking of events in the time domain, most of us have no reservations about the character of an impulse. One can visualize a situation wherein nothing happens until a certain moment when there is a sudden release of energy which is immediately followed by a return to null. The corresponding reciprocal time representation does not have such ready human identification, so a tacit acceptance is made that its characterization is uniform for all values of its parameter. An impulse in the reciprocal time domain, however, is quite recognizable in the time domain as a sine wave which has existed for all time and will continue to exist for all time to come. Because of the uniform periodicity of the time-domain representation for an impulse at a coordinate location in the reciprocal time domain, we have dubbed the coordinate of reciprocal time as frequency. What we mean by frequency, in other words, is that value of coordinate in the reciprocal time domain where an impulse has a sine wave equivalent in the time domain with a given periodicity in reciprocal seconds. So far all of this is a mathematical manipulation of the two major ways in which we may describe an event. Too often we tend to assume the universe must somehow solve the same equations we set up as explanation for the way we perceive the universe at work. Much ado, therefore, is made of the fact that many of the signals used by engineers do not have Fourier transforms, such as the sine wave, square wave, etc. The fact is that the piece of equipment had a date of manufacture and we can be certain that it will some day fail; but while it is available it can suffice perfectly well as a source of signal. The fact that a mathematically perfect sine wave does not exist in no way prevents us from speaking of the impulse response of a loudspeaker in the time domain, or what is the same thing, the frequency response in the reciprocal time domain. Both descriptions are spectra in that the event is functionally dependent upon a single-valued coordinate and is arrayed in terms of that coordinate. If we define, for any reason, a zero coordinate in one domain, we have defined the corresponding epoch in the other domain. Since each domain representation is a spectrum description we could state that this exists in two "sides." One side is that for which the coordinate is less in magnitude than the defined zero. The form of spectral description in the general case is not dependent on the coordinate chosen for the description. Thus we could, by analogy with communication practice which normally deals with frequency spectra in terms of sidebands, say that there are time-domain sidebands. The sideband phenomenon is the description of energy distribution around an epoch in one domain due to operations (e.g., modulation) performed in the other domain. This phenomenon was used by this author to solve for the form of time-delay distortion due to propagation through a dispersive absorptive medium [3].

Fourier transform relations are valid only for infinite limits of integration and work as well for predictive systems as they do for causal. There is no inherent indignation in these transforms for a world with backward running clocks. The clock direction must be found from some other condition such as energy transformation. As pointed out previously [3], this lack of time sense led some investigators to the erroneous conclusion that group delay, a single-valued property, was uniquely related to real-world clock delay for all possible systems and has led others to the equally erroneous conclusion that a uniform group delay always guarantees a distortionless system. When we consider working with causal systems, where our clocks always run forward at constant rate, we must impose a condition on the time-domain representation that is strongly analogous to what the communication engineer calls single sideband when he describes a frequency attribute. We must, in other words, say that the epoch of zero time occurs upon stimulation of the system and that no energy due to that stimulation may occur for negative (prior) time. The conditions imposed on the other domain representation, frequency domain, by this causal requirement are described as Hermitian [21]. That is, both lower and upper sidebands exist about zero frequency and the amplitude spectrum will be even symmetric about zero frequency while the phase spectrum is odd symmetric.

The mathematical simplicity of impulse (and its sinusoidal equivalent transform) calculations has led to a tremendously useful series of analytical tools. Among these are the eigenvalue solutions to the wave equation in the eleven coordinates which yield closed form [10]. However, these are mathematical expansions which, if relating to one domain wholly, may be related to the other domain only if all possible values of coordinate are assumed. Tremendous mathematical frustration has been experienced by those trying to independently manipulate expressions in the two domains without apparently realizing that each was a description of the same event. Having assumed one description, our ground rules of analysis prevent an arbitrary choice of the description in the other domain. Because the time- and frequency-domain representations are two ways of describing the same event, we should not expect that we could maintain indefinite accuracy in a time-domain representation if we obtain this from a restricted frequency-domain measurement, no matter how clever we were. If we restrict the amount of information available to us from one domain, we can reconstruct the other domain only to the extent allowed by the available information. This is another way of expressing the inter-domain dependence, known as the uncertainty principle. Later we shall consider the process of weighting a given spectrum description so as to minimize some undesirable sideband clutter when reconstructing the same information in the complementing spectral description.

When dealing with very simple systems, no difficulty is encountered in using a frequency-only or time-only representation and interpreting joint domain effects. But the very nature of the completeness of a given domain representation leads to extreme difficulty when one asks such seemingly simple questions as, "what is the time delay of a given frequency component passing through a system with nonuniform response?" A prior paper demonstrated that there is a valid third description of an event [3]. This involves a joint time–frequency characterization which can be brought into closed form if one utilizes a special primitive descriptor involving first-order and second-order all-pass transmission systems. In some ways this third description, lying as it does between the two principal descriptions, may be more readily identified with human experience. Everyone familiar with the score of a musical piece would be acutely aware of a piccolo solo which came two measures late. A frequency-only or time-only description of this musical fiasco might be difficult to interpret, even though both contain the information. Other joint domain methods have been undertaken by other investigators [5].

Applying this third description to a loudspeaker provided the model yielding time-delay distortion. It was shown that the answer to the time of emergence of a given frequency component had the surprise that at any given frequency there were multiple arrivals as a function of time. The nature of the third description was such that one could envision each frequency arrival as due to its own special perfect loudspeaker which had a frequency-dependent time delay which was single valued with frequency. If the system processing the information (in this case a loudspeaker) has a simple ordered pole and zero expansion in the frequency domain, then the arrival times for any frequency are discrete. If the expansion has branch points, then the arrival times may be a bounded distribution. The general problem for which this provides a solution is the propagation of information through a dispersive absorptive medium. Even though this characterization of information-bearing medium best fits a loudspeaker in a room, as well as most real-world propagation problems, attempts at solutions have been sparse [6], [7].

The equipment available to us to make measurements on a loudspeaker, such as oscilloscopes and spectrum analyzers, work in either of the primary domains and so do not present the third-domain results directly. This does not mean that other information processing means, perhaps even human perception of sound, work wholly in the primary domains. Within the restrictions of the uncertainty principle, which is after all a mathematical limitation imposed by our own definitions, we shall take a given frequency range and find the time delay of all loudspeaker frequency components within that range. The nature of this type of distortion is illustrated schematically in Fig. 1. If a momentary burst of energy $E(t)$ were fed a perfect loudspeaker, a similar burst of energy $E'(t)$ would be observed by O some time later due to the finite velocity of propagation c. More generally an actual loudspeaker will be observed by O to have a time smeared energy distribution $\epsilon(t)$. As far as the observer is concerned, the actual loudspeaker will have a spatial smear $\epsilon(x)$.

ENERGY, IMPULSE AND DOUBLET

Anyone familiar with analysis equipment realizes that the display of Fig. 1 will take more than some simple assembly of components. In fact, it will take a closer scrutiny of the fundamental concepts of energy, frequency, and time. The frequency-domain representation of an event is a complex quantity embodying an amplitude and a phase description. The time-domain representation of the same event may *also* be expressed as a complex quantity. The scalar representation of time-domain performance of a transmission system based on impulse excitation, which is common coinage in communication engineering [9], is the real part of a more general vector. The imaginary part of that vector is the Hilbert transform [2], [4] of the real part and is associated with a special excitation signal called a doublet by this author. For a nonturbulent (vortex free) medium wherein a vector representation is sufficient, the impulse and doublet responses completely characterize performance under conditions of superposition. For a turbulent medium one must use an additional tensor excitation which in most cases is a quadrupole. For all loudspeaker tests we will perform, we need only concern ourselves with the impulse and doublet response.

Any causal interception of information from a remote source implies an energy density associated with the actions of that source. The energy density represents the amount of useful work which could be obtained by the receiver if he were sufficiently clever. For the cases of interest in this paper, the total energy density at the point of reception is composed of a kinetic and a potential energy density component. These energy density terms relate to the instantaneous state of departure from equilibrium of the medium due to the actions of the remote source.

If we wish to evaluate the amount of total work which could be performed on an observer, whether microphone diaphragm or eardrum, at any moment, such as given in Fig. 1, then we must evaluate the instantaneous total energy density. In order to specify how much energy density is available to us from a loudspeaker, and what time it arrives at our location if it is due to a predetermined portion of the frequency spectrum, we must choose our test signal very carefully and keep track of the ground rules of the equivalence of time and frequency descriptions. We may not, for example, simply insert a narrow pulse of electrical energy into a loudspeaker, hoping that it simulates an impulse, and view the intercepted microphone signal on an oscilloscope. What must be done is to determine first what frequency range is to be of interest; then generate a signal which contains only those frequencies. By the process of generation of the excitation signal for a finite frequency band, we have defined the time epoch for this signal. Interception of the loudspeaker acoustic signal should then be made at the point of desired measurement. This interception should include both kinetic and potential energy densities. The total energy density, obtained as a sum of kinetic and potential energy densities, should then be displayed as a function of the time of interception.

The foregoing simplistic description is exactly what we shall do for actual loudspeakers. Those whose experiments are conducted in terms of the time domain only will immediately recognize that such an experiment is commonly characterized as physically unrealizable in the sense that having once started a time-only process, one cannot arbitrarily stop the clock or run it backward. In order to circumvent this apparent difficulty we shall make use of the proper relation between frequency and time descriptions. Rather than use true physical time for a measured parameter, the time metric will be obtained as a Fourier transform from a frequency-domain measurement. Because we are thus allowed to redefine the time metric to suit our measurement, it is possible to alter the time base in any manner felt suitable. The price paid is a longer physical time for a given measurement.

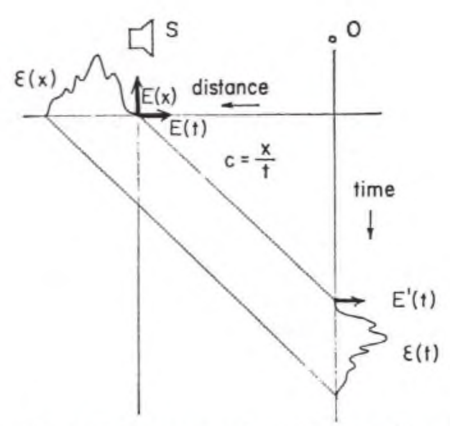

Fig. 1. Symbolic representation of time–distance world line for observer O perceiving energy from loudspeaker S.

A loudspeaker energy plot representing one millisecond may take many seconds of real clock time to process, depending upon the time resolution desired. The process utilized will coincide in large measure with some used in coherent communication practice, and the amount by which the derived time metric exceeds the real clock time will correspond to what is called filter processing gain. The basic signal process for our measurement will start with that of a time delay spectometer (TDS) [1], [16]. This is due not only to the basic simplicity of instrumentation, but the "domain swapping" properties of a TDS which presents a complex frequency measurement as a complex time signal and vice versa.

ANALYSIS, IMPULSE AND DOUBLET

If we consider that a time-dependent disturbance $f(t)$ is observed, we could say that either this was the result of a particular excitation of a general parameter $f(x)$ such that

$$f(t) = \lim_{\lambda\to\infty} \int_{-\infty}^{\infty} f(x) \frac{\sin \lambda (t-x)}{\pi (t-x)} dx \qquad (1)$$

or it was the result of operation on another parameter $F(\omega)$ such that

$$f(t) = \int_{-\infty}^{\infty} F(\omega) e^{i\omega t} d\omega. \qquad (2)$$

Eq. (1) is known as Fourier's single-integral formula [4, p. 3] and is frequently expressed as

$$f(t) = \int_{-\infty}^{\infty} f(x) \delta(t-x) dx \qquad (3)$$

where $\delta(t-x)$ is understood to mean the limiting form shown in Eq. (1) and is designated as an impulse because of its singularity behavior [10].

The functions $f(x)$ and $F(\omega)$ of Eqs. (2) and (3), Fourier's two descriptions of the same event, are thus considered to be paired coefficients in the sense that each of them multiplied by its characteristic "driving function" and integrated over all possible ranges of that driving function yields the same functional dependence $f(t)$. This paired coefficient interpretation is that expressed by Campbell and Foster in their very significant work [8].

Eq. (3) may be looked upon as implying that there was a system, perhaps a loudspeaker, which had a particular characterization $f(x)$. When acted upon by the driving signal $\delta(t-x)$, the response $f(t)$ was the resultant output. Conversely, Eq. (2) implies that there was another equally valid characterization $F(\omega)$ which when acted upon by the driving signal $e^{i\omega t}$ produced the same response $f(t)$.

The functions $f(x)$ and $F(\omega)$ are of course Fourier transforms of each other. The reason for not beginning our discussion by simply writing down the transform relations as is conventional practice is that to do so tends to overlook the real foundations of the principle. To illustrate that these functions are not the only such relations one could use, consider the same system with a driving signal which elicits the response

$$g(t) = \lim_{\lambda\to\infty} \int_{-\infty}^{\infty} f(x) \left\{ \frac{\cos \lambda (t-x) - 1}{\pi (t-x)} \right\} dx \qquad (4)$$

which as before we shall assume to exist as the limiting form

$$g(t) = \int_{-\infty}^{\infty} f(x) d(t-x) dx. \qquad (5)$$

Also,

$$g(t) = \int_{-\infty}^{\infty} F(\omega) \{ -i \operatorname{sgn}(\omega) \} e^{i\omega t} d\omega. \qquad (6)$$

Obviously this is an expression of the Hilbert transform of Eqs. (2) and (3) [2], [4]. It is none the less a legitimate paired coefficient expansion of two ways of describing the same phenomenon $g(t)$. By observing the way in which $\delta(t-x)$ and $d(t-x)$ behave as the limit is approached in Eqs. (1) and (4), it is apparent that both tend to zero everywhere except in a narrow region around the value where $t=x$. Thus there is not one, but at least two driving functions which tend toward a singularity behavior in the limit. As we shall see, these two constitute the most important set of such singularity operators when discussing physical properties of systems such as loudspeakers. Because of the nature of singularity approached by each, we shall define them as impulse and doublet, respectively. The following definitions will be assumed.

The impulse operator is approached as the defined limit

$$\delta(t) = \lim_{a\to\infty} \frac{\sin at}{\pi t}. \qquad (7)$$

The impulse operator is not a function but is defined from Eq. (3) as an operation on the function $f(x)$ to produce the value $f(t)$. $\delta(t)$ is even symmetric. The application of an electrical replica of the impulse operator to any network will produce an output defined as the impulse response of that network. The Fourier transform $F(\omega)$ of this impulse response $f(t)$ is defined as the frequency response of the network and is identical at any frequency to the complex quotient of output to input for that network when excited by a unit amplitude sine-wave signal of the given frequency.

The doublet operator is approached as the defined limit

$$d(t) = \lim_{a\to\infty} \frac{\cos at - 1}{\pi t}. \qquad (8)$$

The doublet operator is not a function but is defined from Eq. (5) as an operation on the function $f(x)$ to produce the value $g(t)$. $d(t)$ is odd symmetric. The application of an electrical replica of the doublet operator to any network will produce an output defined as the doublet response of that network. The doublet response is the Hilbert transform of the impulse response. The Fourier transform of the doublet response is identical to that of the impulse response, with the exception that the doublet phase spectrum is advanced ninety degrees for negative frequencies and retarded ninety degrees for positive frequencies.

In addition to the above definitions, the Fourier transform of the impulse response will be defined as being of minimum phase type in that the accumulation of phase lag for increasing frequency is a minimum for the resultant amplitude spectrum. The Fourier transform of the doublet response will be defined as being of non-minimum phase.

It must be observed that the doublet operator defined here is not identical to that sometimes seen derived from the impulse as a simple derivative and therefore possessing a transform of nonuniform amplitude spectral density [9, p. 542]. The corresponding relation between the doublet operator $d(t)$ and the impulse operator $\delta(t)$ prior to the limiting process is

$$d(t) = -\frac{1}{\pi} \frac{d}{d \cdot} \int_{-\infty}^{\infty} \delta(x) \ln \left| 1 - \frac{t}{x} \right| dx. \qquad (9)$$

The distinction is that the doublet operator defined here has the same power spectral density as the impulse operator. Furthermore as can be seen from Eq. (9), the doublet operator may be envisioned as the limit of a physical doublet, as defined in classical electrodynamics [10].

ANALYTIC SIGNAL

We have thus defined two system driving operators, the impulse and the doublet, which when applied to a system produce a scalar time response. Although the relation between the time-domain responses is that of Hilbert transformation, if one were to view them as an oscilloscope display, he may find it hard to believe they were attributable to the same system. However, the resultant frequency-domain representations are, except for the phase reference, identical in form. We will now develop a generalized response to show that it is not possible to derive a unique time behavior from incomplete knowledge of a restricted portion of the frequency response.

Symbolizing the operation of the Fourier integral transform by the double arrow $\leftrightarrow$, we can rewrite Eqs. (2) and (6) as the paired coefficients

$$f(t) \leftrightarrow F(\omega) \tag{10}$$

$$g(t) \leftrightarrow -i(\operatorname{sgn}\omega)F(\omega). \tag{11}$$

Multiplying Eq. (10) by a factor $\cos\lambda$ and Eq. (11) by $\sin\lambda$ and combining,

$$\cos\lambda \cdot f(t) + \sin\lambda \cdot g(t) \leftrightarrow F(\omega)[\cos\lambda - i(\operatorname{sgn}\omega)\sin\lambda]. \tag{12}$$

The frequency-domain representation is thus

$$F(\omega)e^{-i\lambda}, \qquad 0 < \omega \tag{13}$$
$$F(\omega)e^{i\lambda}, \qquad 0 > \omega.$$

In this form it is apparent that if we were to have an accurate measurement of the amplitude spectrum of the frequency response of a system, such as a loudspeaker, and did not have any information concerning its phase spectrum, we could not uniquely determine either the impulse response or doublet response of that loudspeaker. Such an amplitude-only spectrum would arise from a standard anechoic chamber measurement or from any of the power spectral density measurements using noncoherent random noise. This lack of uniqueness was pointed out in an earlier paper [2]. The best that one could do is to state that the resultant time-domain response is some linear combination of impulse and doublet response.

Because the time domain representation is a scalar, it is seen that Eq. 12 could be interpreted as a scalar operation on a generalized time-domain vector such that

$$\operatorname{Re}[e^{-i\lambda}h(t)] \leftrightarrow F(\omega) \cdot e^{-i\lambda\{\operatorname{sgn}\omega\}} \tag{14}$$

where $\operatorname{Re}(x)$ means real part of and where the vector $h(t)$ is defined as

$$h(t) = f(t) + ig(t). \tag{15}$$

This vector is commonly called the analytic signal in communication theory, where it is normally associated with narrow-band processes [11], [12]. As can be seen from (14) and the Appendix, it is not restricted to narrow-band situations but can arise quite legitimately from considerations of the whole spectrum. This analytic signal is the general time-domain vector which contains the information relating to the magnitude and partitioning of kinetic and potential energy densities.

The impulse and doublet response of a physical system, which in our case is the loudspeaker in a room, is related to the stored and dissipated energy as perceived. This means that if one wishes to evaluate the time history of energy in a loudspeaker, it is better sought from the analytic signal of Eq. (15). It is not sufficient to simply use the conventional impulse response to attempt determination of energy arrivals for speakers. The magnitude of the analytic signal is an indication of the total energy in the signal, while the phase of the analytic signal is an indication of the exchange ratio of kinetic to potential energy. The exchange ratio of kinetic to potential energy determines the upper bound for the local speed of propagation of physical influences capable of producing causal results. We call this local speed the velocity of propagation through the medium. From the basic Lagrange relations for nonconservative systems [13] it may be seen that the dissipation rate of energy is related to the time rate of change of the magnitude of the analytic signal. If the system under analysis is such as to have a source of energy at a given time and is dissipative thereafter with no further sources, the magnitude of the analytic signal will be a maximum at that time corresponding to the moment of energy input and constantly diminishing thereafter. The result of this is that if we wish to know the time history of effective signal sources which contribute to a given portion of the frequency spectrum, it may be obtained by first isolating the frequency spectrum of interest, then evaluating the analytic signal obtained from a Fourier transform of this spectrum, finally noting those portions of the magnitude of the analytic signal which are stationary with time (Hamilton's principle) [10].

It becomes apparent that by this means we will be able to take a physical system such as a loudspeaker in a room and determine not only when the direct and reflected sound arrives at a given point, but the time spread of any given arrival. Because we will be measuring the arrival time pattern for a restricted frequency range, it is important to know what tradeoffs exist because of the spectrum limitations, and how the effects can be minimized.

SPECTRUM WEIGHTING

We will be characterizing the frequency-dependent time delay of a loudspeaker. The nature of the testing signal which we use should be such that minimum energy exists outside the frequency band of interest, while at the same time allowing for a maximum resolution in the time domain. This joint domain occupancy problem has been around for quite some time and analytical solutions exist [14]. For loudspeaker testing where we wish to know the time-domain response for a restricted frequency band, we can use any signal which has a frequency spectrum bounded to the testing band with minimum energy outside this band [19], [20]. An intuitive choice of a signal with a rectangular shaped frequency spectrum which had maximum occupancy of the testing band would not be a good one, because the time-domain characterization while sharply peaked would not fall off very rapidly on either side of the peak. The consequence of this is that a genuine later arrival may be lost in the coherent sideband clutter of a strong signal. A much better choice of band-limited spectrum would be one which places more energy in the midband frequency while reducing the energy at band edge. Such a spectrum is said to be weighted. The weighting function is the frequency-dependent multiplier of the spectral components. An entire uniform spectrum is spoken of as unweighted, and a uniform bounded spectrum is said to be weighted by a rectangular function.

The proper definition of spectrum weighting must take into account both phase and amplitude. If a rectangular amplitude, minimum phase weighting is utilized, the resultant time function will be given by Eq. (1) without the limit taken. The parameter λ will be inversely proportional to the bandwidth. If rectangular amplitude but nonminimum phase weighting is utilized, then the time function will be given by Eq. (4).

It should be obvious that one can weight either the time or frequency domain. Weighting in the time domain is frequently referred to as shaping of the pulse response. The purpose in either case is to bound the resultant distribution of energy. Two types of weighting will be utilized in this paper. The first is a Hamming weighting [14, p. 98] and takes the form shown in Fig. 2**a**. As used for spectral limitation, this is an amplitude-only weighting with no resultant phase shift. Although this type of weighting cannot be generated by linear circuits, it can be obtained in an on-line processor by nonlinear means. The second weighting is a product of two functions. One function is the minimum phase amplitude and phase spectrum of a tuned circuit. The other function is shown in Fig. 2**b**. This second weighting is that utilized in a TDS which will be used as a basic instrument for this paper.

Reference to the earlier paper and its analysis discloses that within the TDS intermediate frequency amplifier (which of course could be centered at zero frequency), the information relating to a specific signal arrival is contained in the form

$$o(t) = [i(t) \otimes w(t)] \otimes S(at) \qquad (16)$$

where $\otimes$ signifies convolution, $S(at)$ is the complex

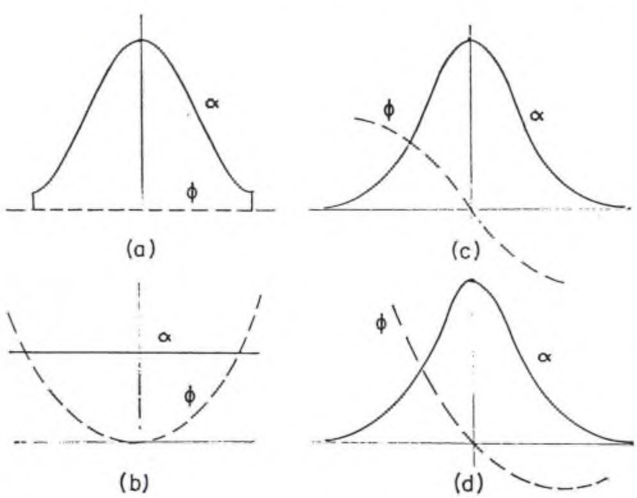

Fig. 2. Various system weighting functions including amplitude (solid line) and phase (dashed line) utilized to bound the resultant energy when taking a Fourier transform. **a.** Hamming weighting. **b.** Quadratic phase all-pass weighting. **c.** Passive simple resonance weighting. **d.** Simple TDS weighting formed as a product of **b** and **c**.

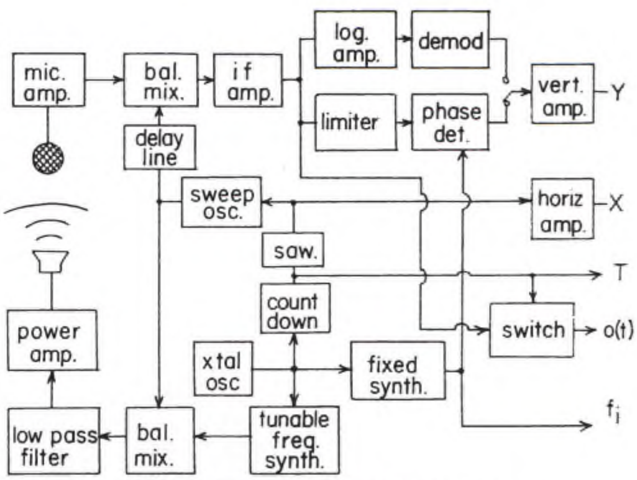

Fig. 3. Block diagram of simple TDS.

Fourier transform of the impulse response of the system under test, $i(t)$ is the impulse response of the intermediate frequency amplifier, and $w(t)$ is the window function defined as the impulse response of a quadratic phase circuit. It can be seen that the TDS provides a weighting of the time domain arrivals of that signal selected in order to give an optimum presentation of the frequency spectrum.

TRANSFORMATION TO TIME

We will now consider how to convert a measured frequency response to the analytic signal. We must start of course by having the loudspeaker frequency response available. Assume then that the system under test is evaluated by the TDS. The block diagram of this is reproduced in Fig. 3 from an earlier paper. As before, a tunable frequency synthesizer is used. Assume that the output of the intermediate frequency amplifier is taken as shown. This output signal will be of the form of Eq. (16), but of course translated to lie at the center of the intermediate frequency. Because the frequency deviation of the sweep oscillator is restricted to that portion of the frequency spectrum of interest (dc to 10 kHz, for example), the signal $o(t)$ is representative of this restricted range. The duration of sweep will be a fixed value of T seconds, so that $o(t)$ is a signal repetetive in the period T. Assume that we close the switch for $o(t)$ shown in Fig. 3 for a period of T seconds and open it prior to and following that time. The signal characterization out of this switch is

$$e^{i\omega_i t} \,|\, i(t) \otimes w(t) \otimes S(at) \,|\, \mathrm{Rect}\,(t-T) \qquad (17)$$

where the rectangular weighting function is defined as,

$$\mathrm{Rect}\,(t-T) = \begin{matrix} 1, & 0<t<T \\ 0, & \text{elsewhere.} \end{matrix} \qquad (18)$$

Assume we now multiply the signal by the complex quantity

$$e^{-i\omega_i t}\, e^{i\Omega t} \qquad (19)$$

and the product is in turn multiplied by a weighting function $A(t)$. If we take the integral of the product of these functions, we have

$$\int_{-\infty}^{\infty}[\mathrm{Rect}(t-T)]\cdot A(t)\cdot\{i(t)\otimes w(t)\otimes S(at)\}\,e^{i\Omega t}dt. \quad (20)$$

The infinite limits are possible because of the rectangular function which vanishes outside the finite time limits. The integral of Eq. (20) may now be recognized as a Fourier transform from the t domain to the Ω domain. This may be expressed in the Ω domain as

$$a(\Omega)\otimes[I(\Omega)\cdot e^{-i\Omega^2/2a}\cdot h(\Omega)] \quad (21)$$

where $a(\Omega)$ is the transform of the weighting in the t domain, $I(\Omega)$ is the frequency response of the intermediate frequency amplifier expressed in the Ω domain, the exponential form is the quadratic phase window function, and $h(\Omega)$ is the Ω domain form of the analytic function shown in Eq. (15). The reason that this is $h(\Omega)$ and not the impulse response $f(\Omega)$ is that we have assumed a TDS frequency sweep from one frequency to another, where both are on the same side of zero frequency.

What we have done by all this is instrument a technique to perform an inverse Fourier transform of a frequency response. The answer appears as a voltage which is a function of an offset frequency Ω. Even though we energized the loudspeaker with a sweeping frequency, we obtain a voltage which corresponds to what we would have had if we fed an infinitely narrow pulse through a perfect frequency-weighted filter to a loudspeaker. We now say that by changing an offset frequency Ω, the answer we see is what would have been observed had we used a pulse and evaluated the response at a particular moment in time. By adjusting the offset frequency we can observe the value that would be seen for successive moments in time.

SINGLE- AND DOUBLE-SIDED SPECTRA

We will be making measurements in the frequency domain and from this calculate the time-domain energy arrivals. If our measurement includes zero frequency, we have shown earlier how one could invoke the odd symmetry requirement to define "absolute" phase [2]. By so doing we have eliminated the parameter λ from Eq. (13). It is thus possible to calculate either the impulse or doublet time response in a unique manner. However, since the lower and upper sidebands are redundant in the frequency domain, care must be taken in using Eq. (20) that only one sideband, for example, positive frequencies, be used and the other sideband rejected. Failure to do this will result in an improper calculation not only of impulse and doublet response, but of the analytic signal as well. If one is aware of this single-sided versus double-sided spectrum pitfall, he may use it to advantage. For example, if a perfectly symmetric double-sided spectrum is used, the analytic signal calculation will yield the impulse response directly, as can be seen from Eq. (12).

Most loudspeaker measurements are made in a single-sided manner. For example, one may wish to know what time distribution arises from the midrange driver which works from 500 Hz to 10 kHz. In this case, because zero frequency is not available it may not be possible to define the parameter λ of Eq. (13), even if both amplitude and phase spectra are measured. Because of this an unequivocal determination of the impulse response (potential energy relation) or of the doublet response (kinetic energy relation) may be impossible. One may always determine the analytic signal magnitude (potential plus kinetic energy relation). The time position of effective energy sources can be determined by noting the moments when the effective signal energy is a maximum.

INSTRUMENTATION

The loudspeaker energy arrivals are obtained from the analytic signal. The signals of interest are first isolated from the remainder of the room reflections by means of a TDS. This measurement is the frequency domain description of the loudspeaker anechoic response, even though the loudspeaker is situated in a room. This loudspeaker description, although mathematically identical to a frequency-domain description, has been made available within the TDS in the time domain. In order to take a Fourier transform of this frequency-domain description to obtain the time-domain analytic signal, we must multiply by a complex sinusoid representing the time epoch, multiply this in turn by a complex weighting function, and then integrate over all possible frequencies (Eq. 20). This process would normally require substantial digital computational facilities for a frequency-domain measurement; however, the "domain swapping" properties of a TDS allow for straightforward continuous signal processing. Fig. 4 is a block diagram of the functions added to the TDS of Fig. 3 to effect this process. The signal from the intermediate frequency amplifier of the TDS is buffered and fed to two balanced mixers. By using an in-phase and quadrature multiplier of the same frequency as the intermediate frequency, the two outputs are obtained which are Hilbert transforms of each other and centered at zero frequency. Each of these is then isolated by low-pass filters and processed by identical switching multipliers controlled by a cosine function of the sweep time to effect a Hamming weighting. The net output of each weighting network is then passed through sampling integrators. At the start of a TDS sweep, the integrators are set to zero by

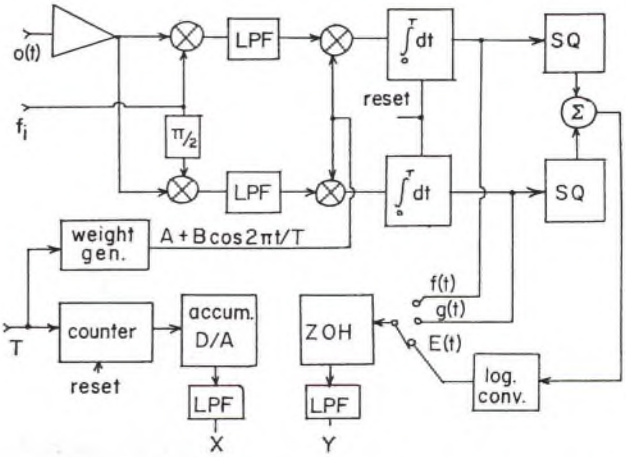

Fig. 4. Block diagram of Fourier transformation equipment attaching to a TDS capable of displaying time-domain plots. **a.** Impulse response $f(t)$. **b.** Doublet response $g(t)$. **c.** Total energy density $E(t)$.

the same clock pulse that phase locks the TDS offset frequency synthesizer so as to preserve phase continuity. Each integrator then functions unimpeded for the duration of the sweep. If the proper phase has been set into the offsetting synthesizer, the output of one integrator at the end of the sweep will correspond to the single value of the impulse response for the moment of epoch chosen, while the other will correspond to doublet response. If the proper phase has not been selected, then one integrator will correspond to a linear combination such as expressed in Eq. (12), while the other integrator will correspond to the quadrature term.

If one desires to plot the impulse or doublet response, the appropriate integrator output may be sent to a zero-order hold circuit which clocks in the calculated value and retains it during the subsequent sweep calculation. This boxcar voltage may then be recorded as the ordinate on a plotter with the abscissa proportional to the epoch. One trick of the trade which is used when horizontal and vertical signals to a plotter are stepped simultaneously, is to low-pass filter both channels with the same cutoff frequency. The plotter will now draw straight lines between interconnecting points.

If one is interested in the magnitude of the analytic signal (14), which from the Appendix, to be published December 1971, is related to energy history, then the most straightforward instrumentation is to square the output from each integrator and linearly add to get the sum of squares. A logarithmic amplifier following the sum of squares will enable a signal strength reading in dB without the need for a square root circuit. By doing this, a burden is placed on this logarithmic amplifier since a 40-dB signal strength variation produces an 80-dB input change to the logarithmic element. Fortunately, such an enormous range may be accommodated readily by conventional logarithmic elements. For graphic recordings of energy arrival, the output of the logarithmic amplifier may be fed through the same zero-order hold circuit as utilized for impulse and doublet response.

The configuration of Fig. 4 including quadrature multipliers, sampled integrators, and sum of squares circuitry is quite often encountered in coherent communication practice [17], [18]. This circuit is known to be an optimum detector in the mean error sense for coherent signals in a uniformly random noise environment. Its use in this paper is that of implementing an inverse Fourier transform for total energy for a single-sided spectrum. An interesting byproduct of its use is thus an assurance that no analytically superior instrumentation as yet exists for extracting the coherent loudspeaker signal from a random room noise environment.

CHOICE OF MICROPHONE

The information which our coherent analysis equipment utilizes is related to the energy density intercepted by the microphone. The total energy density in joules per cubic meter is composed of kinetic energy density E_T and potential energy density E_V, where [22, p. 356]

$$E_T = \frac{1}{2} \rho_0 v^2$$

$$E_V = \frac{1}{2} \rho_0 c^2 s^2. \qquad (22)$$

The equilibrium density is ρ_0, v is the particle velocity, s is condensation or density deviation from equilibrium, and c is velocity of energy propagation.

At first glance it might be assumed that total energy may not be obtained from either a pressure responsive microphone which relates to E_V or a velocity microphone relating to E_T. The answer to this dilemma may be found in the Appendix. One can always determine one energy component, given the other. Hence a determination of acoustic pressure or velocity or an appropriate mixture of pressure and velocity is sufficient to characterize the energy density of the original signal.

This means that any microphone, whether pressure, velocity, or hybrid, may be used for this testing technique, provided that a calibration exists over the frequency range for a given parameter. This also means that any perceptor which is activated by total work done on it by the acoustic signal will not be particular, whether the energy bearing the information is kinetic or potential. There is some reason to believe that human sound preception falls into this category.

REFERENCES

[1] R. C. Heyser, "Acoustical Measurements by Time Delay Spectrometry," *J. Audio Eng. Soc.*, vol. 15, p. 370 (1967).

[2] R. C. Heyser, "Loudspeaker Phase Characteristics and Time Delay Distortion: Part 1," *J. Audio Eng. Soc.*, vol. 17, p. 30 (1969).

[3] R. C. Heyser, "Loudspeaker Phase Characteristics and Time Delay Distortion: Part 2," *J. Audio Eng. Soc.*, vol. 17, p. 130 (1969).

[4] E. C. Titchmarsh, *Introduction to the Theory of Fourier Integrals* (Oxford Press, London, 2nd ed., 1948).

[5] C. H. Page, "Instantaneous Power Spectra," *J. Appl. Phys.*, vol. 23, p. 103 (1952).

[6] L. Brillouin, *Wave Propagation and Group Velocity* (Academic Press, New York, 1960).

[7] H. G. Baerwald, "'Über die Fortpflanzung von Signalen in dispergierenden Systemen," *Ann. Phys.*, vol. 5, p. 295 (1930).

[8] G. A. Campbell and R. M. Foster, *Fourier Integrals* (D. Van Nostrand, Princeton, N.J., 1961).

[9] E. A. Guillemin, *The Mathematics of Circuit Analysis* (M.I.T. Press, Cambridge, Mass., 1965).

[10] E. U. Condon and H. Odishaw, *Handbook of Physics* (McGraw-Hill, New York, 1967).

[11] R. S. Berkowitz, *Modern Radar* (John Wiley, New York, 1967).

[12] L. E. Franks, *Signal Theory* (Prentice-Hall, Englewood Cliffs, N.J., 1969).

[13] T. v. Karman and M. A. Biot, *Mathematical Methods in Engineering* (McGraw-Hill, New York, 1940).

[14] R. B. Blackman and J. W. Tukey, *Measurement of Power Spectra* (Dover, New York, 1959).

[15] A. W. Rihaczek, *Principles of High-Resolution Radar* (McGraw-Hill, New York, 1969).

[16] R. C. Heyser, "Time Delay Spectrometer." U.S. Patent 3,466,652, Sept. 9, 1969.

[17] A. J. Viterbi, *Principles of Coherent Communication* (McGraw-Hill, New York, 1966).

[18] H. L. VanTrees, *Detection, Estimation, and Modulation Theory* (John Wiley, New York, 1968).

[19] T. A. Saponas, R. C. Matson, and J. R. Ashley, "Plain and Fancy Test Signals for Music Reproduction Systems," *J. Audio Eng. Soc.*, vol. 19, pp. 294-305 (Apr. 1971).

[20] A. Schaumberger, "Impulse Measurement Techniques for Quality Determination in Hi-Fi Equipment, with Special Emphasis on Loudspeakers," *J. Audio Eng. Soc.*, vol. 19, pp. 101-107 (Feb. 1971).

[21] R. Bracewell, *The Fourier Transform and Its Applications* (McGraw-Hill, New York, 1965).
[22] D. H. Menzel, *Fundamental Formulas of Physics* (Prentice-Hall, Englewood Cliffs, N.J., 1965).
[23] E. A. Guillemin, *Theory of Linear Physical Systems* (John Wiley, New York, 1963).
[24] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series and Products* (Academic Press, New York, 1965), Eq. 3.782.2.
[25] B. B. Baker and E. T. Copson, *The Mathematical Theory of Huygens' Principle* (Oxford Press, London, 1953).

THE AUTHOR

Richard C. Heyser received his B.S.E.E. degree from the University of Arizona in 1953. Awarded the AIEE Charles LeGeyt Fortescue Fellowship for advanced studies he received his M.S.E.E. from the California Institute of Technology in 1954. The following two years were spent in post-graduate work at Cal Tech leading toward a doctorate. During the summer months of 1954 and 1955, Mr. Heyser was a research engineer specializing in transistor circuits with the Motorola Research Laboratory, Phoenix, Arizona. From 1956 until the present time he has been associated with the California Institute of Technology Jet Propulsion Laboratory in Pasadena, California where he is a senior member of the JPL Technical Staff.

Mr. Heyser has presented several papers before the AES and is a member and Fellow of the Audio Engineering Society, the Institute of Electrical and Electronic Engineers, as well as Tau Beta Pi, Pi Mu Epsilon, Phi Kappa Phi, Sigma Pi Sigma, and Sigma Xi. He is also a member of Theta Tau.

Reprinted from *JAES*, vol. 19, no. 10, pp. 829–834, 1971 November.

Determination of Loudspeaker Signal Arrival Times*
Part II

RICHARD C. HEYSER

Jet Propulsion Laboratory, California Institute of Technology, Pasadena, Calif.

EXPERIMENT

The information to be determined is the time delay of total acoustic energy that would be received from a loudspeaker if fed from an impulse of electrical energy. Because we are interested in that energy due to a preselected portion of the frequency band, we may assume that the impulse is band limited by a special shaping filter prior to being sent to the loudspeaker. This filter would not be physically realizable if we actually used an impulse for our test; but since we are using a method of coherent communication technology, we will be able to circumvent that obstacle. Fig. 5 shows three responses for a midrange horn loaded loudspeaker. The frequency band is dc to 10 kHz, and each response is measured on the same time scale with zero milliseconds corresponding to the moment of speaker excitation. The driver unit was three feet from the microphone. Curve (a) is the measured impulse response and is what one would see for microphone pressure response, had the loudspeaker been driven by a voltage impulse. Curve (b) is the measured doublet response and is the Hilbert transform of (a). In both (a) and (b) the measured ordinate is linear voltage. Curve (c) is the total received energy on a logarithmic scale. Here the interplay of impulse, doublet, and total energy is evident.

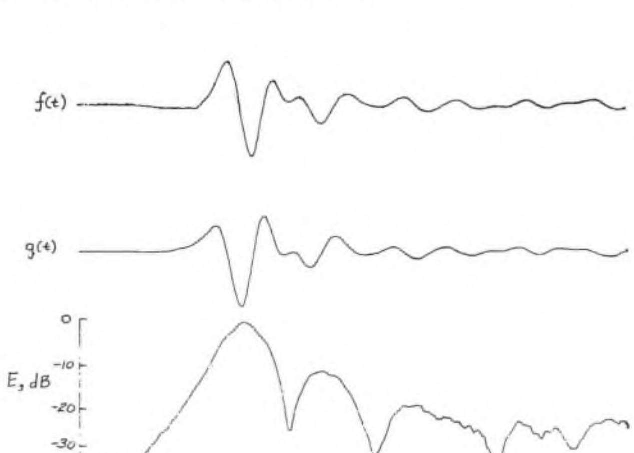

Fig. 5. Measured plots of impulse response $f(t)$, doublet response $g(t)$, and total energy density E for a midrange horn loudspeaker for spectral components from dc to 10 kHz.

* Presented April 30, 1971, at the 40th Convention of the Audio Engineering Society, Los Angeles. For Part I, please see pp. 734-743 of the October 1971 issue of this *Journal*.

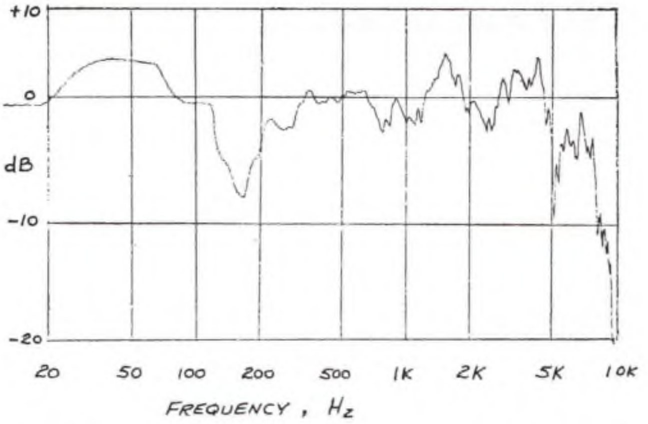

Fig. 6. TDS plot of frequency response (amplitude only) of an eight-inch open-back cabinet mounted loudspeaker with microphone to cone air path spacing of three feet.

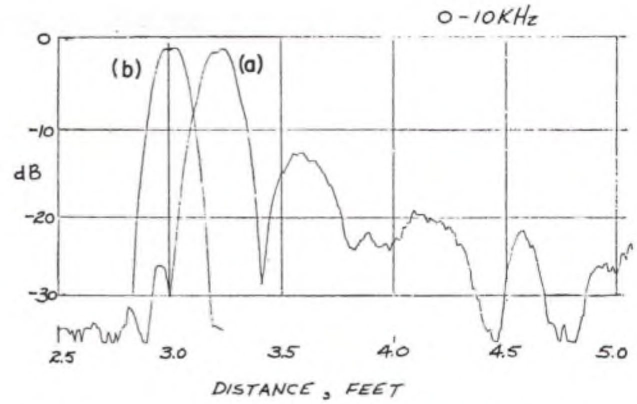

Fig. 7. Curve (a)—Energy-time arrival for loudspeaker of Fig. 6, taking all frequency components from dc to 10 kHz. Curve (b)—Superimposed measured curve to be expected if loudspeaker did not have time-delay distortion.

Fig. 6 is the TDS measured amplitude frequency response of a good quality eight-inch loudspeaker mounted in a small open-back cabinet. For simplicity the phase spectrum is not included. Fig. 7 is the time delay of energy for the speaker of Fig. 6. Superimposed on this record is a plot of what the time response would have been, had the actual loudspeaker position and acoustic position coincided and if there were no time-delay distortion. It is clear from this record that time-delay distortion truly exists. If one considers all response within 20 dB of peak, it is evident that this loudspeaker is smeared out by about one foot behind its apparent physical location. It should be observed that the response dropoff above 5 kHz coincides with a gross time delay of about three inches, as predicted by earlier analysis (Part I, [2]).

Fig. 8 is a plot of energy versus equivalent distance in feet for a high-efficiency midrange horn loaded driver. The band covered is 500 Hz to 1500 Hz and includes the region from low-frequency cutoff to midrange. The physical location of the driver phase plug is shown, and it may be seen that the acoustic and physical positions differ by nearly one foot. Fig. 9 is the same driver, but the band is 1000 Hz to 2000 Hz. The acoustic position is now closer to the phase plug and a hint of a double hump in delay is evident. Because the bandwidth is 1000 Hz, the spatial resolution available does not allow for more complete definition of acoustic position for this type of display.

Fig. 10 is a data run on an eight-inch wide range loudspeaker without baffle. A scaled pseudo cross section of the loudspeaker is shown for reference. The frequency range covered is dc to 20 kHz. Although the time-delay value may vary from one part of the spectrum to another, it is apparent that a wide frequency range percussive signal may suffer a spatial smear of the order of six inches.

Fig. 11 is a medium-quality six-inch loudspeaker mounted in an open-back cabinet. The position of main energy is, as predicted, quite close to that which would be assumed for a cutoff at about 5 kHz. The secondary hump of energy from 3 to 3.5 milliseconds is not due to acoustic energy spilling around the side of the enclosure, but is a time delay inherent in the loudspeaker itself.

Fig. 12 is the energy received from an unterminated midrange driver excited from dc to 10 kHz. The effect of untermination is seen as a superposition of an exponential time delay, due to a relatively high Q resonance, and internal reverberation with a 0.3-millisecond period.

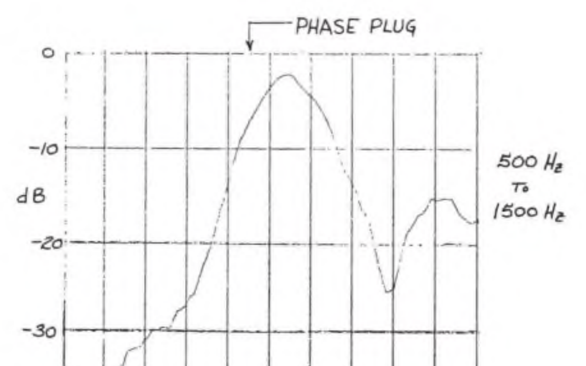

Fig. 8. Energy-time arrival for high-quality midrange horn loudspeaker for all components from 500 Hz to 1500 Hz. Measured position of phase plug is shown.

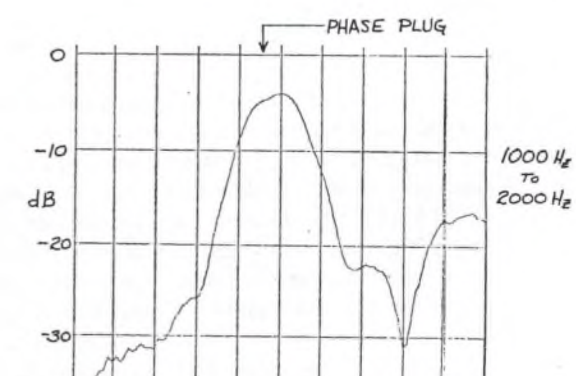

Fig. 9. Same loudspeaker as Fig. 8, but excitation is from 1000 Hz to 2000 Hz.

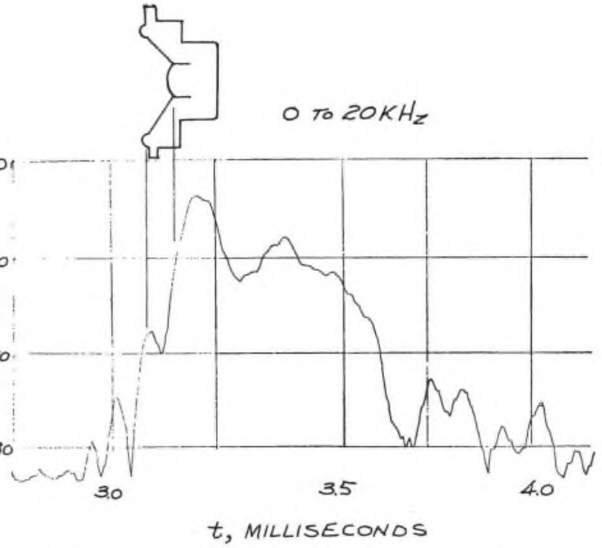

Fig. 10. Energy-time arrivals for unbaffled high-quality eight-inch loudspeaker. Frequency band is dc to 20 kHz and phantom sketch of loudspeaker physical location is included for identification of amount of time-delay distortion relative to speaker dimensions.

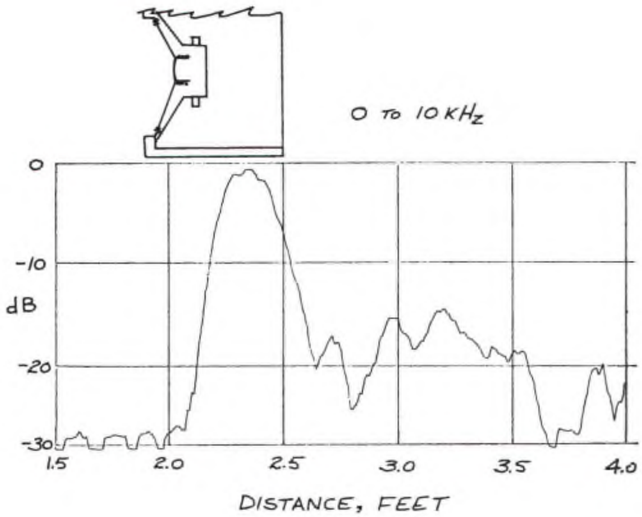

Fig. 11. Energy-time arrivals for open-back cabinet mounted medium-quality eight-inch loudspeaker. Frequency band is dc to 10 kHz and loudspeaker position shown to approximate scale.

Fig. 13 shows the effect of improper termination by a horn with too high a flare rate. The resonance is more efficiently damped, but the internal reverberation due to acoustic mismatch still exists. This is a low-quality driver unit.

Fig. 14 is the time delay distortion of the midrange driver discussed at some length elsewhere (Part I [2, Fig. 4]. The internal delayed voices are plainly in evidence.

Fig. 15 is a high-quality paper cone tweeter showing the time-delay distortion for the dc to 10-kHz frequency range. The multiplicity of reverberent energy peaks with about a 0.13-millisecond period is due to internal scattering within the tweeter. It is not at all clear from the frequency response taken alone that such an effect exists; however, by observing the time-delay characteristic it is possible to know what indicators to look for upon reexamination of the complete frequency response.

Fig. 16 is the time display of a multiple-panel high-quality electrostatic loudspeaker. This is a 1–5-kHz response taken along the geometric axis of symmetry, coinciding with the on-axis response. The physical position of the closest portion of radiating element occurs at a distance equivalent to 2 milliseconds air path delay. Fig. 17 is the same speaker 15 degrees off axis. Not only is the total energy down, but the contribution of adjacent panels is now evident.

Figs. 18 and 19 are dc to 25-kHz on-axis and 15-degree off-axis runs on a high-quality horn loaded compression tweeter. The positions of mouth, throat, and voice coil are shown in the on-axis record and several interesting effects are observable which do not show up in normal analysis. There appears to be a small acoustic contribution due to the horn mouth. This effect has been repeatedly seen by this author in such units. One possible explanation is that a compressional or shear body wave is actually introduced in the material of the horn (or cone in direct radiators) which travels at least as fast as the air compressional wave and causes an acoustic radiation from the bell of the horn itself. Also an

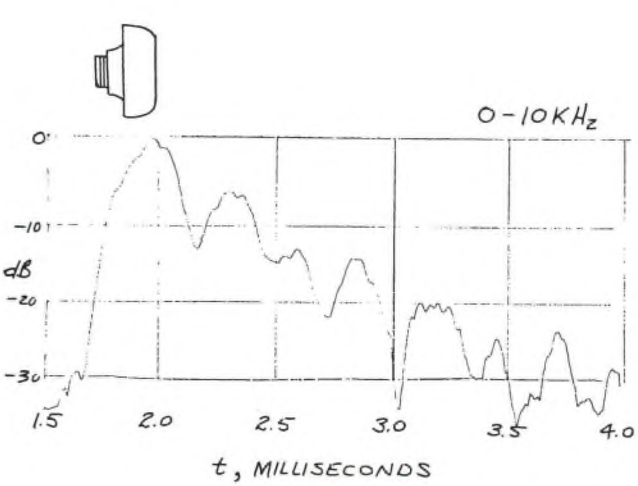

Fig. 12. Energy-time arrivals of unterminated low-quality midrange driver unit. Frequency range is dc to 10 kHz and physical position of driver shown.

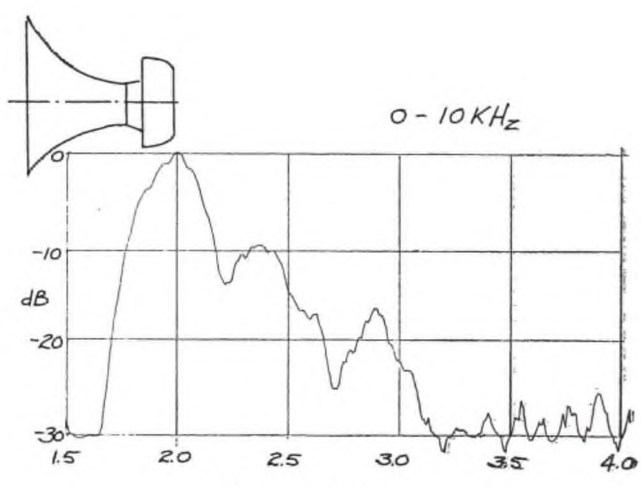

Fig. 13. Energy-time arrivals for improperly terminated driver unit of Fig. 12.

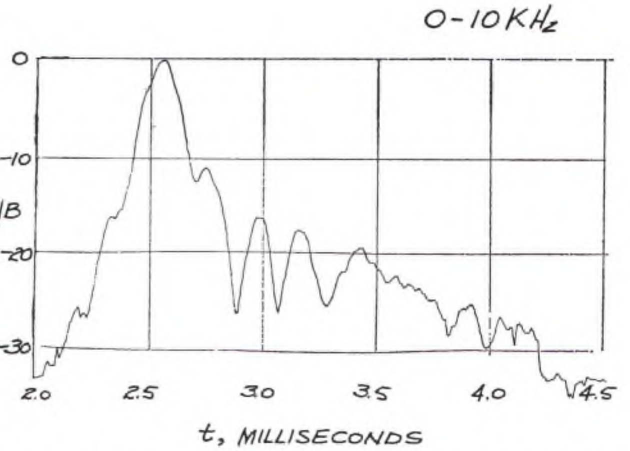

Fig. 14. Dc to 10 kHz energy-time arrivals for midrange horn loaded loudspeaker exhibiting distinct nonminimum phase frequency response.

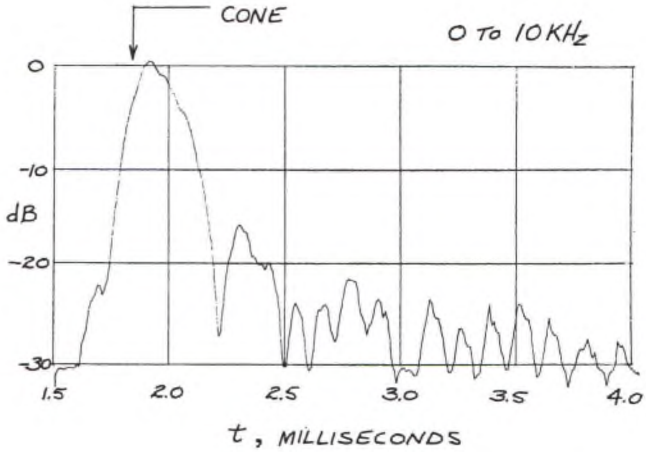

Fig. 15 Dc to 10 kHz energy-time arrivals of paper cone tweeter exhibiting distinct reverberation characteristic.

internal reverberation is observable following emergence of the main loudspeaker energy. This reverberation appears to be due to acoustic scattering off the sides of the internal structure of the horn itself. This may be inferred from the 0.12-millisecond period seen in Fig. 18, which coincides with the on-axis geometry, together with the replacement by a different behavior 15 degrees off axis as seen in Fig. 19. This suggests that closer attention might be paid to the details of mechanical layout of such horns whose acoustic properties may have been compromised for improved cosmetic appeal.

It has been noted by several authors that a network which introduces frequency-dependent phase shift, only without amplitude variation, quite often cannot be detected in an audio circuit, even when the phase shift is quite substantial. Because such networks create severe waveform distortion for transient signals while not apparently effecting the listening quality of such signals, it is assumed by inference that phase distortion must be inaudible for most systems. Fig. 20 is a measurement made through a nominal 2-millisecond electrical delay line with and without a series all-pass lattice. The network used is a passive four-terminal second-order lattice with a 1-kHz frequency of maximum phase rate. The frequency range is dc to 5 kHz, and an electronic delay is used to show the overall time delay on a scale comparable to that used for loudspeaker measurements. Although the lattice does indeed severely disturb the impulse response waveform, it is interesting to note that the total energy is not greatly effected when one considers a reasonable band of frequencies. Since this time-delay distortion, which agrees with calculated values, is due to an analytically perfect signal, it is not at all unlikely that a multimiked program heard over any loudspeaker possessing the degree of time-delay distortion measured in this paper would not appear to show this particular phase-only distortion. In view of the amount of time-delay distortion evident in most loudspeakers, it might be presumptuous to assume that this effect is totally inaudible in all systems.

SUMMARY AND CONCLUSION

A ground rule has been utilized in assessing the linear performance of a loudspeaker in a room. This rule is that the quality of performance may be associated with the accuracy with which the direct sound wave at the position of an observer duplicates the electrical signals presented to the loudspeaker terminals. Although it is realized that there are many criteria of performance,

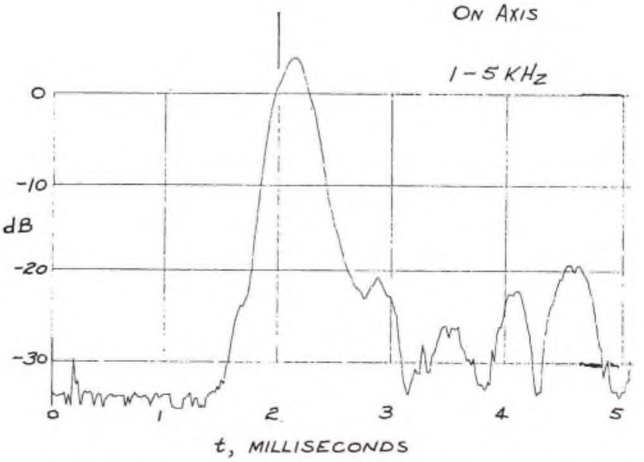

Fig. 16. On-axis energy-time response of high-quality electrostatic multi-panel midrange speaker with position of closest panel equivalent to air path delay of 2 milliseconds and 1–5-kHz excitation.

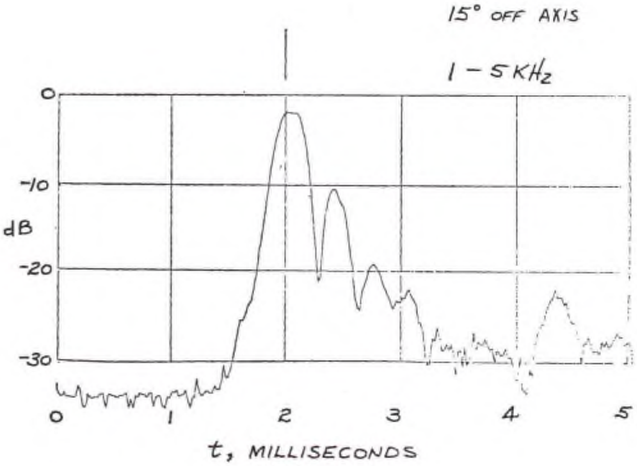

Fig. 17. 15-degree off-axis energy-time response of electrostatic loudspeaker of Fig. 16.

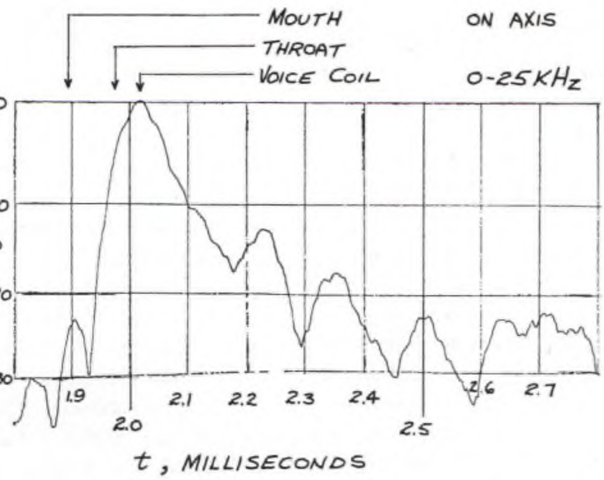

Fig. 18. On-axis energy-time response of a high-quality horn loaded compression tweeter. Frequency excitation is dc to 25 kHz and positions of mouth, throat, and voice coil shown.

this assumption of equivalence of acoustic effect resulting from an electrical cause has the advantage that it yields to objective analysis and test. The difference between the total sound due to a loudspeaker in a room and the same loudspeaker in an anechoic environment, for example, may be simplified to the following model. In an anechoic environment we have one loudspeaker at a fixed range, azimuth, and elevation with respect to an observer. In a room we have the original anechoic

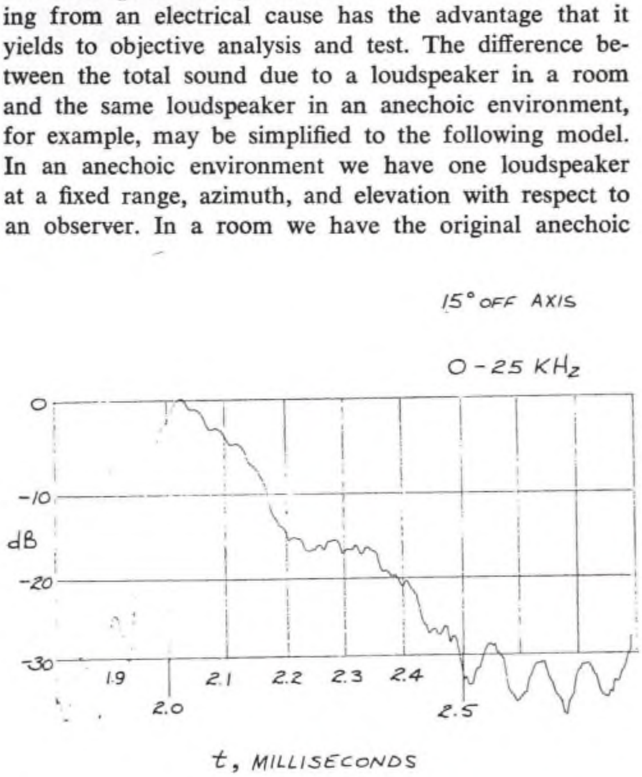

Fig. 19. 15-degree off-axis energy-time response of speaker of Fig. 18.

loudspeaker, but in addition we have a multiplicity of equivalent loudspeakers assuming various positions of range, azimuth, and elevation. The additional loudspeakers, in this room model, all have the same program material as the anechoic loudspeaker, but of course suffer time delays in excess of the direct path delay of the anechoic loudspeaker. Also each room model loudspeaker has a frequency response unique to itself. The groundrule of loudspeaker quality may be applied to each equivalent source in turn and the composite effect analyzed for total quality of response in the room.

A purely mathematical analysis of any single loudspeaker in this room model disclosed that there is a direct tie between frequency response and time smear of signal received by an observer. The analysis showed that if we were to isolate any speaker to an anechoic environment, we could duplicate the acoustic response as closely as we desired for any given observer by replacing the original speaker and its frequency response aberrations with a number of perfect response loudspeakers. Each of these perfect response loudspeakers in this mathematical model occupies its own special frequency-dependent position in space behind the apparent physical position of the original imperfect loudspeaker. The result of this is that the acoustic image of a sound source is smeared in space behind the originating speaker. Perhaps another way of looking at this is that even in an otherwise anechoic environment an actual loudspeaker could be considered to be a perfect transducer imbedded in its own special "room" which creates an ensemble of equivalent sources. The type of distortion caused by this multiplicity of delayed equivalent sources has been called time-delay distortion.

A measurement of the amount of time-delay distortion in an actual loudspeaker in a room has now been made. The anechoic frequency response, both amplitude and phase, was first isolated by time-delay spectrometry for the specific portion of frequency spectrum of interest. The complex frequency response was then processed by real-time continuous circuitry to yield the complex time response.

Plots of the complex time vector components as a function of equivalent time of arrival for a variety of loudspeakers have been presented. The existence of time-delay distortion has been verified by this direct experimental evidence. It has been shown that the equivalent spatial smear for even the better class of loudspeaker may amount to many inches and that the equivalent acoustic source is always behind the apparent physical source location. It has not been possible to plot the individual joint time–frequency components predicted mathematically. This is because these components overlap in the time and frequency domains and a single-domain time presentation, even though band limited, cannot separate simultaneous arrival components. Sufficient experimental evidence has been presented to show that these components do exist to an extent necessary to create

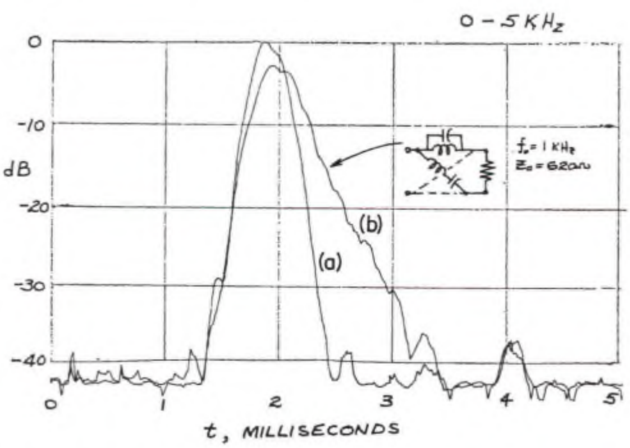

Fig. 20. Energy-time response. Curve (a)—Electrical delay line with 2-millisecond delay and excitation from dc to 5 kHz. Curve (b)—Delay line of (a) in series with second-order all-pass lattice which exhibits severe impulse response distortion due to rapid phase shift at 1 kHz.

the acoustic image smear detected by an observer.

Several energy principles have been originated and proved. While originally developed to determine techniques for investigating time-delay distortion, these principles reach far beyond simple loudspeaker testing. It has been shown that the unit impulse is but one component of a more generalized tensor. For nonturbulent systems the tensor becomes a simple two-component vector. This is the case for most acoustic and electronic situations of energy propagation. The conjugate term to the unit impulse is the unit doublet. In an acoustic field generated by a loudspeaker, one can associate the potential energy density with the impulse response of the loudspeaker. When one does this he may then associate the kinetic energy density with the loudspeaker doublet response. The total energy density may be associated with the vector sum of impulse and doublet response. Inasmuch as it is the total energy density which is available to perform work on an eardrum or microphone diaphragm, the majority of experimental data presented in this paper has been the time of arrival of this parameter.

It has also been shown that potential and kinetic energy densities are not mathematically independent if one is careful with his energy bookkeeping. What this means for acoustic radiation from a loudspeaker is that either the impulse or doublet response is sufficient to determine total performance if one has the proper tools at his disposal. But one should be cautious of gross simplification in the event that impulse or doublet response is utilized independently. As with any incomplete analysis, certain truths may not be self-evident.

An interesting area of speculation is opened up when one realizes that any reasonably well-behaved acoustic transducer placed in a sound field is capable of yielding information concerning the total energy density if associated with a suitable means of data processing. One cannot help but incautiously suggest that a closer look at the human hearing mechanism might be justified to determine whether total sound energy detection rather than potential energy (pressure) could shed a light on some as yet unexplained capabilities we seem to possess in the perception of sound.

Note: Mr. Heyser's biography appeared in the October 1971 issue.

Reprinted from *JAES*, vol. 19, no. 11, pp. 902–905, 1971 December.

Determination of Loudspeaker Signal Arrival Times* Part III

RICHARD C. HEYSER

Jet Propulsion Laboratory, California Institute of Technology, Pasadena, Calif.

APPENDIX

Energy Relations as Hilbert Transforms

A fundamental approach to a complicated system may be made through that system's energy relations. Accordingly we present the following principles.

1) In a bounded system the internal energy density E is related to its potential and kinetic energy density components V and T by the vector relation

$$\sqrt{E} = \sqrt{V} + i\sqrt{T}$$

where the vector components are Hilbert transforms of each other.

2) In a bounded system a complete description of either the kinetic or potential energy density is sufficient to determine the total internal energy density.

3) By appropriate choice of coordinates within a bounded system, the available energy at a point of perception due to a signal source at a point of transmission may be partitioned as follows.

a) The potential energy density is proportional to the square of the convolution integral of the signal with the system impulse response.

b) The kinetic energy density is proportional to the square of the convolution integral of the signal with the system doublet response.

The first law of thermodynamics defines an exact differential function known as the internal energy (Part I, [10])

$$dE = dQ - dW, \qquad \text{joules} \qquad (23)$$

which equals the heat absorbed by the system less the work done by the system. By integration the energy may be obtained as a function of the state variables, and in particular for the class of electroacoustic situations of concern for this paper, it may be composed of kinetic energy and potential energy T and V,

$$E = T + V, \qquad \text{joules.} \qquad (24)$$

By taking the time rate of change of the components of (23) and expressing this in engineering terms, we have (Part I, [23, p. 124]).

$$\frac{d(T + V)}{dt} = P - 2F, \qquad \text{watts} \qquad (25)$$

which asserts that the time rate of change of energy equals the power drawn from the system less the energy dissipated as heat within the system.

Properly speaking, the internal energy of a system is that property which is changed as a causal result of work done on or by that system. Energy, per se, is not generally measured. We may, however, describe and measure the energy density. Energy density is a measure of the instantaneous work which is available to be done by a system at a particular point in space and time if the total energy partitioned among the state variables

* Presented April 30, 1971, at the 40th Convention of the Audio Engineering Society, Los Angeles. For Parts I and II, please see pp. 734-743 and pp. 829-834 of the October and November issues of this *Journal*.

could be annihilated. Energy density for state variables s is expressed as $E(s)$ and has the dimensions of joules per unit of s. The energy densities of joules per second and joules per cubic meter will be utilized in this paper. Energy density may be partitioned, for nonturbulent systems, into kinetic and potential densities. The methods by which we measure energy density, even for acoustic systems, may take the form of mechanical, electrical, or chemical means. The dynamical considerations which gave rise to Eqs. (23) and (24) naturally led to the terms kinetic and potential. When dealing with electrical or chemical characterizations, such terms are difficult to identify with the processes involved. This author has found it convenient to identify potential energy as the energy of coordinate configuration and kinetic energy as the energy of coordinate transformation.

Assume that the ratio of total kinetic to potential energy density at any moment is related to a parameter θ such that

$$\sqrt{T} / \sqrt{V} = \tan \theta. \quad (26)$$

From (24) and (26) it is possible to define the vector

$$\sqrt{E} = \sqrt{V} + i \sqrt{T}. \quad (27)$$

This is shown in Fig. A-1. We know from physical considerations that the internal energy of any bounded system is not only finite but traceable to a reasonable distribution of energy sources and sinks. If, for example, we measure the acoustic field radiated from a loudspeaker, we know that the value of that field at any point does not depend upon the way in which we defined our coordinate system. We can state, therefore, that $\sqrt{E}$ is analytic in the parameter t and is of class $L^2\ (-\infty, \infty)$ such that

$$\varepsilon = \int_{-\infty}^{\infty} |\sqrt{E}|^2 \, dt < \infty. \quad (28)$$

When conditions (27) and (28) are met it is known that the vector components of (27) are related by Hilbert transformation (Part I, [4, p. 122]). Furthermore,

$$\int_{-\infty}^{\infty} (\sqrt{T})^2 \, dt = \int_{-\infty}^{\infty} (\sqrt{V})^2 \, dt = \varepsilon/2 \quad (29)$$

which means that not only is it possible to express the kinetic and potential energy determining time components as Hilbert transforms, but when all time is considered, there is an equipartition of energy.

The relationship between kinetic and potential energy density is true for a bounded system, that is, one in which a boundary may be envisioned of such an extent as to totally enclose at any moment the total energy due to a particular signal of interest. A proper summation of the energy terms within that boundary for the signal of interest would then disclose a partitioning in accordance with principle 1). A measurement of the energy density at a microphone location due to a remote source will only yield a part of the total energy density of that source. The relation (27) will therefore not necessarily be observed by the microphone at any given moment. Thus, for example, the pressure and velocity components at a point in an expanding sound wave from a source will be related by what is called the acoustic impedance

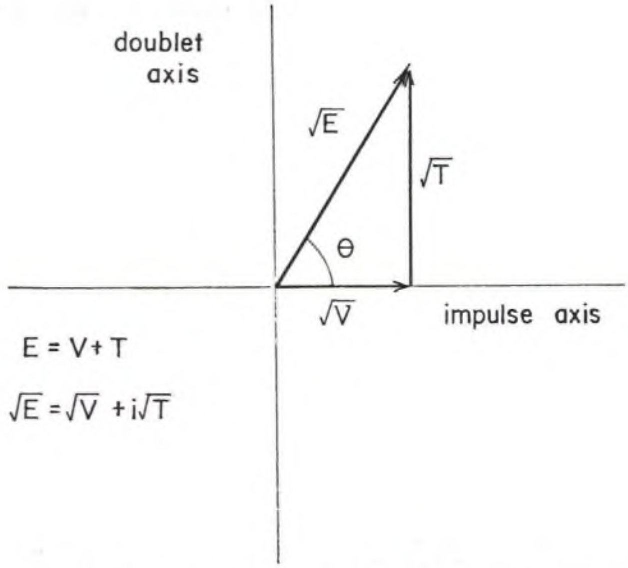

Fig. A-1. Root energy density plane defined such that one axis is system impulse response while quadrature (Hilbert) axis is doublet response.

of the medium and are not necessarily at that point related by Hilbert transformation. However, we know that the source of sound was, at the moment of energization, a bounded system and was therefore governed by the physics of (27). If the medium of propagation is such that a given energy component imparted by the source is preserved in form between source and microphone, we may take that microphone measurement and reconstruct the total energy-time profile of the source by analytical means. This observation is the basis for the measurements of this paper.

Any vector obtained from (Eq. 27) by a process of rotation of coordinates must possess the same properties. This surprisingly enough is fortunate, for although from dynamical arguments the components shown in (Eq. 27) are the most significant, it quite frequently happens that an experiment may be unable to isolate a purely kinetic component. This does not inhibit a system analysis based on total energy since we can obtain the total vector by adding our measured quantity to a quadrature Hilbert transform and be assured of a proper answer.

For verification of principle 3) we must consider that class of kinetic and potential energy related signals which could serve as stimulus to a system for resultant analysis. In particular we seek a signal form which when used as a system stimulus will suffice to define within a proportionality constant the system vector (27) by an integral process. This is done so as to parallel the analytical techniques which use a Green's function solution to an impulse (Part 1, [10]) and of course the powerful Dirac delta. Because we are dealing with quadrature terms we have not one but two possible energy stimuli. Consider the special representation of (27),

$$\sqrt{V(x)} = \frac{1}{\sqrt{2\pi a}} \frac{\sin ax}{x} \quad (30)$$

$$\sqrt{T(x)} = \frac{1}{\sqrt{2\pi a}} \frac{\cos ax - 1}{x}.$$

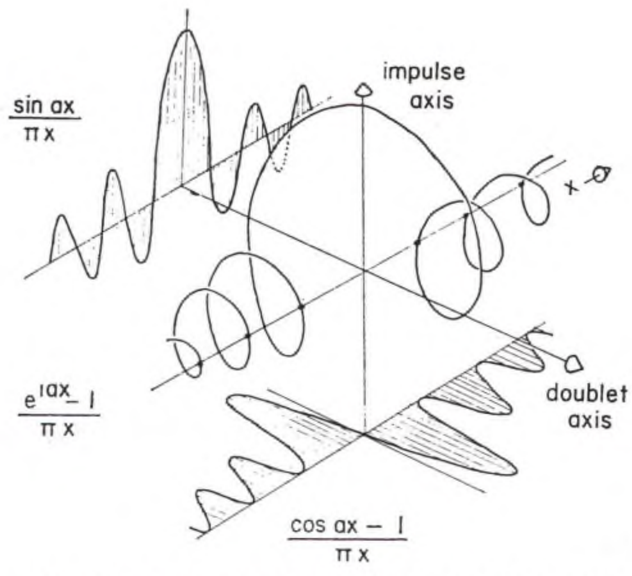

Fig. A-2. Sketch of defined complex energy vector prior to allowing the parameter a to become large without limit. Impulse (7) and doublet (8) are shown as orthogonal projections from this vector.

The energy density represented by this is obtained from (24) as

$$E(x) = \frac{\sin^2 ax + 1 - 2\cos ax + \cos^2 ax}{2\pi a x^2} = \frac{1 - \cos ax}{\pi a x^2}. \tag{31}$$

The total energy represented by (31) as a becomes large without limit is (Part I, [24])

$$\varepsilon = \lim_{a\to\infty} \int_{-\infty}^{\infty} \frac{1 - \cos ax}{\pi a x^2}\, dx = 1. \tag{32}$$

Thus in the limit the quadrature terms of (30) produce a representation of unit total energy which exists only for $x = 0$ and is null elsewhere. To see this more clearly rewrite (27) with (30) components as

$$\epsilon(y) = \frac{1}{\sqrt{2\pi}} \frac{\sin \sqrt{a} y}{y} + i \frac{1}{\sqrt{2\pi}} \frac{\cos \sqrt{a} y - 1}{y} \tag{33}$$

where $y = \sqrt{a}\, x$. The vector (33) is as shown in Fig. A-2 with its quadrature components as projections. If we took the limiting form of (33) as $\sqrt{a}$ became large without limit, this would approach the impulsive vector

$$\epsilon(y) = \delta(y) + i d(y) \tag{34}$$

where by definition

$$\delta(y) = \text{unit impulse} = \lim_{\lambda\to\infty} \frac{\sin \lambda y}{\pi y}$$

$$d(y) = \text{unit doublet} = \lim_{\lambda\to\infty} \frac{\cos \lambda y - 1}{\pi y} \tag{35}$$

This impulsive vector is symbolized in Fig. A-3 as its quadrature projections. It may be readily seen that $\delta(y)$ is identical to the impulse commonly referred to as the Dirac delta (Part I, [10, p. I-168]). To this author's knowledge this particular unit doublet has not received previous recognition.

In order to justify the designation of the energy-related vector $\epsilon(y)$ as impulsive, consider the magnitude squared form shown in (31). It is known that (Part I, [4, p. 35])

$$\lim_{\lambda\to\infty} \int_{-\infty}^{\infty} f(y) \frac{1 - \cos \lambda(x-y)}{\pi\lambda(x-y)^2}\, dy = f(x). \tag{36}$$

In the limit, utilizing (33),

$$f(x) = \int_{-\infty}^{\infty} f(y)\{\epsilon(x-y)\cdot\epsilon^*(x-y)\}\, dy \tag{37}$$

where the asterisk denotes complex conjugation. Thus the magnitude squared of the vector (34) is an impulse in the Dirac delta sense, although the generating vector is composed of an impulse and a doublet.

We know from classical analysis that the response at a receiving point due to injection of a Dirac delta at a transmitting point is a general system describing function. If the system is such as to allow superposition of solutions, then we can state that the total energy density at the receiving point due to an arbitrary forcing function $x(t)$ at the transmitting point is obtained from

$$\sqrt{E(t)} = \int_{-\infty}^{\infty} x(\tau)\, h(t-\tau)\, d\tau \tag{38}$$

where the describing function $h(t)$ is the normalized system response to the Dirac delta of total energy (34). Likewise the potential energy component $V(t)$, also obtainable from a Dirac delta, has a similar form with its own describing function. It must therefore follow that there is a kinetic energy describing function obtained as the response to the unit doublet as assumed from the generating form of (30). By this argument $\epsilon(y)$ could be regarded as a unit energy impulsive vector composed of equal portions of potential energy producing impulse and kinetic energy producing doublet. The assumption that the impulse is related to potential energy is drawn by analogy of form from classical mechanics in the assumption that the difference in state following an offsetting impulse of position is positional displacement, while the difference of state following the doublet is velocity. Relating to circuit theory, suppose a single resonance circuit is excited by a unit impulse of voltage.

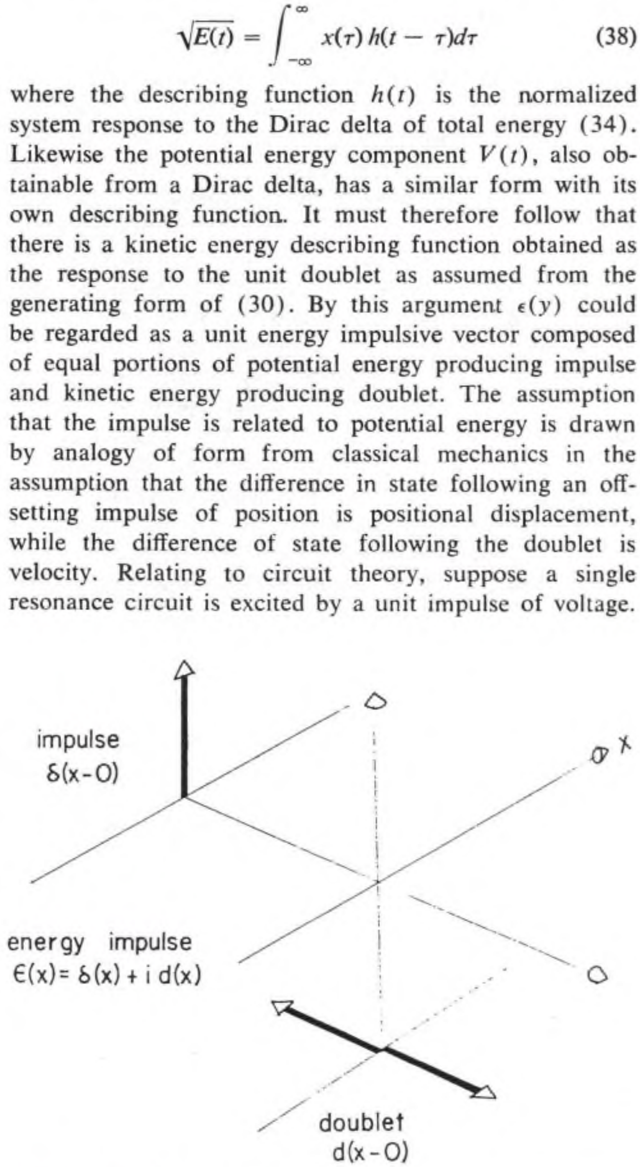

Fig. A-3. Sketch of limiting form assumed by components of Fig. A-2 when a is allowed to go to the limit. Note that the defining envelope of both impulse and doublet even as the limit is approached is proportional to the reciprocal of coordinate x.

At the instant following application of the impulse the capacitor has a stored charge (potential energy $\frac{1}{2}CV^2$) while the inductor has no current (kinetic energy $\frac{1}{2}LI^2$). Thereafter, the circuit exchanges energy under the relations (24) and (26). Should a unit doublet of voltage be applied, one would have as initial conditions a current in the inductor with no net charge in the capacitor. If one did not choose to identify the impulse with potential energy solely, he could multiply (34) by the unit vector of Eq. (14) to obtain

$$e^{i\lambda}\{\delta(t) + id(t)\} \qquad (39)$$

so as to redistribute the initially applied energy in the proper manner. Regardless of how one does this, it should be evident that a general description of system energy density must involve both the impulse and doublet response, not just the impulse response.

It might logically be asked why the need for a doublet response has not been previously felt with sufficient force to generate prior analysis. The answer is found in principle 2). An analysis based on either the impulse or doublet can be used to derive a complete system analysis by appropriate manipulation. The physical reason why one cannot use solely the impulse response or doublet response is that a measurement made on one system parameter, such as voltage, velocity, or pressure, can only express the momentary state of energy measured by that parameter. One scalar parameter of the type available from linear system operation does not represent the total system energy. A complete mathematics of analysis could be generated based completely on the doublet driving function and obtain the same results as a mathematics based on the impulse. This is because in order to get a complete answer, the complementing response must be calculated for either approach. Among the examples which spring to mind for the need of impulse and doublet analysis jointly is Kirchoff's formulation of Huygens' principle for acoustics (Part I, [25, p. 43]) and the impulse and doublet source solutions for electric and magnetic waves (Part I, [10]).

The response of a system $h(t)$ to the unit energy operation (34) is, from Eqs. (3), (5), and (15),

$$h(t) = f(t) + ig(t) = \int_{-\infty}^{\infty} f(x)\,\epsilon(t-x)\,dx. \qquad (40)$$

The system response $h(t)$ is the analytic signal composed of the impulse response $f(t)$ and the doublet response $g(t)$. From (25) the time position of energy sinks and sources is found from the local minima and maxima of

$$\frac{d}{dt}|h(t)|. \qquad (41)$$

While it is readily proved that the analytic signal $h(t)$ has a single-sided spectrum, this fact is of little value to our present consideration of energy. We assume that the parameter under analysis is a scalar or may be derived from a scalar potential. We assert that the sources of energy, which relate to the effective sources of sound, may be determined by considering both the kinetic and potential energies. These may be obtained separately as scalar components of the vector analytic signal. A local maximum in the magnitude of the analytic signal is due to a local source of energy and not an energy exchange.

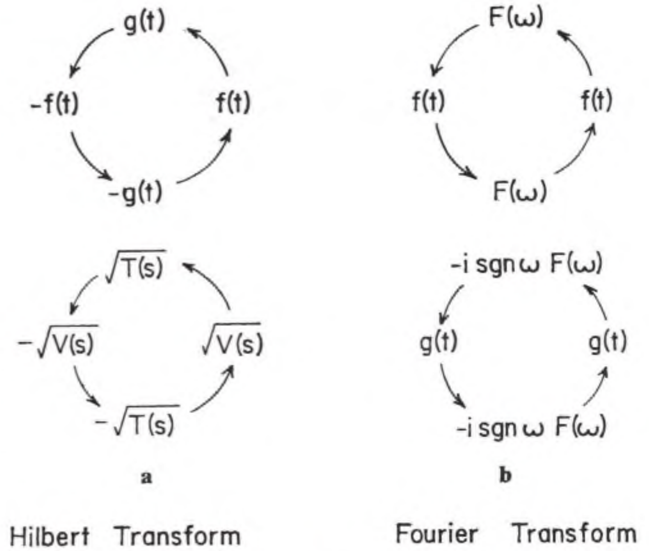

Fig. A-4. Symbolic representation of functional changes brought about by successive applications of **a.** Hilbert transformation to conjugate functions of same dimensional parameter; **b.** Fourier transformation to functions of reciprocal dimensional parameter.

The total energy of (24) is a scalar obtained by squaring the defined vector components of (27). The Hilbert transform relations exist between the vector components of (27) and subsequently of (40). Although two successive applications of Hilbert transformation produce the negative value of the original function (skew reciprocity), the energy being obtained as a square is uneffected. The Fourier transform relating two descriptions of the same event is reciprocal in order that no preference be displayed in converting from one domain to the other (Fig. A-4). The Hilbert transform, being skew reciprocal, does show a preference. This is also separately derived from the Cauchy–Riemann relations for the analytic function (27). It should be observed that the geometric relation between analytic functions derived in [2, Appendix A] of Part I must hold between the impulse and doublet response. Hence it is possible to generate a reasonably accurate sketch of the form of a doublet response from an accurate impulse response measurement. From these one could infer the form of total energy of a given system.

There is a strong generic tie between the imaginary unit $i = \sqrt{-1}$ and the generalized Hilbert transform in that two iterations of the operation produce a change of sign while four iterations completely restore the original function. One must surely be struck by the analogy of the energy related vectors (27) and (40) to the quadrature operation of the imaginary unit which is known to be related to the system-describing operations of differentiation and integration.

Note: Mr. Heyser's biography appeared in the October 1971 issue.