An anthology of the works of Richard C. Heyser on measurement, analysis, and perception

AUDIO ES ®

preface

Richard Charles Heyser, 1931–1987: A Retrospect

I was a college student and a neophyte in the audio world when I was introduced to Dick Heyser at an AES Los Angeles Section meeting in 1966. It was immediately apparent to me that he was someone special. Dick went out of his way to share his knowledge and enthusiasm with everyone, inspiring us to continually learn and understand more. Invariably he would stay late after the meetings to answer our questions about audio and, of course, to talk about time delay spectrometry (TDS). He would often begin "over our heads," but once he realized that he would shift gears to a more basic approach.

Dick was immensely gifted and, at the same time, genuinely modest about his talents and extensive achievements. In 1953, he obtained a B.S.E.E. degree from the University of Arizona. That year he was also awarded the Charles LeGeyt Fortescue Fellowship for advanced studies, which led to three years of postgraduate work at the California Institute of Technology. The fellowship was just the first of many honors Dick would receive during his 31-year career as a research engineer at Cal Tech's Jet Propulsion Laboratory (JPL).

His countless accolades and accomplishments have been cited in *Who's Who in Technology*, *Distinguished Americans of the West and Southwest*, and *Who's Who in the West*. He held nine patents in the audio and medical fields. A fellow of both the Audio Engineering Society and the Acoustical Society of America, Dick was also a member of IEEE, Tau Beta Phi, the Hollywood Sapphire Club, and a number of other organizations. He was a recipient of the AES Silver Medal for his development of TDS and selected as the 1986 *Sound and Video Contractor*'s Person of the Year. Shortly before the illness that culminated in his death, he became president-elect of the AES.

Dick was a prolific writer whose work appeared in such publications as *The Proceedings of the IREE (Australia)*, *Radiology*, and *The Proceedings of the Aerospace Medical Association*. He wrote many reviews and articles as a senior editor of *Audio* magazine, and a considerable number of his papers were published in the *Journal of the Audio Engineering Society*. It was in the October 1967 issue of the *Journal* that he introduced TDS—and the Heyser Transform, as it is becoming known—to the world.

TDS instrumentation has since proved its validity over time and in a variety of applications as a powerful tool for acoustical analysis and design. In medical research, TDS has been used to make ultrasound measurements for the identification of diseased soft tissue. Dick's ultrasound imaging project garnered several major grants and was staffed by five senior people at JPL. As a result, JPL's ultrasound lab was ranked one of the best in the country.

TDS has also been utilized in ultrasound measurements of the ocean floor. On a cruise of the Pacific, the ship Sea Sounder sailed along the Western U.S. Coast, from California to Alaska, to conduct an experiment for the National Science Foundation. A synthetic aperture device, dubbed a "fish," was lowered into the water for the purpose of collecting detailed images and information on ocean bottom sediment, troughs, and earthquake faults.

It is in the fields of acoustics and audio engineering that TDS has been applied most extensively. TDS has had a powerful effect on the design of diffusors; individual pieces of equipment, such as loudspeakers, microphones, and amplifiers; and the initial design and redesign of sound reinforcement systems, as well as audio control rooms, recording studios, and concert halls. In short, TDS has helped to unlock the definition of good acoustics.

This anthology consists of 32 writings on TDS audio. Of the works known to exist, only those clearly redundant—for example, a preprint that later became a *JAES* article—are not included. Three of the writings are, in actuality, authors' replies to comments on papers previously published in the *Journal*. For the sake of clarity, the comments are also published here.

An anthology can be organized in several ways. Since so many of Dick's articles are interrelated, I have chosen to present them chronologically. This is intended to facilitate the study of readers interested in the historical perspective—the evolvement of the technique and its ramifications—and to present no great hindrance to those wishing to explore specific aspects of TDS. The bibliography, which covers Dick's known works and related writings, has been divided into major subject categories and is chronological within those categories.

I deeply appreciate the efforts of those who assisted in the collection of material for the bibliography: Don and Carolyn Davis, who issued a request for listings of Dick's papers and forwarded the replies to me; Phillippe C. Troilliet, who generously responded with 10 pages of titles, dates, and pertinent facts; and others who took time to send similar information.

A special debt is owed to Dr. Dennis H. Le Croissette, Dick's closest and oldest colleague at JPL. Dr. Le Croissette urged him to begin writing a composite paper that would summarize his latest thoughts on TDS and would better serve as a reference paper. After Dick's death, Dr. Le Croissette was asked and graciously agreed to serve as editor of this paper prior to its inclusion in this anthology.

Along with the edited manuscript, Dr. Le Croissette sent a letter describing his perspective on Dick's TDS work. I have, with his permission, reproduced here much of what he expressed. Dr. Le Croissette wrote:

> The paper, "Fundamental Principles and Some Applications of Time Delay Spectrometry," by the late Richard C. Heyser, summarizes 25 years of his work in the fields of audio engineering measurement and analysis.
>
> In the early 1960s, Dick Heyser became interested in measuring the characteristics of loudspeakers in his personal laboratory. Since he did not have access to an anechoic chamber, he devised an electronic method of measuring the free-field response of the speaker in a reverberant setting. This ingenious method employed a coherent swept-frequency technique which he called time delay spectrometry.
>
> Over the years, Heyser recognized that his original concept had opened up a number of avenues leading to improved measurement and analysis of wave propagation. Although his original intent was the purely practical one of loudspeaker measurements, he soon began to sense and delve into the mathematical implications of his system. The relationship between time and frequency domains was his initial concern, but he also realized that TDS offered a window to view events in many different frames of reference. Following this same train of thought he investigated the differences that had been noted between so-called "objective" conventional signal analysis and the "subjective" appreciation of recorded sound.
>
> From 1955 until his untimely death in 1987 Dick Heyser was employed by JPL. In his work environment, he was involved for many years in the design and development of television systems used on the nation's unmanned space probes in the exploration of deep space. In recent years he applied the TDS technique, developed in his "spare time" to medical ultrasound imaging and to a series of underwater sound measurements.
>
> His composite paper explains his initial interest in the field, shows how it developed, and contains his latest thoughts on his multiple domain theory and his concept of alternatives. In addition, he summarized his work on four applications of TDS: loudspeaker evaluation, testing of microphones, underseas imaging, and medical ultrasound.
>
> TDS was first published in the *Journal* in 1967. I believe that his last words on this subject as given in this paper are also worthy of study by the acoustic community: Dick Heyser was always ahead of his time.

This anthology serves not only as a memorial to Dick's work with TDS but also as fundamental material for future developments in audio. Let us make the most of his legacy.

John R. Prohs
Guest Editor
1988 September

contents

PREFACE

PATENTS

biography

Richard C. Heyser was an active member of the AES for almost three decades. He served as a governor of the society from 1983 to 1984 and was prominent in numerous society activities. At various times he held all of the elected positions of the Los Angeles Section. In 1986 he was voted president-elect of the society.

Heyser was born in 1931 in Chicago, Illinois. He attended the University of Arizona, where he received a B.S.E.E. degree in 1953. Awarded a Charles LeGeyt Fortescue Fellowship for advanced studies, he earned an M.S.E.E. degree from the California Institute of Technology in 1954. He spent the next two years doing postgraduate work at the California Institute of Technology. In 1956 he joined the Jet Propulsion Laboratory of the California Institute of Technology in Pasadena, California, where he became a member of the technical staff. His work involved communication and instrumentation design for all major space programs at JPL, beginning with the conceptual design of America's first satellite, Explorer I. Later, he was involved in the application of coherent spread spectrum techniques to improving underwater sound research and medical ultrasound imaging.

In addition to his work at JPL, Mr. Heyser maintained a personal laboratory where he conducted research on audio and acoustic measurement techniques. This effort resulted in a number of papers published in the *Journal of the Audio Engineering Society* and elsewhere. He was awarded nine patents in the field of audio and communication techniques, including time delay spectrometry. Heyser was a reviewer for the *Journal* and a member of the Publications Policy Committee of the AES. He also was active in the society's standards work, serving as chairman of the Audio Polarity Committee. As a senior editor of *Audio* magazine he was responsible for the loudspeaker reviews of that publication for the past 12 years. He was a member of the IEEE, a fellow of the Audio Engineering Society, and the recipient of its Silver Medal Award in 1983. He was also a fellow of the Acoustical Society of America and a member of the Hollywood Sapphire Club. He is listed in *Who's Who in the West, Who's Who in Technology*, and *Distinguished Americans of the West and Southwest*.

Richard C. Heyser **1931–1987**

Reprinted from *JAES*, vol. 15, no. 4, pp. 370–382, 1967 October.

Acoustical Measurements by Time Delay Spectrometry*

RICHARD C. HEYSER

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

A new acoustical measurement technique has been developed that provides a solution for the conflicting requirements of anechoic spectral measurements in the presence of a reverberant environment. This technique, called time delay spectrometry, recognizes that a system-forcing function linearly relating frequency with time provides spatial discrimination of signals of variable path length when perceived by a frequency-tracking spectrum analyzer.

INTRODUCTION It is a credit to technical perseverance that the electronic subsystems which make up an audio installation have been brought to a high state of perfection. It is possible not only to predict the theoretical performance of a perfect electronic subsystem, but to measure the deviation of the performance of an existing subsystem from that perfect goal. The capability of measuring performance against an ideal model and of predicting the outcome for arbitrary signals is taken for granted. Yet the very acoustical signals which are both the source and product of our labors seldom have sufficient analysis to predict performance with comparable analytical validity. The measurement of even the simpler parameters in an actual acoustical system may be laborious at best. There exists, in fact, very little instrumentation for fundamental measurements which will allow prediction of performance under random stimuli. It is the intent of this paper to describe a new acoustical measurement technique which allows "on-location" measurement of many acoustical properties that normally require the use of anechoic facilities. A new acoustical model of a room is also introduced as a natural by-product of this technique. With this model substantial objects may be selectively analyzed for their effect on sound in the room.

The acoustic testing process to be described relies heavily on electronic circuit techniques which may not be familiar to acousticians. As a brief review of fundamental principles it will be recalled that in linear electronic circuit analysis the concept of superposition permits complete analytical description of circuit response under the influence of any driving function which is describable as a distribution of sinusoids. Each sinusoid in the distribution will possess a unique amplitude and time rate of change of angle or frequency, and will produce a network response which is in no way dependent on the existence of any other sinusoid. The response of a network at any particular frequency is therefore obtained quite simply by feeding in a sinusoid of the desired frequency and comparing the phase and amplitude of the output of a network with its input. The response of a network to all frequencies in a distribution will then be the linear superposition of the network response to each frequency. The response of a network to all possible sinusoids of constant amplitude and phase is called the frequency response of that network. The frequency response is in effect the spectrum of frequency distribution of network response to a normalized input. Cascading of linear networks will involve complex multiplication of the frequency response of each included network to obtain an overall response. An analytical solution to such a combination may then be readily obtained from this overall frequency response.

The equivalent frequency response of an acoustical system should, in principle, be the comparable value in analyzing resultant performance. The processes of sound generation, reflection, transmission, and absorption all have their counterparts in network theory. It is well known, however, that any attempt at utilizing a simple sinusoid driving function on a real-world acoustical system will lead to more confusion than insight. Any object with dimensions comparable to a spatial wavelength of the sinusoid signal will react to the signal and become an undesired partner in the experiment. Furthermore, since the velocity of sound in the various media prevents instantaneous communication, time enters into the measurement in the form of standing wave patterns which will be different for each applied frequency. For those acoustical subsystems for which a single frequency response might be meaningful, such as loudspeakers, microphones, or certain acoustical surfaces, special (and expensive) anechoic test areas are utilized in an attempt

* Presented October 16, 1967 at the 33rd Convention of the Audio Engineering Society, New York.

to remove acoustically any object that might interfere with the measurement. Unfortunately there are many acoustical situations for which such a measurement appears to be impossible or even meaningless. A single frequency response of an auditorium, for example, will be of no use in evaluating the "sound" of the auditorium as perceived by an observer. The large number of reflecting surfaces give a time-of-arrival pattern to any attempted steady-state measurement which prevents analytical prediction of response to time-varying sound sources. Clearly an auditorium has not a single frequency response, or spectral signature, but rather a linear superposition of a large number of spectral responses, each possessing a different time of arrival.

Here we have the essence of many real-world acoustical measurement problems. It is not a single frequency response that must be measured but a multiplicity of responses each of which possesses a different time delay. The room in which an acoustical measurement is made simply adds its own series of spectral responses, which may mask the desired measurement. Selection of the proper responses will yield a set completely defining the acoustical system under test so long as superposition is valid. Conversely, once the entire set of spectra are known along with the time delay for each spectrum it should be analytically possible to characterize the acoustical system for any applied stimulus. Traditional steady-state techniques of measurement are unable to separate the spectra present in a normal environment since a signal response due to a reflection off a surface differs in character from a more direct response only in the time of arrival following a deliberately injected transient. In the discussion to follow a method of time-delay spectrometry will be developed which allows separation and measurement of any particular spectral response of an acoustic system possessing a multiplicity of time-dependent spectral responses. A practical implementation of this technique will be outlined using presently available instruments. Analytical verification of the technique will be developed and a discussion included on acoustical measurements now made possible by this technique.

EVOLUTION OF MEASUREMENT TECHNIQUE

In evolving the concept of time-delay spectrometry it will be instructive first to consider a very simple measurement and then to progress by intuitive reasoning to the general case. Assume that it is desired to obtain the free-field response of a loudspeaker situated in a known reverberant environment. A calibrated microphone will be placed at a convenient distance from the speaker in the direction of the desired response. With the acoustical environment initially quiescent, let the speaker suddenly be energized by a sinusoid signal. As the speaker activates the air a pressure wavefront will propagate outward at a constant velocity. This pressure wavefront will of course not only travel toward the microphone but also in all other directions with more or less energy. Assume that the microphone is connected through a relatively narrow-bandwidth filter to an indicating device, and that this filter furthermore is tuned to the exact frequency sent to the loudspeaker. As the leading edge of the pressure wave passes the microphone, the only contributor to this wave could be the loudspeaker since all other paths from reflective surfaces to the microphone are longer than the direct path. The loudspeaker will take some time to build up to its "steady-state" excitation value and the microphone and tuned circuit will similarly have a time constant. If the system arrives at a steady-state value before the first reflected sound arrives at the microphone, this steady-state is in effect a free-field measurement at the frequency of the impressed sinewave. Note that because of the broad spectrum of a suddenly applied sinewave, a tuned filter circuit is necessary to prevent shock-excited speaker or microphone resonances from confusing the desired signal.

The measurement thus described is quite simple and has actually been used by some investigators.[1-3] As long as a fixed filter is used, the measurement must be terminated prior to receipt of the first reflected "false" signal and the system must be de-energized prior to a subsequent measurement. Suppose, however, that the fixed-frequency sinewave is applied to the speaker only long enough to give a steady-state reading prior to the first false signal, then suddenly shifted to a new frequency outside the filter bandwidth. Assume also that by appropriate switching logic, a filter tuned to this new frequency is inserted after the microphone at the precise time that the sound wave *perceived by the microphone* changes frequency. The microphone circuit will thus be tuned to this new frequency and the later reflected false signals of the first frequency will not be able to pass through the new filter. If one continues this process through the desired spectrum it is apparent that the indicator circuit will never "know" that the measurement was performed in a reverberant environment and a legitimate frequency response may thus be measured.

The practical economics of inserting fixed filters and waiting for the starting transient at each frequency to die down weigh heavily against such a system, so an alternative may be considered. Project a smooth glide tone to the speaker and utilize a continuous tracking filter after the microphone. If the tracking filter is tuned to the frequency of the emitted glide tone as perceived by the microphone and if the glide tone has moved in frequency by at least the bandwidth of the tracking filter before the first reflected signal is perceived, no buildup transient is encountered and the measurement will be anechoic even though performed in a reverberant environment. The nature of the glide tone may readily be ascertained by intuitive reasoning. If there is no relative motion between speaker and microphone it can be stated with absolute certainty that the time delay between speaker and microphone is a constant. There is, in other words, a unique and linear relationship between time and distance traveled by the pressure wave. Each reflecting surface will appear to be a new sound source with a time delay corresponding to path length. If we specify that the sweeping tone and the tracking filter combination be capable of maximizing response for all frequencies from any given apparent source then we have required an equivocation of frequency, room spacing, and time. The glide tone satisfying this requirement possesses a constant slope of frequency versus time. If all reverberant energy due to any given frequency has died to an acceptable level a fixed time following excitation, say T seconds, then the glide tone may be allowed to repeat its linear sweep in a sawtooth fashion with a period of no less than T seconds.

While we began by postulating direct loudspeaker measurement it is apparent that we could by suitable choice of sweep rate, bandwidth, and time delay, "tune" in on at least first-generation reflections with the selective exclusion of others, even the direct loudspeaker response. Because the output of the tracking filter yields the spectral signature of the perceived signal with a frequency proportional to time and since selective spatial isolation of the desired signal is obtained by utilizing the fixed time delay between source and microphone, the rationale of the name *time-delay spectrometry* becomes apparent.

Some simple relationships may be directly derived from the basic geometry of a practical situation. Consider the representation of Fig. 1 in which a microphone is connected to a tracking filter "tuned" to perceive a source at a distance X on a direct path. The filter has a bandwidth B Hz; it is seen that the sweep tone will traverse some ΔX in space while within a band B of any given frequency. Define ΔX as the region in space, along the direction of propagation of the acoustic signal, within which the selected signal power will be no less than half the maximum selected value. This is the spatial analog of the half-power bandwidth B of the tracking filter, and will therefore be referred to as the space-equivalent bandwidth. This space-equivalent bandwidth ΔX is related to the tracking filter electrical bandwidth B, the velocity of sound c and the rate of change of frequency $\Delta F/\Delta t$, by

$$\Delta X = B[c/(\Delta F/\Delta t)] \approx c/B. \qquad (1)$$

The last relation is based upon an optimized bandwidth which is the square root of the sweep rate. While not immediately obvious, this optimized bandwidth is common in sweeping analyzers of the variety recommended for this measurement.[4]

The signal perceived by the microphone is that emitted by the speaker some time in the past. To visualize the relationship consider Fig. 2 diagramming the behavior of a sweep tone repetitive in a time T. The signal emitted from the source or transmitter will be denoted by F_t while that received by the microphone is F_r. It is usually desired to sweep through zero frequency; this is shown in the diagram as a signal dropping in frequency uniformly with time until zero frequency and then rising again. We are thus going through zero beat. In this diagram the direct signal possessing the shortest time delay is shown dashed. It is clear that the relationship between distance, transmitted and received frequencies at any instant, and sweep rate is

$$X = (F_r - F_t)[c/(\Delta F/\Delta t)]. \qquad (2)$$

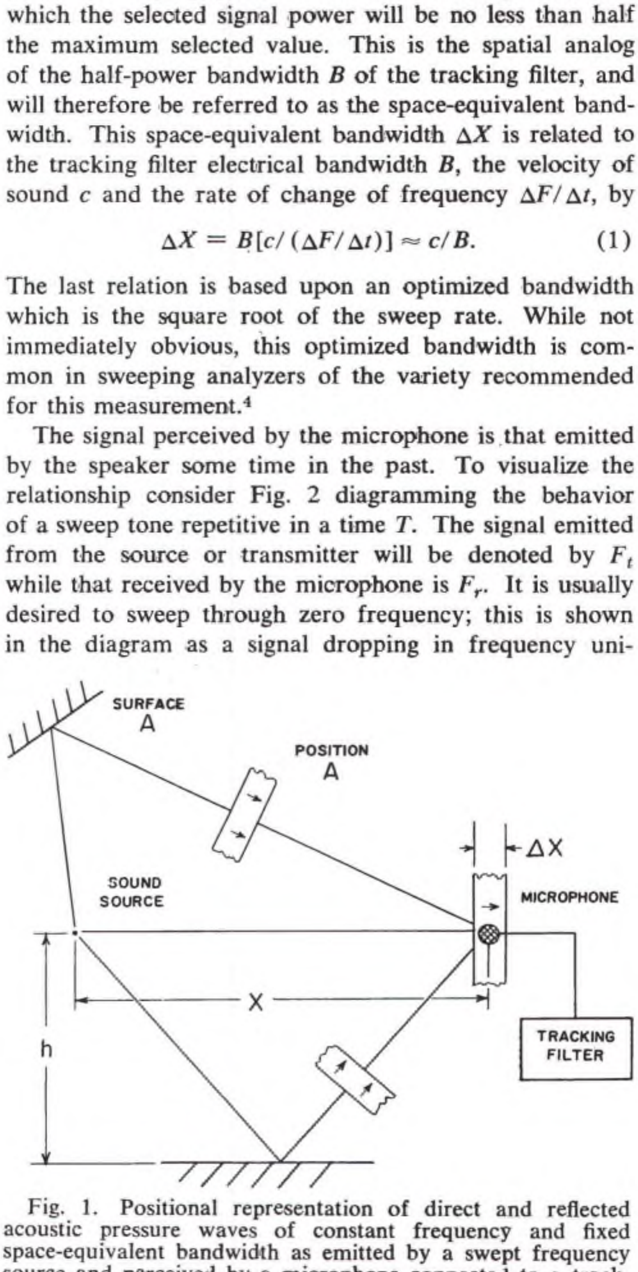


Fig. 1. Positional representation of direct and reflected acoustic pressure waves of constant frequency and fixed space-equivalent bandwidth as emitted by a swept frequency source and perceived by a microphone connected to a tracking filter tuned to maximize the response at a distance X.

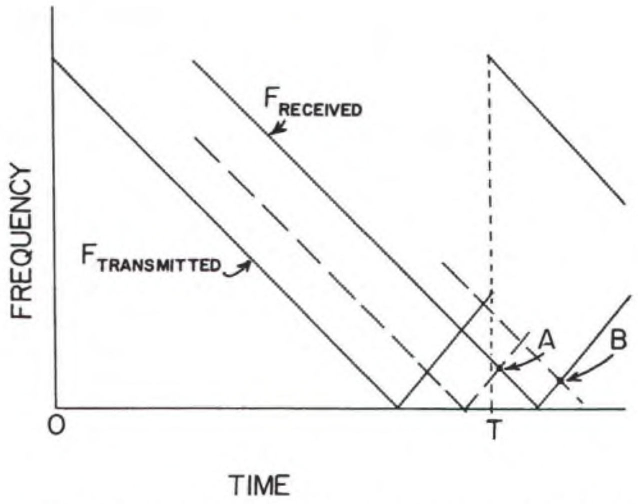


Fig. 2. Frequency plot of the sweep tone passing through zero frequency, showing room signals received as a function of time.

This states the fact that to "tune" to the response of a signal X feet away from the microphone it is only necessary to offset the frequencies of glide tone and tracking filter by a fixed difference. Referring again to Fig. 1 we could check the response of the reflective surface A in one of two ways. First, we might offset the source and received frequencies to yield the primary sound signal as shown and then, without changing this frequency offset, physically transport the microphone to position A. It is assumed that a position may be found for which no contour of undesired reflected or direct sound comes into space tune at position A. The second way of adjusting for surface A would be to maintain the existing microphone position and change the offset between transmitted and received frequencies to account for the longer path of A. Both techniques allow for probing the effective acoustic surfaces in the room. The room is, of course, filled with sound, but since we have an instrument uniquely relating time, space, and frequency we have in effect "frozen" the space contours of reflected sound. We may probe in space merely by adjusting a frequency offset. Analytically we may state that we have effected a coordinate conversion which trades spatial offset for driving frequency offset, but which retains intact all standard acoustical properties including those due to the frequency of the system-driving function. This is the power of this technique, for so long as the acoustic properties remain substantially linear, a hopeless signal combination in normal spatial coordinates transforms to a frequency coordinate generally more tractable to analyze.

In considering the uniqueness of signals perceived by frequency offsetting it may be observed that a transmitted signal passing through zero frequency will reverse phase at the zero frequency point and continue as a real frequency sweep with reversed slope of frequency *vs* time.

Thus, near zero there will be two transmissions of a given frequency during one sweep. The signals corresponding to the same frequency slope as that of the tracking filter will be accepted or rejected in total as the proper offset is entered for the appropriate time delay. The duplicate frequencies near zero possessing a different slope will cause a "ghost" impulse to appear when the instantaneous frequency perceived by the microphone corresponds to that of the tracking filter, regardless of the offset. Similarly, a tracking filter may be set to "look" for frequencies on both sides of zero. If the signal under analysis is a first-generation reflection, the main signal from the speaker will arrive sooner than this reflection and an impulse will occur at position *A* in Fig. 2. If, on the other hand, we are looking for a signal prior to the main signal, the impulse might appear at position *B*. Thus a repetitive sweep passing through zero frequency may contain image impulses due to signal paths other than that to which the tracking filter has been tuned. All sweeps with sufficiently long periods which do not pass through zero frequency will avoid any such ambiguities.

Another relationship worthy of consideration is the distance one may separate speaker and microphone without interference from a large object offset from the direct line of sight. Such considerations arise when measurements are made near floors, walls and the like. Surprisingly, an off-center reflecting object such as a floor limits the maximum distance of speaker-microphone separation. Consider Fig. 1 where the speaker and microphone are assumed to be a height h off the floor. The maximum distance will be that for which the path length from the image speaker is just equal to $X+(\Delta X/2)$. The maximum usable line-of-sight distance between speaker and microphone, X_{max}, is related to height h and space equivalent bandwidth ΔX by

$$X_{max} = (\Delta X/4)\{[h/(\Delta X/4)]^2 - 1\}. \qquad (3)$$

Note that if it is necessary to make lower frequency measurements or, what is the same thing, higher definition measurements, the distances scale up to values comparable to the wavelength of that frequency for which the period is the rise time of the tracking filter. This follows from observing that Eq. (1) may be rewritten as

$$\Delta X \cdot B \approx c \qquad (4)$$

PRACTICAL IMPLEMENTATION

While it might appear at first glance that the acoustical measurements outlined in the previous paragraph require specialized apparatus, equipment necessary to perform the described measurement is readily assembled from commercially available instruments. The tracking filter is a portion of an audio spectrum analyzer. This instrument is a narrow-band superheterodyne receiver tuned through the audio spectrum with a local oscillator swept linearly in time—the needed frequency characteristic. Depending upon the resolution required, the commercially available analyzers will sweep through a variety of given frequency dispersions in a repetitive sawtooth fashion. A sweep rate of one per second is typical for spectrum analyzers covering the full audio spectrum. The output of this audio tracking filter is rectified and applied to the vertical axis of a self-contained oscilloscope with the linear time axis swept horizontally. Because of the repetitive nature of the display, this instrument provides a visual presentation of signal energy *vs* frequency. The bandwidth of the spectrum analyzer is usually chosen to be the narrowest possible without losing information when swept past a complex spectrum.

The sweep tone for driving the loudspeaker may be obtained by down-converting the spectrum analyzer local oscillator to the audio band. If the local oscillator is heterodyned with another oscillator equal to the analyzer intermediate frequency, the difference frequency will always be at the precise frequency to which the analyzer is tuned. The proper offsetting frequency for spatial tuning of acoustical signals is obtained by detuning the fixed oscillator from the intermediate frequency. A down-converting synchronously sweeping generator is usually available as an accessory to the spectrum analyzer. To perform time-delay spectrometry it is only necessary to substitute a stable tunable oscillator for the fixed crystal oscillator. Where extreme accuracy is required it may be necessary to use a frequency counter to monitor the proper offset frequency. The remainder of the equipment consists of the usual power amplifier, microphone, and preamplifier.

Figure 3 diagrams a complete setup capable of performing any or all of the measurements outlined. While the cost of such assembled equipment is greater than

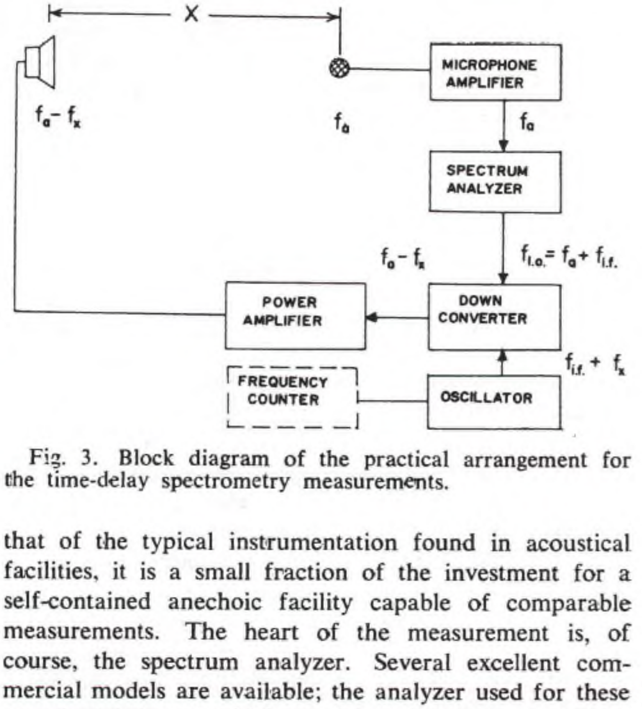


Fig. 3. Block diagram of the practical arrangement for the time-delay spectrometry measurements.

that of the typical instrumentation found in acoustical facilities, it is a small fraction of the investment for a self-contained anechoic facility capable of comparable measurements. The heart of the measurement is, of course, the spectrum analyzer. Several excellent commercial models are available; the analyzer used for these experiments is one which provides a continuously adjustable sweep width from 200 Hz to 20 KHz with a sweep center frequency separately adjustable from dc to 100 KHz. The bandwidth is tracked with the sweep width control to yield optimum resolution at the sweep rate of one per second. According to Eq. (1), this means that at maximum dispersion a 20 KHz spectrum may be obtained at 141 Hz resolution at 7.8 ft space-equivalent bandwidth. If measurements are desired valid to 32 Hz, for example, the sweep width would be set to 1000 Hz and the space-equivalent bandwith would be 34 ft. Where smaller space-equivalent bandwidths are desired and the decreased resolution can be tolerated, it is possible to increase the internal analyzer sweep rate by as much as a factor of five by minor circuit modification without compromising the validity of the measurement.

APPLICATIONS OF TIME DELAY SPECTROMETRY

Perhaps the best way of illustrating the use of time-delay spectrometry would be a detailed look at a typical measurement. Assume that an on-axis pressure response frequency characteristic is desired for a direct radiator loudspeaker. The entire spectrum from zero to 20 KHz is desired, with a frequency resolution of 140 Hz acceptable. The room selected for measurement has an 8 ft floor-to-ceiling height and a reverberation time of less than one second.

The space-equivalent bandwidth is approximately 7.8 ft (from Eq. 1), which means that no substantial object should be placed within 4 ft of the pickup microphone. An acceptable sweep rate is one 20 KHz sweep per second. From Eq. (3) it can be seen that if the speaker is placed halfway between floor and ceiling, the maximum distance the microphone should be placed from the speaker is six feet. This assumes, of course, a worst-case specular reflection from both floor and ceiling as well as a uniform polar response from the speaker.

The speaker under test is then placed 4 ft off the floor and more than 4 ft from any substantial object. The pickup microphone should be placed on-axis within 6 ft of the speaker and at least 4 ft from any reflective surface. The microphone is electrically connected to a suitable preamplifier which in turn feeds into the spectrum analyzer input. The down-converted local oscillator signal from the tracking oscillator is fed to an appropriate power amplifier driving the speaker. The offset oscillator which replaces the crystal in the tracking oscillator is connected to a frequency counter to complete the electrical setup.

With the deviation and deviation rate set to the desired test limits in the spectrum analyzer, a sweeping tone will be heard from the speaker. If the microphone-to-speaker distance is known, the offset oscillator should be set in accordance with Eq. (2) to achieve maximum deflection of the spectrum analyzer display. For example, a 6 ft separation will require a 114 Hz offset. It is at this point that the advantage of a tracking analyzer is evident: not only may the intensity of sound from the speaker be modest, but also it is not necessary to cease all sound and motion in the room while measurement is in progress. Any reasonable extraneous sound may be tolerated, and mobility need only be restricted to the extent that travel between loudspeaker and microphone is discouraged.

The spectrum analyzer display should remain stationary in vertical deflection. To ascertain that the proper frequency offset is used, the offset oscillator may be detuned on both the high and low side; the resultant display should show a reduction in amplitude. The peaked stationary pattern on the screen of the spectrum analyzer will be a plot of pressure response *vs* frequency with a smoothing bandwidth of 0.7% of the displayed dispersion.

Without altering the test setup several types of acoustical measurements may be made. Off-axis response of the speaker may be obtained by rotating the speaker by the required angle and immediately viewing the results on the analyzer. If it is desired to determine the sound transmission characteristic of a particular material, for example, one first obtains the on-axis speaker response and then interposes a 4 ft sample between speaker and microphone, noting the new response. If the spectrum analyzer is set for logarithmic deflection, the sound absorption in decibels is the difference in the two responses independent of speaker or microphone response.

The reflection coefficient may similarly be obtained by positioning the material behind the microphone by at least 4 ft and retuning the offset oscillator for the reflected signal. If the microphone is rotated 180° to remove its polar characteristics from the measurement, and if the proper space loss is entered, the reflection coefficient is obtained directly as a function of frequency, independent of microphone and loudspeaker characteristic. If one does not wish to compute the space loss, and space permits, the microphone may be physically transported to the position of maximum on-axis response with the reflective surface removed, and space loss obtained as response difference from the earlier position.

ROOM SPECTRUM SIGNATURES

The analysis up to this point has been directed at producing an electrical signal which when fed to a speaker will allow two very important measurements to be made on the sound picked up by a microphone. First, only the selected direct or reflected signal will be pulled out and displayed. Second, the spectral response of the selected signal will be directly presented. With this in mind, consideration will now be given to a characterization of sound in a room by time-delayed spectra.

Assume that an observer is positioned in a room with a localized source of sound such as a loudspeaker. This loudspeaker furthermore has a pressure frequency response characteristic, such as Curve A in Fig. 4. As the loudspeaker is energized by any arbitrary signal, the first sound heard by an observer will be characterized by Curve A. (This does not imply that all the frequencies

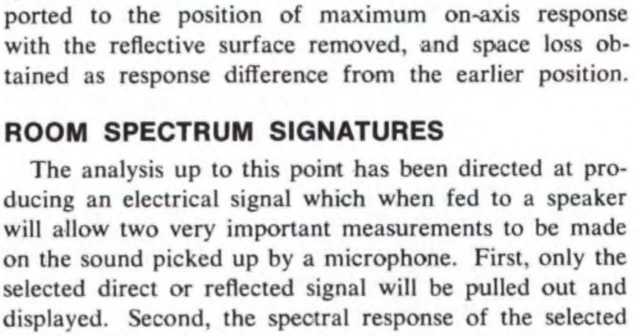


Fig. 4. Acoustic model of room spectrum signatures as perceived by an observer listening to a sound source with spectral distribution *A*.

are present, but only that those electrical signal components which are present are modified in accordance with Curve A.) Since the loudspeaker output varies as a function of time it will be assumed that Curve A is representative of the pressure output at a particular instant in time. We have in effect plucked out that signal which is characteristic of a given instant and will follow this pressure wave as it travels through the room and intercepts the observer.

A few milliseconds after the principal pressure wave A has passed, a second pressure wave B will be perceived by the observer. This is a reflected wave and is made up of the principal wave modified by the frequency response of the reflecting surface. As in electronic circuit theory, the observed resultant frequency response of B is the product of the frequency response of A and the frequency response of the reflecting surface. Beside the obvious time delay introduced by the longer path length, the surface creating the wave B may introduce a dispersive smearing of portions of the frequency range. This may be due to surface irregularities which act as local scattering centers or to acoustic impedance variations causing effective displacement of the position of optical and acoustic surfaces. If the reflecting surface has very little depth variation and is primarily specularly reflective, such as a hard wall or ceiling, the time-pressure profile might appear as wave B. If on the other hand the reflecting object has depth, such as a chair, the time-pressure profile may take on the character of the third pressure wave C. Here the effect is a distinct broadening of the time during which the observer perceives the signal. Perhaps a better name than reflection coefficient in the case of such an object might be scattering coefficient.

As time progresses the observer will perceive more scattering spectra with increasing density and generally lower amplitude until all sound due to the initial loudspeaker excitation has fallen below a predetermined threshold. The complete pressure-frequency-time profile will be a unique sound signature of the room as perceived by an observer listening to the loudspeaker. This is, after all, the way in which the sound reaches the observer. A more general characterization of the room itself would replace the loudspeaker and its unequal polar pressure response by an analytically uniform pressure transducer. Each position of observer and transducer will have its unique pressure-frequency-time profile. It quite frequently happens, however, that such a general characteristic is of little concern and what is desired is the observer-loudspeaker situation of Fig. 4.

The complete pressure-frequency-time profile represented by Fig. 4 is a useful acoustic model of a reverberant room in which an observer perceives a localized source of sound. As far as the observer is concerned there are many apparent sound sources. Each of these has the same "program content" as the primary source of sound but possesses its own spectral energy distribution and unique time delay, corresponding to its apparent location with respect to the observer. The effect of any given object in the acoustic environment may be determined in this model by noting the equivalent time delay from source to object to observer, and analyzing the scattering spectrum corresponding to this delay. The entire set of time-dependent scattering spectra should allow detailed analysis of this system for any time-dependent stimulus applied to localized source of sound.

This acoustic model of a room is seen to tie directly to time-delay spectrometry. The electronic sweep tone fed to the loudspeaker will in effect represent all possible frequencies in the chosen sweep range. The time axis of Fig. 4 could also be labeled offset oscillator frequency; the signal displayed on the spectrum analyzer will be seen to be the frequency spectrum of the selected time delay. The space-equivalent bandwidth, ΔX, will establish the ability to resolve independent scattering spectra since time-delay spectrometry forms an equivocation of time, offset frequency, and distance. It is immediately apparent that the selection of loudspeaker on-axis response used as an example for generation of time-delay spectrometry entails selection of Curve A and is only a special case of a more general acoustical measurement technique. It is possible by suitable choice of loudspeaker and microphone position to "pull out" important acoustic properties of a room without destroying the room or utilizing large-scale digital computer techniques.

Analytically, a surface at a distance corresponding to a time delay of t_k seconds may be characterized by a frequency spectrum multiplied by a linear phase coefficient. The cumulative distribution of these sources is represented by Fig. 4 and may be expressed as

$$R(\omega) = \sum_k S_k(\omega)\, e^{-i\omega t_k}, \tag{5}$$

where $R(\omega)$ is the cumulative distribution of all sources, both real and apparent, as perceived by an observer listening to a localized source of sound in a reverberant environment, and $S_k(\omega)$ is the spectral energy distribution of the source at a distance corresponding to a time delay of t_k seconds. The angular frequency ω, expressed in radians per second, will be used in this paper for analytical simplicity. Since the absolute magnitude of $S_k(\omega)$ is measured by time-delay spectrometry it is not necessary to introduce a space loss term. If the source has a spectral distribution $P(\omega)$, then the observer will perceive a signal which has a frequency spectrum that is the product of $P(\omega)$ and $R(\omega)$. The impulse response of the room, $r(t)$, is simply the Fourier Transform of $R(\omega)$:

$$r(t) = (1/2\pi) \int_{-\infty}^{\infty} R(\omega)\, e^{+i\omega t}\, d\omega \tag{6}$$

and the time-varying signal $o(t)$ perceived by an observer for a program source $p(t)$ is the convolution integral of program source and room response

$$o(t) = \int_{-\infty}^{\infty} p(\tau) r(t-\tau) d\tau \tag{7}$$

The concept expressed by Eqs. (6) and (7) is certainly not new to the field of acoustics. The difficulty in applying these relations lies in the massive amounts of data that must be processed before the effect of a particular object in a room may be evaluated. The acoustic model of Eq. (5) as represented in Fig. 4 leads to a simpler analysis since time-delay spectrometry isolates the $S_k(\omega)$ function of acoustic surfaces and presents the information in a meaningful form which requires no further reduction. While not a cure-all for acoustic analysis, time-delay spectrometry provides a good insight to actual acoustic situations.

EXPERIMENTAL RESULTS

Sufficient experimental evidence has been collected to assure validity of the technique. The examples below give some insight into the nature of the spectrum display.

Loudspeaker Testing

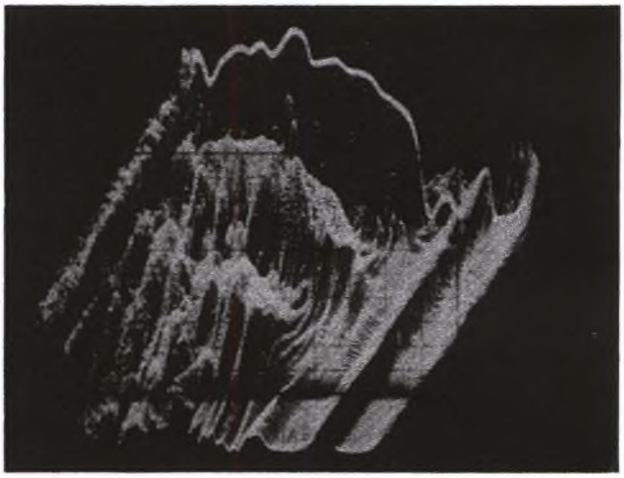
A single-cone 8-in. loudspeaker housed in a ported box was chosen for test, in a room deliberately selected to be a poor environment for loudspeaker testing. The room measures 24 × 10 × 7 ft. The floor is hard vinyl on cement and the ceiling is plaster, while all other surfaces are hard wood. Miscellaneous objects throughout the room reduce the open area so that when the loudspeaker-to-microphone distance is at the calculated maximum of Eq. (3) the smallest separation from the microphone to any substantial object is one half the space-equivalent bandwidth. (The room spacing is chosen to be no larger than the calculated minimum dimension in order that the concept may be better demonstrated.) The spectrum measurement was made to include dc to 10 KHz at a dispersion rate of 20 KHz per second. Fig. 5a is a photograph of the on-axis pressure response in decibels as a function of linear frequency. Since the analyzer sweeps through zero-beat to bring dc at the left-hand edge, a mirror image response is seen for those frequencies to the left of dc.

In order to illustrate the nature of the response to be obtained for the steady-state application of sinewaves, the test setup was left intact and the spectrum analyzer driven at such a slow rate that the room acoustics entered into the measurement. Figure 5b is the result of this measurement. The sweep width, vertical sensitivity, and analyzer bandwidth are identical to those of Fig. 5a, but the sweep rate is such that it takes over three minutes to go from zero frequency to 10 KHz. The camera aperture was optimized to give a readable display, and consequently the large number of standing wave nulls and peaks did not register well. The wild fluctuation in reading as a function of frequency is quite expected and lends credence to the futility of steady-state loudspeaker measurement in the acoustic equivalent of a shower stall.

One certain way of verifying that a "free-field" measurement is made is to impose inverse square law.[5] Figure 5c is a time-delay spectrograph of an inverse square law measurement. The setup of Fig. 5a and 5b was spectrographed and appears as the lower trace in Fig. 5c. The loudspeaker was then moved toward the microphone to one half its former distance, and a second exposure was taken with the proper offset oscillator setting. Since the display is logarithmic, the two traces should be parallel and separated by 6 dB, and indeed this is seen to be the case. The minor discrepancies are due to the fact that the first measurement was made at 5 ft, which is the maximum theoretical separation for the room and dispersion rate. The closer measurement is made at 2½ ft where the loudspeaker does not exactly appear as a point source since the separation-to-diameter ratio is only four.

a.

b.



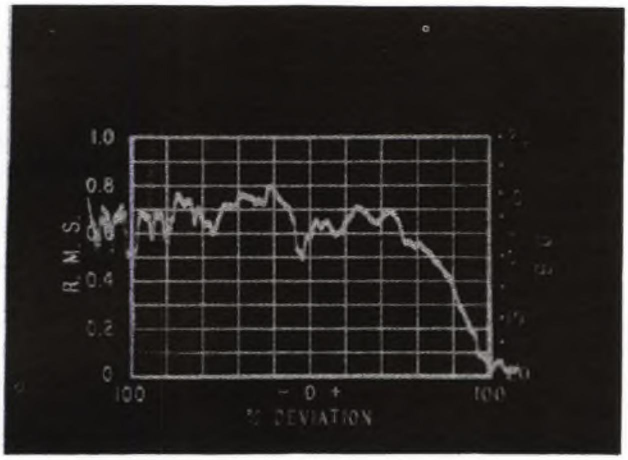
c.

Fig. 5. Oscillographs of logarithmic pressure response measurements on a single-cone loudspeaker from dc to 10 KHz, measured in a reverberant environment by time-delay spectrometry. **a.** Using rapid sweep. **b.** Using slow sinewave sweep. **c.** Showing inverse square response as a function of distance.

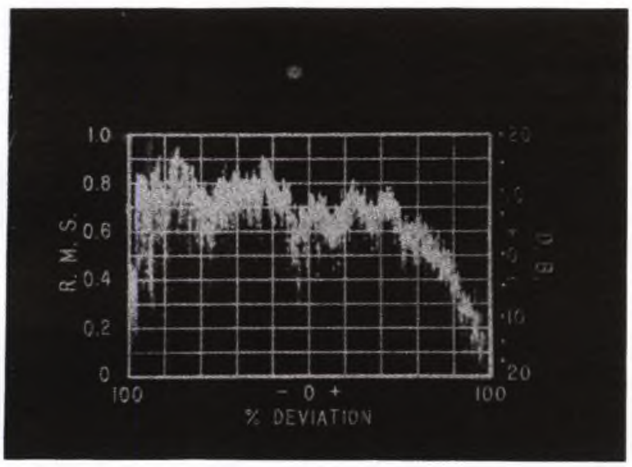
Isometric Display of Room Signature

The acoustic model of a room proposed above resulted in a three-dimensional plot of pressure, frequency, and time. Time-delay spectrometry is a means of probing and presenting this model. In order to visualize this phenomenon, the spectrum analyzer display has been modified to accept what may be called Isometric Scan. The offset oscillator is in this instance a linear voltage-controlled oscillator. Each sweep of a time delay spectrograph is presented normally as a plot of pressure *vs* frequency. Isometric scan advances the offset oscillator following each sweep and creates the effect of the third dimension of offset frequency exactly as one illustrates a three-axis system on two-dimensional paper, that is, by a combined horizontal and vertical offset displacement of the sweep. The net effect is a presentation identical in form to Fig. 4.

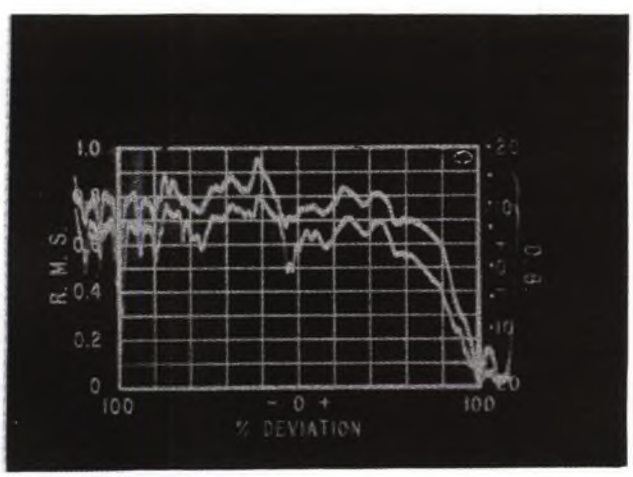
Figure 6a is an isometric scan of the room and setup of the preceding loudspeaker test. The time axis commences at zero when the principal on-axis signal is received and advances through 43 msec which is equivalent to twice the longest room dimension. All frequencies from zero through 10 KHz are displayed. The exceedingly large number of "alpine peaks" in evidence are only partially visible in the photograph since it is difficult to capture the visual impression an observer sees as this three-dimensional plot unfolds on the oscilloscope screen.



a.



b.



c.

Fig. 6. Oscillographs of isometric displays of spectral signatures using time-delay spectrometry. **a.** Room signatures of the setup of Fig. 5. **b.** Principal response of a multicellular horn from dc to 20 KHz. **c.** Spectral signatures obtained with loudspeaker system pointed at ceiling, including direct response, ceiling reflection, and speaker reflection of ceiling wave.

Figure 6b is an isometric scan of a multicellular horn from dc to 20 KHz and is a good representation of the principal pressure wave A of Fig. 4 with the finite space-equivalent bandwidth in this case equal to ½ ft. Figure 6c is an interesting example of multiple reflections. An acoustic suspension loudspeaker system was placed on a floor pointing upward at a hard ceiling, with an omnidirectional microphone halfway between speaker and ceiling. The sweep extends from dc to 20 KHz with a 20 KHz marker "fence" in evidence in the spectrograph. The space-equivalent bandwidth is ½ ft. Time starts at the receipt of the principal wave. The second peak corresponds to the ceiling reflection. The third peak is the reflection of the ceiling wave off the loudspeaker itself. Some of the energy is reflected by the loudspeaker grille and some by the speaker cones themselves yielding beautiful doppler data, and some energy penetrates through the loudspeaker to the floor.

ANALYTICAL VERIFICATION

The measurement technique of the previous paragraphs was developed in a highly intuitive manner. Although experimental measurements tend to verify the technique, it is still necessary to analyze two very important considerations: first, the nature of a repetitive glide tone, and second, the validity of identifying the spectrum analyzer display with the acoustic spectrum.

Fourier Spectrum of a Repetitive Linear Glide Tone

The system-forcing function developed for time-delay spectrometry has the characteristic shown in Fig. 7. Loosely speaking, it is a signal which has an instantaneous frequency linearly proportional to time for a total period of T seconds and then repeats the cycle indefinitely at this period. Clearly the signal must have a Fourier series spectrum of terms with periods which are integral submultiples of T. It cannot, in other words, be a continuum of frequencies. Yet the intuitive development presupposed a forcing function which not only did not have "holes" in the spectrum but also was of constant amplitude.

Readers familiar with frequency modulation will recognize that the signal of Fig. 7 is in reality a linear sawtooth frequency modulating a carrier halfway between the peak deviation frequencies. Let the maximum and minimum deviation frequencies be ω_2 and ω_1 respectively and the period of sweep T seconds. The function to be defined then has an instantaneous frequency ω_{inst} given by

$$\omega_{inst} = \left(\frac{\omega_2 - \omega_1}{T}\right)t + \left(\frac{\omega_2 + \omega_1}{2}\right) \triangleq \frac{D}{T}t + \omega_c \quad (8)$$

where the angular dispersion D and effective carrier frequency ω_c are introduced for simplicity. Since we are interested in time-dependent factors, the time phase dependence becomes

$$\phi(t) = \int_0^t \omega_{inst} dt = (D/2T)t^2 + \omega_c t. \quad (9)$$

The actual function of time which is to be expanded

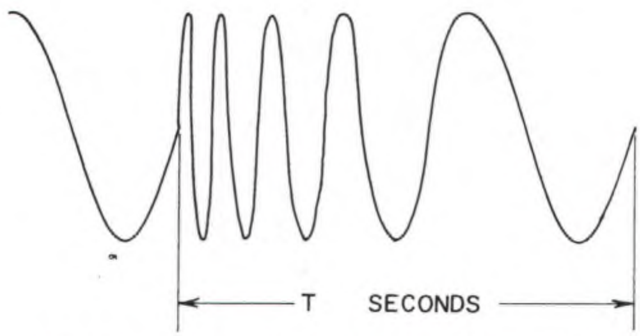


Fig. 7. Graphic representation of the repetitive sweep tone used as system-forcing function.

in a Fourier series is that signal possessing the phase $\phi(t)$, or

$$E(t) = \cos\phi(t) = \tfrac{1}{2}[e^{i\phi(t)} + e^{-i\phi(t)}]. \quad (10)$$

From an analytical standpoint it is simpler to use the exponential form. Since each complex exponential is separable into the product of a steady-state term and a term periodic in T, it is sufficient to expand this periodic term in a Fourier series. Furthermore, the positive frequency portion will be analyzed first and the negative frequency terms considered from this; thus the Fourier frequencies and coefficients become, in exponential form:

$$e^{i(D/2T)t^2} = \sum_{N=-\infty}^{\infty} C_N e^{iN\omega_0 t} \quad (11)$$

$$C_N = (1/T) \int_{-T/2}^{T/2} e^{i(D/2T)t^2} e^{-iN\omega_0 t}\, dt \quad (12)$$

where

$$\omega_0 = 2\pi/T.$$

By completing the square, multiplying and then dividing by a normalizing factor $(2/\pi)^{1/2}$, C_N becomes

$$C_N = \quad (13)$$

$$\frac{1}{T}\left(\frac{\pi T}{D}\right)^{1/2} e^{-i(T/2D)(N\omega_0)^2} \left(\frac{2}{\pi}\right)^{1/2} \int_{-T/2}^{T/2} e^{ix^2} dx$$

where

$$x = (D/2T)^{1/2}[t - (T/D)N\omega_0]$$

or, by noting Eq. (8),

$$C_N = \frac{1}{T}\left(\frac{\pi T}{D}\right)^{1/2} e^{-i(T/2D)(N\omega_0)^2} \quad (14)$$

$$\left[\left(\frac{2}{\pi}\right)^{1/2} \int_0^{\omega_2} e^{ix^2} dx - \left(\frac{2}{\pi}\right)^{1/2} \int_0^{\omega_1} e^{ix^2} dx\right]$$

$$x = (T/2D)^{1/2}[\omega_{inst} - (\omega_c + N\omega_0)].$$

Normally analysis would stop here because the definite integrals can be evaluated only as an infinite series. However, the complex expansion of the exponential integral

$$\left(\frac{2}{\pi}\right)^{1/2} \int_0^{u} e^{ix^2} dx = C(u) + iS(u) \quad (15)$$

is made up of the Fresnel cosine integral $C(u)$ and Fresnel sine integral $S(u)$, both of which have been tabulated.[6]

Noting the fact that the argument changes sign in the second integral, we can write the resultant Fourier coefficient for the N^{th} sideband term as

$$C_N = (1/T)(\pi T/D)^{1/2} e^{-i(T/2D)(N\omega_0)^2} \quad (16)$$
$$[C(\omega_1) + C(\omega_2) + iS(\omega_1) + iS(\omega_2)].$$

The C_n's are the coefficients of the Fourier components or, considered another way, are the sideband terms and consist of the product of the quantized spectrum of a continuous swept tone and a complex Fresnel integral modifying term involving the finite frequency terminations. Figure 8 illustrates how to calculate the coefficient for the N^{th} sideband; Fig. 9 is a plot of the locus of the complex value of the Fresnel integral term for a high deviation ratio, which is common for time-delay spectrometry. Note that the spectrum is completely symmetrical about the effective carrier and that while the dropoff is shown for one band edge, the other band edge is identical in form. It is seen that the magnitude is essentially constant throughout the spectrum, as expected, and that within the swept band the phase departure is limited to 15° and rapidly approaches 0° at center frequency.

The function

$$e^{ix^2} \quad (17)$$

goes through a first zero crossing for the real value, where $u = 1$ when

$$x^2 = \pi/2. \quad (18)$$

From Eq. (14), this occurs when

$$\Delta\omega = (2D/T)^{1/2}(\pi/2)^{1/2} = (\pi D/T)^{1/2}. \quad (19)$$

Since the relation between optimized bandwidth B and dispersion rate D/T for the analyzer is

$$B \cong (D/2\pi T)^{1/2} = [1/(2\pi)^{1/2}](D/T)^{1/2}, \quad (20)$$

then

$$\Delta\omega \cong B/\sqrt{2}. \quad (21)$$

In other words, even though Fig. 9 shows an overshoot at band edge of diminishing amplitude and increasing frequency as one approaches the effective carrier, these ripples are not distinguishable to a tracking analyzer since they are smoothed by the proper filter bandwidth and the analyzer cannot distinguish this distribution from a true

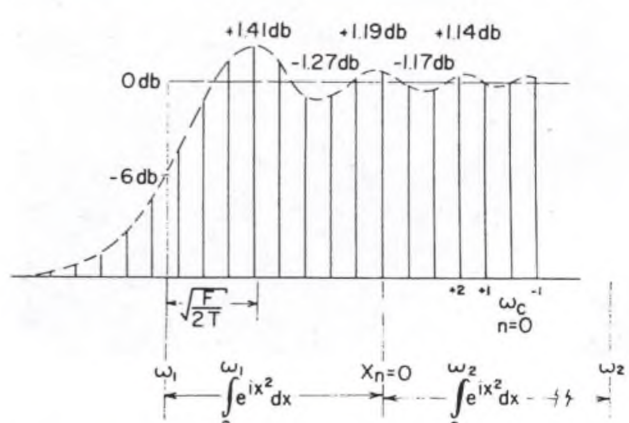


Fig. 8. Method of calculating sideband coefficients for linear sawtooth frequency-modulated carrier with peak angular frequencies ω_1 and ω_2, dispersion $D = 2\pi F = \omega_2 - \omega_1$, and sweep time T.

constant amplitude. Thus the desire for an effective constant amplitude continuum is satisfied.

While the conventional rates of time-delay spectrometry dictate a high effective modulation index yielding the near band-edge distribution of Fig. 9, it may be of some interest to investigators to determine the distribution for low effective indexes. This may be done quite simply by noting that a complex plot of Eq. (15) yields a Cornu spiral. This is shown in Jahnke and Emde[6] and appears

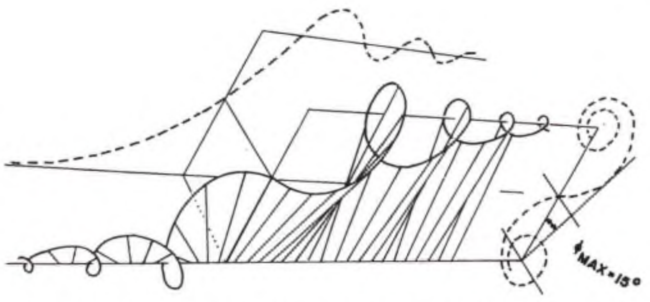


Fig. 9. Locus of the complex value, including magnitude and phase, of the Fresnel Integral term in the coefficient of the sweep tone Fourier series for high deviation ratios.

as the phase projection in Fig. 9. The desired sideband amplitude is the length of segment joining the appropriate points on the Cornu spiral. It should also be pointed out that for high indexes the spectral distribution will approach the amplitude probability distribution of the modulating waveform. For a perfect sawtooth this will be a square spectrum with the characteristic frequency terminating overshoot analogous to Gibb's phenomenon, while for a slightly exponential sawtooth the spectrum will appear to be tilted.

Several important points may now be established concerning this forcing function. First, the spectrum is quite well behaved for all modulation indexes. As one starts with an unmodulated carrier and begins modulation at progressively higher indexes, it will be noted that there is no carrier or sideband nulling as is the case with sinusoidal frequency modulation. Thus any deviation ratio is valid for time-delay spectrometry. Second, the spectral sidebands are down 6 dB at the deviation band-edge and are falling off at the rapid rate of about 10 dB per unit analyzer bandwidth. Thus the spectrum is quite confined and no unusual system bandwidth is required. Third, an expansion of the negative frequencies will show a comparable symmetric spectrum, which means that spectrum foldover effects such as those due to passing through zero beat do not in any way compromise the use of this type of modulation for time-delay spectrometry.

A final point in this spectral analysis: one tends to take for granted the validity of the analytical result without much consideration for the physical mechanism. Equation (16) shows that this glide tone can be generated by connecting a large number of signal generators with almost identical amplitude to a common summing junction. Furthermore, the amplitude of the resultant glide tone will be equal to that of any one of the separate generators. At first glance this would not only seem to violate conservation of energy but it would definitely be hard to convince an observer that when he heard a smooth glide tone with definite pitch as a function of time he was actually hearing all possible frequencies. Inspection of the phases as controlled by the partial Fresnel integrals reveals that at any given time all generators except those in the vicinity of the instantaneous glide tone frequency will cancel to zero, and that the distribution of those generators which are not cancelled is approximately Gaussian about the perceived tone. In addition, the "bandwidth" of this distribution is of the order of the analyzer bandwidth. The result of this is that an analyzer slightly detuned from the glide tone will not only be down in response but will be down by the same amount at all frequencies.

Spectrum Analyzer Response

In considering the nature of the display presented by a tracking spectrum analyzer used for time-delay spectrometry, the generalized circuit of Fig. 10 will be used. The portion within the dashed contour characterizes the general spectrum analyzer. The sweeping local oscillator signal is brought out and mixed in a balanced modulator with a fixed local oscillator which has a slight offset from the analyzer intermediate frequency. The filtered difference output of the mixing process is the audio glide tone previously discussed, which is used as a forcing function for the acoustical system under analysis. The acoustical response of this system as perceived by a microphone will consist of a multiplicity of time-delayed responses. The entire signal from the microphone is sent to the spectrum analyzer. The purpose of the analysis below is to demonstrate that any given delayed response may be selected from all inputs by appropriate selection of the glide tone generating offset oscillator, and to de-

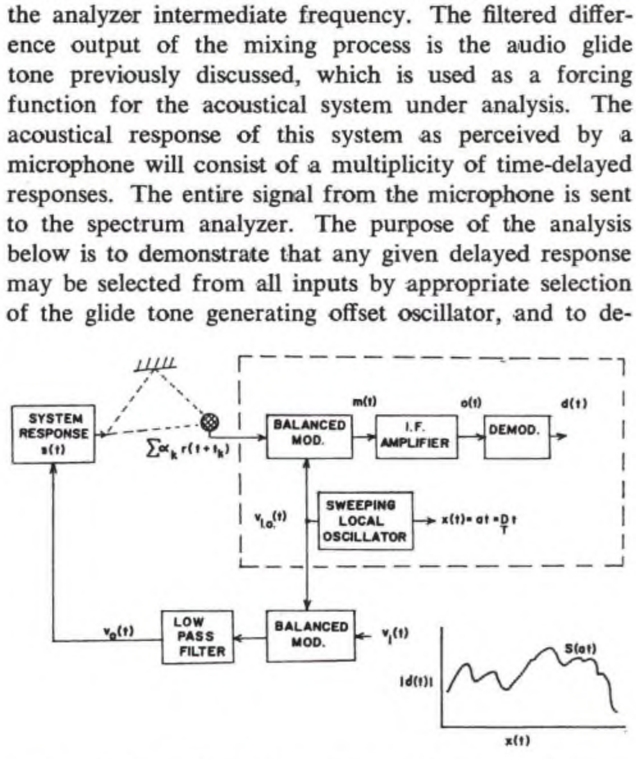


Fig. 10. Generalized block diagram of the acoustical system test using time-delay spectrometry.

velop conditions and limitations of analyzer display.

The analytical description of a sweeping tone used as a system-forcing function leads to a remarkable property: The sweep tone is the complex conjugate of its own Fourier Transform, with time in place of frequency. Just as a steady sinusoid applied to a network provides a narrow frequency window which pulls out the frequency response of the network, the sweep tone may be shown to provide a spectrum window pulling out the frequency spectrum of the network. As a consequence of this property this sweep function $w(t)$ will be designated as a window function,

$$w(t) \triangleq e^{i\frac{1}{2}at^2}, \qquad (22)$$

where a represents the angular dispersion rate D/T.

Proceeding to the analysis it will be observed that the analyzer local oscillator consists of a sweeping tone $w(t)$ heterodyned to the intermediate frequency ω_i. It is a real time function and is thus

$$v_{1.0}(t) = [w(t)e^{i\omega_i t} + w^*(t)e^{-i\omega_i t}]/2 \quad (23)$$

where the starred operation indicates complex conjugation. Similarly, the down-converting oscillator consists of the intermediate frequency ω_i offset by a fixed value ω_o and is

$$v_1(t) = [e^{i(\omega_i+\omega_o)t} + e^{-i(\omega_i+\omega_o)t}]/2. \quad (24)$$

After low-pass filtering, the system-driving function, neglecting constant gain terms, is

$$v_0(t) = w(t)e^{-i\omega_o t} + w^*(t)e^{+i\omega_o t}. \quad (25)$$

Assuming the system response is linear, the system response $r(t)$ for this driving function is the convolution integral of the driving function $v_0(t)$ and the system time response $s(t)$:

$$r(t) = \int_{-\infty}^{\infty} s(\tau)v_0(t-\tau)\,d\tau \triangleq s(t) \otimes v_0(t). \quad (26)$$

The second symbolism will be utilized because of its simplified form.

Because of the finite time delay, t_k, between system output and microphone input for each probable path K, the microphone response $r(t)$ consists of the sum of all the signal path inputs,

$$r(t) = \sum_K a_K r(t+t_k) \quad (27)$$

where a_K is the strength of the K^{th} signal. Also, each separate path length signal is expressible as

$$r(t+t_k) = [s(t+t_k) \otimes w(t+t_k)e^{-i\omega_o(t+t_k)}] + [s(t+t_k) \otimes w^*(t+t_k)e^{i\omega_o(t+t_k)}]. \quad (28)$$

The process of balanced modulation is a simple multiplication in the time domain. The modulator output $m(t)$ is then, neglecting constant multipliers,

$$m(t) = \sum_K a_K r(t+t_k)\cdot(w(t)e^{i\omega_i t} + w^*(t)e^{-i\omega_i t}). \quad (29)$$

The frequency spectrum of the modulator output $M(\omega)$, considering only positive frequencies, is

$$M(\omega) = \sum_K a_K \int e^{-i(\omega-\omega_i)t} w(t)r(t+t_k)\,dt. \quad (30)$$

(The convention followed for the purpose of this analysis is that lower-case functional notation designates time dependence while upper-case notation designates frequency dependence.)

From Eqs. (A5), (A6), and (A9) of the Appendix, Eq. (30) may be rewritten as

$$M(\omega) = \sum_K a_k \int e^{-i(\omega-\omega_i+at_k-\omega_o)t} W(at_k-\omega_o)(a/2\pi i) \{W(\omega_o)w(\xi)[s(\xi)\otimes w(\xi)] + W^*(\omega_o)w(\xi)[s(\xi)\otimes w^*(\xi)]\}\,dt \quad (31)$$

where $\xi = t+t_k-(\omega_o/a)$.

From Eqs. (A7) and (A8) in the Appendix, we may expand this into two integral summations

$$M(\omega) = \sum_K a_K \int e^{-i(\omega-\omega_i+at_k-\omega_o)t} W(at_k-\omega_o)(a/2\pi i)^{3/2} \{W(\omega_o)w(\xi)w(\xi)[S(a\xi)\otimes W(a\xi)]\}\,dt + \sum_K a_K \int e^{-i(\omega-\omega_i+at_k-\omega_o)t} W(at_k-\omega_o)(a/2\pi i)^{3/2} \{W^*(\omega_o)[S(a\xi)\otimes W^*(a\xi)]\}\,dt. \quad (32)$$

The first integral is the balanced modulator "upper sideband" and is the transform of a scanned time spectrum multiplied by two window functions. The transform will therefore be in the vicinity of the intermediate frequency only for $\xi = O$ and will constantly retreat from ω_i for all other times at a rate twice that of the sweeping local oscillator. As a consequence, this integral need not be considered further.

The second integral is the "lower sideband" and will transform the scanned spectrum to lie directly on the frequency $(\omega_i+\omega_o-at_k)$. For both integrals the "bandwidth" of any substantial energy to be found in the transformed spectrum will be determined by the convolution of $W(at)$ and $W^*(at)$ respectively with the system frequency spectrum, with the frequency parameter replaced by time. Figure 11 symbolizes the spectral energy distribution given by Eq. (32).

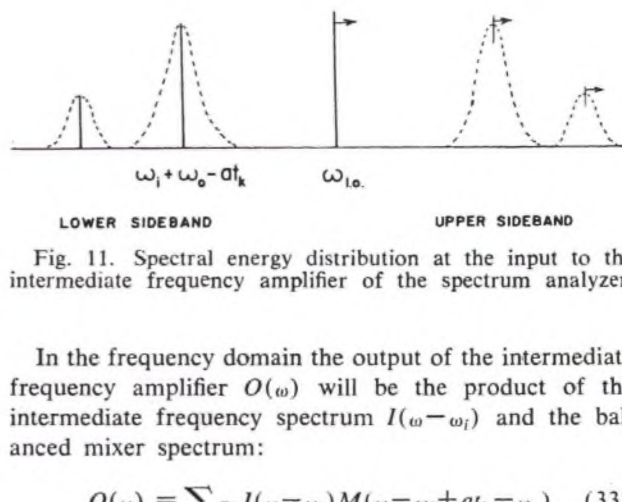


Fig. 11. Spectral energy distribution at the input to the intermediate frequency amplifier of the spectrum analyzer.

In the frequency domain the output of the intermediate frequency amplifier $O(\omega)$ will be the product of the intermediate frequency spectrum $I(\omega-\omega_i)$ and the balanced mixer spectrum:

$$O(\omega) = \sum_K a_k I(\omega-\omega_i)M(\omega-\omega_i+at_k-\omega_o). \quad (33)$$

If the delay times t_k are not so close together as to overlap the functions $M(\omega-\omega_i+at_k-\omega_o)$ in the vicinity of $(\omega-\omega_i)$, then any particular response may be selected by adjusting ω_o such that

$$\omega_o = at_k \quad (34)$$

for the desired path time delay t_k. Then Eq. (33) becomes

$$O(\omega) = G_K I(\omega-\omega_i)M(\omega-\omega_i). \quad (35)$$

The time spectrum of the output of the intermediate frequency amplifier is thus

$$o(t) = g_k[i(t)\otimes W^*(at)\otimes S(at)](a/2\pi i)^{3/2} \quad (36)$$

or, noting Eq. (A3) in the Appendix and regrouping in accordance with the associative property of convolution,

$$o(t) = g_k(a/2\pi i)\,[i(t) \otimes w(t)] \otimes S(at). \quad (37)$$

The demodulator output $d(t)$ is the signal displayed on the vertical axis and is thus

$$d(t) = (constant) \cdot |o(t)|. \quad (38)$$

Inspection of Eqs. (37) and (38) discloses that the proper spectrum has indeed been displayed. It should be recognized that while the analysis was done on a continuous basis no loss of validity is experienced for repetitive sweeps.

In interpreting Eq. (37) it is apparent that the proper spectrum is modified by the window function and time response of the intermediate-frequency amplifier of the analyzer. The action of these latter terms is a smoothing upon the actual spectrum. This smoothing is similar in character to a simple low-pass filtering of the perfect spectrum.

To demonstrate that no spectrum bias is introduced in Eq. (37) by $w(t)$ and $i(t)$, consider the case of a perfectly flat spectrum where

$$S(at) = Constant \quad (39)$$

and where a finite time is assumed. In this case the convolution integrals collapse to simple integrals yielding

$$o(t) = Constant. \quad (40)$$

If, in other words, the system spectrum is independent of frequency, the spectrum analyzer display will show a straight line whose height above the baseline will be proportional to the gain or loss through the system.

Looking at the other end of the functional dependence, consider the case of a system spectrum response which is zero for all frequencies but one, and is infinite at that singular frequency. What, in other words, is the response to a network of infinite Q. Such a function is a Dirac delta

$$S(at) = \delta(t - t_o) \quad (41)$$

simple substitution into Eq. (37) yields

$$o(t) = K[i(t) \otimes w(t - t_o)]. \quad (42)$$

This is exactly the same response as experienced by a spectrum analyzer viewing a single sinewave spectrum.[4] This zero width spectrum is displayed as a smoothed function, closely approximating the intermediate frequency response for sweep rates such that

$$B^2 \geqq dF/dt. \quad (43)$$

As the sweep rate increases above the inequality of Eq. (43), the displayed function broadens and the "phase tail" of $w(t) \otimes i(t)$ evidences an increasing ring on the trailing edge. Thus for the normal spectrum analyzer the infinite Q response is smoothed to the order of the analyzer bandwidth.

One exceedingly important result follows from Eq. (37). This is that the spectrum is preserved not only in amplitude but also in phase. Thus the actual spectrum experiences both an amplitude and a phase smoothing. This gives us a measurement tool which could not have been predicted from the intuitive reasoning utilized. The implications are significant, since it is now possible to isolate the influence of an acoustic subsystem without removing that subsystem from its natural environment and analyze complex behavior, amplitude, and phase, in response to an applied stimulus.

CONCLUSION

What has been described is an acoustical testing procedure which allows selective spatial probing of a natural environment by commercially available equipment. The results are displayed immediately in the form of pressure response as a function of frequency, and require no further processing for interpretation. There are of course resolution limitations imposed by the dimensions of the testing area, but these are easily calculated and are small penalties to pay for the privilege of making formerly unobtainable on-location tests.

An acoustic model of reflecting objects has been introduced which identifies the object with an equivalent frequency response expressible as a frequency-dependent scattering coefficient. This frequency response modifies the frequency response of a sound source in a manner analogous to its electronic circuit counterpart, but possesses a time delay corresponding to its position relative to source and auditor. The effect of an assemblage of reflecting objects can be represented as a three-dimensional plot of pressure, frequency and time.

The utilization of time-delay spectrometry is limited primarily by the imagination and ingenuity of the experimenter. The work described in this paper attempted to yield detailed insight into the technique yet not prevent its application because specialized equipment was unavailable. It must be clear that there is still a substantial class of meaningful measurements which not only require extensive modification of the equipment but also demand further understanding and analysis, in particular the phase angle which is analytically available as well as the amplitude of response. Electronic circuit concepts were used to develop the technique as well as the acoustic model. Perhaps once we possess detailed complex spectra of various acoustic systems we may be able to utilize the existing wealth of circuit techniques to reduce the multi-dimensional real-world acoustic problems to a form more amenable to analysis. It may well be that time-delay spectrometry is a tool which will contribute to this goal.

APPENDIX

Summary of Important Relations

1. $f(t) = (1/2\pi) \int_{-\infty}^{\infty} F(\omega) e^{i\omega t} d\omega,$

$$F(\omega) = \int_{-\infty}^{\infty} f(t) e^{-i\omega t} dt$$

2. $w(t) = e^{i\frac{1}{2}at^2} \qquad W(\omega) = (2\pi i/a)^{\frac{1}{2}} e^{-i(\omega^2\, 2a)}$

3. $w(t) = (a/2\pi i)^{\frac{1}{2}} W^*(at)$

4. $\int_{-\infty}^{\infty} f(t) g(x-t)\, dt \triangleq f(x) \otimes g(x)$

5. $s(t+t_k) \otimes w(t+t_k)\, e^{-i\omega_o(t+t_k)} = (a/2\pi i)^{1/2} W(\omega_o)[s(t+t_k-\omega_o/a) \otimes w(t+t_k-\omega_o/a)]$

6. $s(t+t_k) \otimes w^*(t+t_k)\, e^{i\omega_o(t+t_k)} = (a/2\pi i)^{1/2} W^*(\omega_o)[s(t+t_k-\omega_o/a) \otimes w^*(t+t_k-\omega_o/a)]$

7. $w(t)[s(t) \otimes w(t)] = w(t)\,w(t)\,[S(at) \otimes W(at)](a/2\pi i)^{1/2}$

8. $w(t)[s(t) \otimes w^*(t)] = [S(at) \otimes W^*(at)](a/2\pi i)^{1/2}$

9. $w(t) = w(t+t_k-\omega_o/a)\, W(at_k-\omega_o)\, e^{-i(at_k-\omega_o)t}(a/2\pi i)^{1/2}$

REFERENCES

1. W. B. Snow, "Loudspeaker Testing in Rooms," *J. Audio Eng. Soc.* **9**, 54 (1961).
2. D. H. Bastin, "Microphone Calibrator," *Electronics* (Nov. 1948).
3. T. J. Schultz and B. G. Watters, "Propagation of Sound Across Audience Seating," *J. Acoust. Soc. Am.* **36**, 885 (1964).
4. H. W. Batten, R. A. Jorgensen, A. B. MacNee and W. W. Peterson, "The Response of a Panoramic Analyzer to C.W. and Pulse Signals," *Proc. I.R.E.*, 948 (1954).
5. G. E. Peterson, G. A. Hellwarth, and H. K. Dunn, "An Anechoic Chamber with Blanket Wedge Construction," *J. Audio Eng. Soc.* **15**, 67 (1967).
6. For example, E. Jahnke and F. Emde, *Tables of Functions* (Dover Publications, New York, 1945).

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 Cal Tech Jet Propulsion Laboratory in Pasadena, California. He is a senior member of the JPL Technical Staff presently engaged in the lunar television operation on Project Surveyor.

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, National Professional Engineering Fraternity.

Reprinted from *JAES*, vol. 17, no. 1, pp. 30–41, 1969 January.

Loudspeaker Phase Characteristics and Time Delay Distortion: Part 1

RICHARD C. HEYSER

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

A technique is described for measurement of the complete frequency response of a loudspeaker, including amplitude and phase. A concept of time delay is introduced which provides a physical description of the effect of phase and amplitude variations as a frequency-dependent spatial smearing of the effective acoustic position of a loudspeaker.

INTRODUCTION In a previous paper a derivation was given for a new acoustic testing technique which allowed anechoic measurements to be taken in a normally reverberant environment [1]. The quantity measured was shown analytically to be the complex Fourier transform of the system impulse response of any signal with a fixed time delay. When applied to loudspeakers, this means that beside the conventionally measured pressure amplitude spectrum there is a pressure phase spectrum, which to the best information available to this author has not been as well investigated. In measuring both the amplitude and phase spectrum of some common loudspeaker types it was immediately evident that some peculiarities in the behavior of phase were not apparent from an inspection of amplitude alone. Since the measurement technique allows a subtraction of time delay incurred between the electrical stimulus applied to the loudspeaker terminals and the pressure wave incident upon the test microphone. it was apparent from the beginning that there was in many cases a frequency-dependent time delay in excess of that caused by the travelling of the pressure wave over the known distance from voice coil to microphone diaphragm. Attempts at understanding this delay in terms of an equivalent electronic transfer function were not satisfactory: the only concept of circuit time delay which might prove useful is that of group delay. also called envelope delay, and it was not evident exactly how one might go about application of this concept. particularly when open literature definitions are accompanied with disquieting phrases such as ". . . when the amplitude does not change rapidly with frequency . . ." and ". . . if there is no absorption . . .". In fact it is rather disturbing that a substitute more in alignment with nature is not used in those cases when this so-called time delay becomes negative, a condition found to be quite common in loudspeaker measurements. For this reason it became necessary to derive a time delay more in keeping with concepts of causality as well as to investigate the amplitude and phase relationships in loudspeakers. The resulting concept of a spatial "spreading out" of the effective acoustic position of a loudspeaker behind its physical position would appear to be a more unified approach to time delay in such a complicated network. This paper is a documentation of some characteristics which have been measured on typical loudspeakers, as well as the derivation of a concept of time delay useful for interpreting these characteristics. The ultimate purpose is to provide audio engineers working on loudspeaker design a criterion by which a loudspeaker may be equalized to provide a more perfect response than would be possible by the use of pressure response measurements alone.

THE LOUDSPEAKER AS A NETWORK ELEMENT

In considering the role of a loudspeaker in the reproduction of sound. it would appear logical to describe this device as a network element. Although a transducer of electrical to acoustic energy, a loudspeaker may dissipate power, store energy, modify frequency response, introduce distortion, and in general impart the same aberrations in its duty as any conventional electronic network. Unlike a conventional network, however, a loudspeaker interfaces with the spatial medium of a human interpreter and may possess characteristics in this medium, such as polar response and a time-delayed phenomenon, which are unlike a normal electronic network element. While recognition of these latter effects exists, measurement has proven cumbersome.

In the discussion to follow a loudspeaker will be considered as a general network element. This element will have some transfer function relating the output sound pressure wave to the applied electrical stimulus. The output of the loudspeaker will be considered to be its free-field response. Where severe interaction with its environment is desired, as for example in the case of a corner horn or wall-mounted dipole, then that portion of the environment will be included. No simplification to an equivalent circuit is sought; indeed, the system is assumed to be so complicated that the only knowledge one has is a direct acoustic response measurement.

DEFINITION OF FREQUENCY RESPONSE

The transfer function of a loudspeaker will be assumed independent of signal level. This considerably simplifies the analysis and is justified when one considers that deviations from linearity of amplitude are much less than deviations from uniform frequency response. This means that, as in linear circuit theory, one can now define the transfer function to be a function of frequency. Our concern then rests with determining the frequency response of a loudspeaker.

The frequency response of a loudspeaker will be defined as the complex Fourier transform of its response to an impulse of electrical energy. If one considers the magnitude of the frequency dependent pressure response, this definition coincides with the practice of using a slow sinewave sweep in an anechoic chamber and plotting the pressure response as a function of frequency [17]. This definition, however, also includes the phase angle as a frequency-dependent term. Thus if a loudspeaker has an amplitude response $A(\omega)$ and a phase response $\phi(\omega)$ the loudspeaker response $S(\omega)$ is

$$S(\omega) = A(\omega)e^{i\phi(\omega)}. \quad (1)$$

It quite frequently happens that the amplitude response is better characterized in decibels as a logarithmic function, $a(\omega)$, which leads to the simplified form

$$S(\omega) = e^{a(\omega)}e^{i\phi(\omega)}. \quad (2)$$

The frequency response of a loudspeaker will then consist of two plots. The plot of magnitude of sound pressure in decibels as a function of frequency, $a(\omega)$, will simply be called amplitude. The plot of phase angle of sound pressure as a function of frequency, $\phi(\omega)$, will be called phase. These taken together will completely characterize the linear loudspeaker frequency response by the relation

$$\ln S(\omega) = a(\omega) + i\phi(\omega). \quad (3)$$

Equation (3) allows us to state that the amplitude and phase are the real and imaginary components of a function of a complex variable. This complex function is thus a more complete definition of what is meant by the frequency response of a loudspeaker. From this we see that measuring amplitude alone may not be sufficient to specify total speaker behavior.

THE LOUDSPEAKER AS A MINIMUM PHASE NETWORK

The measurement of loudspeaker pressure response is not a new art, and substantial definitive data has existed for over 40 years [17]. With a few very rare exceptions these measurements have been of the amplitude response [23, 24, 25]. Indeed, it is common practice to call this the frequency response of a loudspeaker.

It is of more than casual concern to investigate the conditions under which the measurement of amplitude response is sufficient to characterize the complete behavior of a loudspeaker. Obviously, whenever this is the case, the measurement of phase is academic, and a conventional amplitude response measurement should continue as a mainstay of data. Referring to Eq. (3) one may inquire into the conditions under which a prescribed magnitude function results in a definite phase angle function and conversely.

For a complex function such as expressed in Eq. (3) there will exist a unique relationship between the real and imaginary parts, determined by a contour integral along the frequency axis and around the right half-plane, if the logarithm is analytic in this right half-plane [2]. A network which meets this requirement is called a minimum phase network. Thus, if a loudspeaker is a minimum phase network, the measurement of either phase or amplitude is sufficient to characterize the frequency response completely.

When one has a minimum phase network the amplitude and phase are Hilbert transforms of each other related by [2, 3, 5, 6]

$$a(\omega) = \frac{1}{\pi} P \int_{-\infty}^{\infty} \frac{\phi(x)}{\omega - x}\, dx \quad (4)$$

$$-\phi(\omega) = \frac{1}{\pi} P \int_{-\infty}^{\infty} \frac{a(x)}{\omega - x}\, dx, \quad (5)$$

where P indicates that the principal part of the integral is to be taken at the pole ω. These integrals are the counterpart of Cauchy's differential relations on the frequency axis and are a sufficient condition to ensure that the loudspeaker does not have an output prior to an input signal, that is, $f(t) = 0$ for $t < 0$. While it may appear absurd to even concern oneself with the obvious fact that a minimum phase network cannot predict the output, it is quite important to the concept of group delay for a minimum phase function. It can be shown that group delay is not coincident with signal delay for a minimum phase network [26].

Given that one seeks confirmation of minimum phase behavior in a loudspeaker, how can one determine when this condition is actually achieved? It is unfortunate that one cannot determine this from either amplitude or phase alone. This follows either from an analysis of Eqs. (4) and (5) or the Cauchy-Riemann equations (Eq. A2 in Appendix A) or from a consideration of the factors needed to say that there are no zeros in the right half-plane. Thus it would seem that it is necessary to measure both amplitude and phase in order to determine if a measurement of amplitude alone would have been sufficient. On its face value, this is a convincing argument for measuring phase.

Hard on the heels of such an analysis comes the query as to why one would be concerned whether a loudspeaker is minimum phase or not. The answer comes from a very important property of minimum phase networks which may be paraphrased [4] as: If a loudspeaker is a minimum phase network then it can be characterized as a ladder network composed of resistors, capacitors, and inductors, and there will always exist a complementing network which will correct the loudspeaker frequency distortion as closely as one chooses. Thus, if one assumes linearity, a minimum phase loudspeaker could be completely compensated to become a distortionless transducer.

Consider what this means if one does not have a minimum phase loudspeaker. Assume that a pressure response (amplitude) measurement is made on a loudspeaker in an anechoic chamber. The speaker will of course have peaks and dips in its response. One might naturally assume that by diligent work it would be possible to level out the peaks and fill in the dips by a combination of electrical and mechanical means to produce a loudspeaker with a considerably smoother measured response. However, if one does this with a non-minimum phase loudspeaker, even if it is non-minimum phase over only a portion of the spectrum, one will have a transducer with a smoother $a(\omega)$ but quite probably a considerably distorted $\phi(\omega)$, having inadvertently created a complicated all-pass lattice network. As shown in Appendix B and elsewhere [26], this is equivalent to a dispersive medium. When program material is fed into this "equalized" loudspeaker the odds are that the sound heard by a listener will be considerably more unpleasant than that coming through the loudspeaker without benefit of such equalization.

The unhappy consequence of this type of experience will be an assumption that the anechoic chamber response bears little relationship to the quality of reproduction, and that perhaps the best loudspeakers were indeed made in a previous generation. If, on the other hand, one knew what portions of the frequency spectrum were minimum phase, equalization of that portion *would* improve the response, provided one did not attempt equalization on the remaining non-minimum phase portions. Thus if one is concerned with improving the quality of sound reproduction by means of passive or active equalization one is immediately interested in phase response of a loudspeaker and determination of minimum phase criteria.

TIME DELAY IN A LOUDSPEAKER

A network is said to be dispersive if the phase velocity is frequency-dependent, which means that the phase plot is not a linear function of frequency [11]. In pursuing the mathematical analysis of network transfer functions one eventually arrives at considerations of the effect of dispersion on time delay of a signal through a network. Perhaps the best that can be said of such considerations is that they are seldom lucid and almost never appear applicable to the problem at hand. One must nonetheless recognize that group delay (envelope delay) and phase delay are legitimate manifestations of the perturbations produced on a signal by a network. In this section we shall investigate some general considerations of time delay and apply them to the loudspeaker.

There is, of course, a very real time delay incurred by an acoustic pressure wave as it travels from loudspeaker to listener. The velocity of sound in air at a fixed temperature is constant, at least to the precision required for this discussion. Furthermore, this velocity is not a function of frequency. If a pressure wave travels a given distance at a given velocity it takes a period of time expressed as the quotient of distance to velocity.

When one considers the loudspeaker one can no longer assume a non-dispersive behavior. Appendix B contains a derivation of the effect of a dispersive medium (without absorption) on the transfer function of a network. It is shown that if the time lag is a function of frequency the effect will be an excess phase lag in the transducer transfer function. A very important result of this is that one should look first at the phase terms when considering time delay. Equations (4) and (5) indicate that, at least for a minimum phase network, either phase or amplitude give the same information. Bode [4] has shown that for a minimum phase network the phase lag, and hence to some measure time lag, of the low-frequency portions of the spectrum are governed by the amplitude function at high frequencies. Since it is known that the amplitude response of a loudspeaker must fall off at some sufficiently high frequency, there is some justification to believe that there will be an additional time delay in a loudspeaker due to the rolloff of high-frequency response. Stated another way, the acoustic position of a loudspeaker should, on the average, lie behind its physical position by an amount that is some inverse function of its high-frequency cutoff. The acoustic position of a woofer will be further behind the physical transducer than that of a tweeter. This important fact is quite frequently overlooked by engineers who consider that spatial alignment of voice coils is sufficient to provide equal-time path signals from multirange loudspeakers.

Having thus considered air path delay and time lag in the loudspeaker due to high frequency cutoff, one comes to the seemingly nebulous concept of dispersive lag in the transducer. Two types of so-called time delays have been defined for the purpose of expressing the distortion of a signal passing through a medium. These are phase delay and group delay, also called envelope delay [7-8]. Phase delay expressed in seconds is a measure of the amount by which a sinusoid disturbance of fixed frequency is shifted in phase after passing through a network. This conception is applicable only for a sinewave and only after total equilibrium is achieved, and is consequently of no use in considerations of the realistic problem of aperiodic disturbances. Therefore, we will not consider phase delay any further.

Group delay, also expressed in seconds, is an attempt at expression of the relative time shift of signal frequency components adjacent to a reference frequency. When used in this manner and applied to a medium with a substantial time delay (in the classic sense of distance traversed at a finite velocity) group delay is of benefit in distortion analysis. Group delay has caused a considerable confusion among engineers who have attempted to stretch the definition beyond the frequency range within which it is valid. Group delay is expressible very simply as the slope of the phase-frequency distribution; thus for a network with the transfer function of Eq. (2),

$$t_{Group} = -d\phi(\omega)/d\omega. \tag{6}$$

Group delay, although expressed in the dimension of time, is not a satisfactory substitute for the engineering concept of time delay which one intuitively feels must be present in a medium. There is a strong desire to ascribe a causal relation between an applied stimulus to a loudspeaker, for example, and the emergent pressure wave which as a premise, must have some time lag. There are, however, enough examples of total inapplicability—so called anomalous dispersion—to create suspicion regarding the primacy of group delay.

It can be shown that there is indeed a time delay phenomenon in a network more in alignment with engineering experience [26]. This delay is not necessarily single-valued, but may be multiple-valued or possibly a time distribution. An engineering interpretation which can be put on the time delay of a network may be secured by investigating the behavior of a given frequency component of an input signal as it finally emerges in the output. It is true that a sudden change in a parameter at the input will in effect create a broad spectrum around the particular frequency the delay of which we wish to characterize. However, this does not invalidate the premise that there will still exist a spectral component of this frequency, and one may legitimately ask what happens to that spectral component. It is shown that it may take some finite time for the component to appear at the output and that for a general network there may be many components of the same frequency arriving with different time delays.

This multiplicity of delayed outputs at a given frequency means that in the case of a loudspeaker one could also think in terms of a number of loudspeakers arrayed in space behind the physical transducer in such a way that the air-path delay of each produces the appropriate value of delay. Each frequency in the reproduced spectrum will then possess some unique spatial distribution of the equivalent loudspeakers. The emergent sound pressure wave will be a perfect replica of electrical signal only if these equivalent loudspeakers merge into one position for all frequencies since only then will there be no frequency-dependent amplitude or phase terms and all signal components will arrive at the same time [26].

Figure 1a is a highly schematic attempt at illustrating the phenomenon of loudspeaker time delay. A single loudspeaker is assumed, with a physical position in space indicated by the solid line. The effective position of the source of sound as a function of frequency is shown by speaker symbols which lie behind the physical location of the actual speaker by an amount determined by the delay time and velocity of sound in air. There may be many such equivalent perfect speakers distributed in space with more or less energy from each. The average position of the distribution of these equivalent speakers is

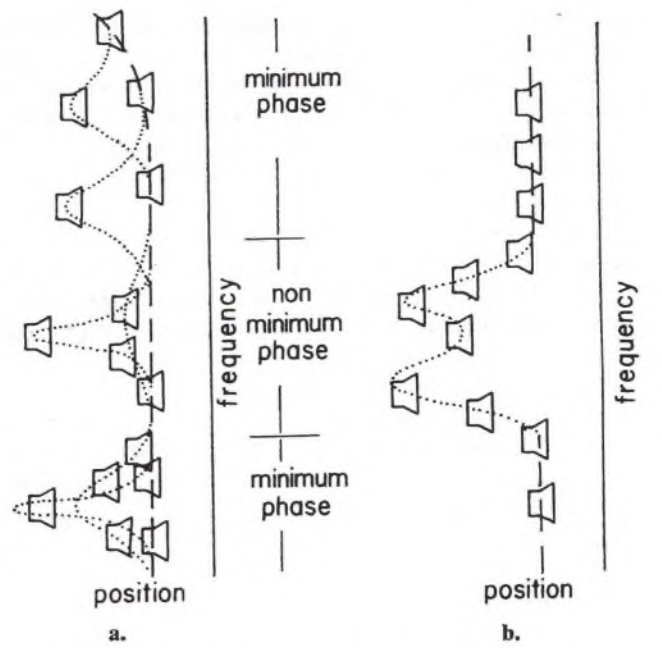


Fig. 1. **a.** Symbolic representation of the effect of frequency-dependent amplitude and phase variations of a single loudspeaker as equivalent to an assemblage of perfect loudspeakers distributed in space behind the physical position of the single loudspeaker. Each of the equivalent perfect loudspeakers has a flat amplitude and phase pressure response but assumes a frequency-dependent position indicated by the **dotted lines**, and in general differs in amount and polarity of energy radiated. The physical position of the actual radiator is indicated by the **solid line** while the average acoustic position determined by high-frequency cutoff lies behind this at the position of the **dashed line. b.** Symbolic representation of the acoustic effect of a response equalization based entirely on correcting the amplitude variations of the speaker of Fig. 1a. The frequencies at which the unequalized speaker is minimum phase may be corrected to perfect response indicated by a space-fixed perfect loudspeaker, but excess time delay results for non-minimum phase frequencies.

indicated by the dashed line, and corresponds to the delay attributed to high-frequency rolloff. Those frequencies at which the original loudspeaker is minimum phase are shown. Note that at no frequency will an equivalent speaker be found in front of the physical speaker. This is a result of the obvious fact that a loudspeaker can have no response prior to an input.

In Fig. 1b the result of amplitude equalization alone is symbolized. The equivalent speakers coalesce into one for the minimum-phase frequencies but spread out for all others.

TECHNIQUE OF MEASUREMENT

Analysis of the technique of time delay spectrometry has shown that the intermediate frequency amplifier contains the complex response of Eq. (2) with the frequency parameter ω replaced by a time parameter at [1]. This complex spectrum is convolved with a modifying term which is composed in turn of the convolution of the impulse response $i(t)$ of the intermediate frequency amplifier with the sweeping window function $w(t)$. The derivation of this relation is presented in a previous paper [1] and is sufficiently lengthy that it will not be reproduced here.

Equation (37) of that paper shows that if a loudspeaker is the subject of this test, the time function one finds in the intermediate frequency amplifier is $o(t)$, where

$$o(t) = (Gain\ Factor)[i(t) \otimes w(t)] \otimes S(at). \quad (7)$$

The effect of the convoluting, or folding integrals, indicated by the symbol $\otimes$ is a scanning and smoothing of the response $S(at)$. This means that if one observes appropriate precautions in sweep rate a, the intermediate frequency amplifier contains a signal which may be interpreted as

$$o(t) = Smoothed\ S(at). \quad (8)$$

Expressing this differently, one might say that the desired spectrum (Eq. 2) is contained in the equipment but all "sharp edges" have been smoothed off. The amount of smoothing is a function of the bandwidth of the equipment and the rapidity of the sweep. What is important is that there are no surprises or genuinely false patterns created by time delay spectrometry. If the loudspeaker contains a peak in response, then this peak will show in the analysis.

Having established the fact that time delay spectrometry will yield the complex transfer function of Eq. (2), it remains to see how both amplitude and phase may be extracted. It is assumed that the previous paper [1] will be used as reference for the basic technique. Figure 2 is a simplified block diagram of a time delay spectrometry configuration capable of measuring both amplitude and phase of a loudspeaker response.

In Fig. 2 a crystal oscillator is used as a source of stable fixed frequency. A countdown circuit derives a rate of one pulse per second which triggers an extremely linear sawtooth generator. This sawtooth provides a horizontal display to an oscilloscope as well as drive to a voltage-controlled sweep oscillator. The crystal oscillator is also used as a frequency source for two frequency synthesizers. A fixed frequency synthesizer converts the crystal frequency to the frequency of the intermediate frequency amplifier for the purpose of providing a reference to the phase detector. A tunable frequency synthesizer with digital frequency control is used to provide a precise offset frequency to down-convert the sweeping oscillator to the audio range; the filtered audio-range sweeping tone is then used to drive the speaker under test. A delay line is used between the sweeping oscillator and the microphone-channel balanced mixer for the purpose of precise cancellation of the time lag of the down-conversion lowpass filter. The intermediate-frequency amplifier containing the up-converted microphone spectrum feeds both a limiter-phase detector channel and a logarithmic amplifier-amplitude demodulator channel. The output of either the amplitude or phase detectors may be selected for display on the vertical axis of the oscilloscope which has the sawtooth horizontal drive.

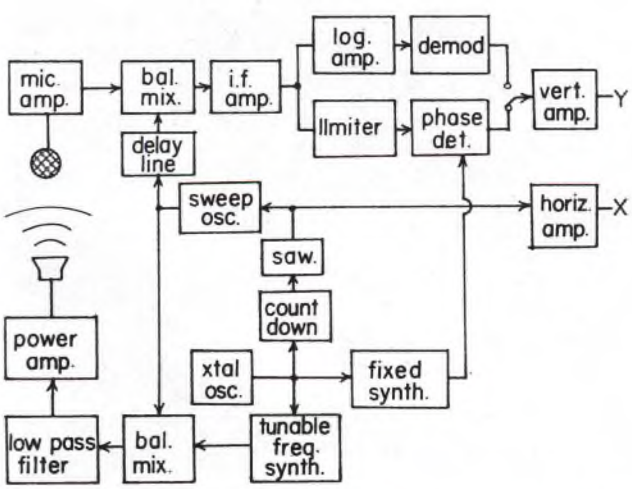


Fig. 2. Block diagram of a time delay spectrometry configuration capable of measuring both amplitude and phase frequency response of a loudspeaker.

It will be recognized that the demodulation of phase information has necessitated a considerable increase in circuit complexity over that required for amplitude alone. Particularly important is the fact that subtraction of the free air-path time lag and a stable display of phase not only requires a fixed offset frequency precisely divisible by the frequency of the sawtooth sweep but also necessitates a coherent phase-stable locked loop involving the time lag of the measurement air path.

As an illustration of the precision required, assume that one is measuring a response from dc to 20 kHz in a one second sweep and that an intermediate frequency of 100 kHz is being used. An offset of one part in 100,000 of the oscillator with respect to the phase standard will produce 360° of phase drift in one second. With a high-persistence phosphor, standard with spectrum analyzers, a distinct image blur will occur with a drift of 3.6° per sweep, yet this requires the oscillators to be offset by no more than one part in ten million. The use of a sampling attachment with an x-y plotter might require two orders of magnitude of stability over this value, i.e., oscillators which if multiplied up to 1,000 MHz would differ by no more than one Hz. This requires a self-consistent set of frequencies provided by obtaining all frequencies from a single oscillator.

Distance-measuring capability is similarly impressive. If one is measuring the phase response in the assumed 20 kHz band, one has a wavelength of approximately 650 mil at the highest frequency. A physical offset in distance of only 6 mil between loudspeaker and microphone will produce a 3.6° phase change at 20 kHz. This corresponds to a ½ μsec time difference in path length.

While the phase measurement provides an enormously sensitive measure of time delay, it should not be implied that it is at all difficult to obtain good data. It is unusual when there is sufficient motion of the air path or transducers to upset a phase measurement, and such an effect is visible within the time of two sweeps of the display.

In making measurements on a loudspeaker, the microphone is positioned at the desired speaker polar angle and at a distance consistent with the spectrum sweep rate and closest reflecting object. With the loudspeaker energized by the sweeping tone, the closest integer offset frequency is dialed which corresponds to the relation

$$F_0 = (X/c)(dF/dt) \quad (9)$$

where F_0 is the synthesizer offset in Hz, X is the distance from speaker to microphone in feet, c is the velocity of sound in feet per second, and the sweep rate is measured in Hz per second.

The amplitude response should be at or near its peak value and require no further adjustment. The displayed phase response will quite generally be tilted with respect to the frequency axis at a slope which is proportional to the deviation of the offset oscillator from the proper value. An offset discrepancy of only two Hz will accumulate 720° of additional shift in a one second sweep. The proper offset frequency is easily found by observing the phase display and using that value which produces the most nearly horizontal phase plot. If the value is between two integer frequencies with a one second sweep, it may be helpful to slightly reposition the microphone toward or away from the speaker. Having found the required offset frequency it is more than likely that the display will pass through 360° at one or more frequencies, producing phase ambiguities. The phase of the offset synthesizer must then be slipped until the most nearly optimum display results. This is most readily done by dialing an offset which is 0.1 Hz above the integer Hz value, which will cause the phase display to move vertically by 36° per sweep. When the proper pattern is obtained the 0.1 Hz digit is removed, leaving a stationary pattern of phase *vs* frequency.

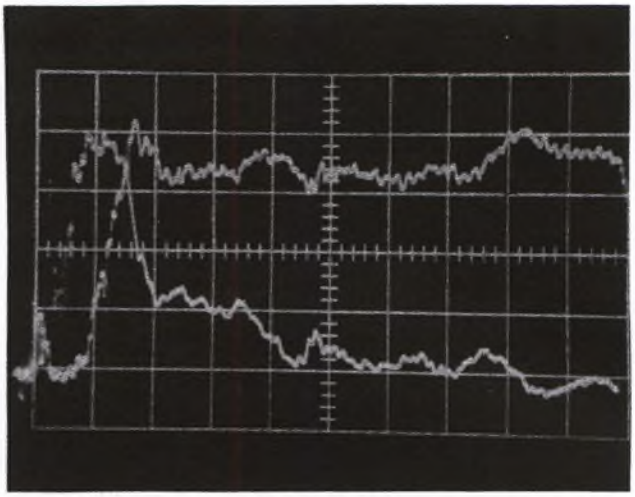


Fig. 3. Simultaneous amplitude (**upper**) and phase (**lower**) response of an inexpensive horn-loaded compression tweeter. Response is shown from dc to 10 kHz at 1 kHz per horizontal division. Amplitude is 10 dB per division and phase 60° per division.

MEASUREMENTS ON TYPICAL LOUDSPEAKERS

In the analysis of typical loudspeaker amplitude and phase characteristics it is desirable that a simultaneous display of both functions be made so that variations in either parameter at any given frequency may be compared. The oscillographs of this section are therefore dual-trace plots with the following characteristics: 1. The horizontal axis represents increasing frequency to the right with a linear scale; 2. Amplitude is the upper plot with logarithmic response and increasing signal as an upward deflection; 3. Phase is the lower plot with phase lead as an upward deflection. These characteristics, with the possible exception of a linear frequency scale, are among the standard conventions of network analysis.

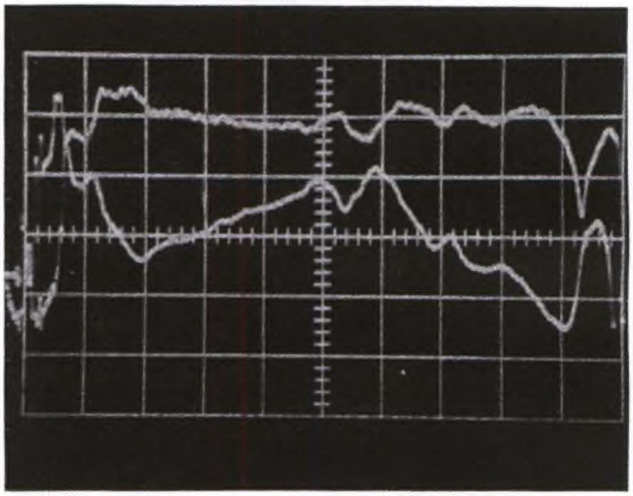


Fig. 4. Amplitude (**upper**) and phase (**lower**) response of midrange horn-loaded driver. Coordinates identical to those of Fig. 3.

Figure 3 is an example of an inexpensive horn-loaded compression tweeter. The plot extends from zero to 10 kHz with a frequency deflection factor of 1 kHz per cm (one cm is represented by a major division). The amplitude deflection factor is 10 dB per cm and the phase is 60° per cm. By referring to Fig. A-2 in the appendix it can be seen that this speaker is of the minimum phase type. The peaks at 1.6 kHz, 4 kHz, 7 kHz, and 8 kHz, and the dips at 4.6 kHz and 7.3 kHz have the appropriate phase fluctuation. Of particular interest is the peak at 1.6 kHz, identified by the phase change as a primary resonance in the driver occurring above horn cutoff.

Figure 4 is an example of a midrange horn-loaded speaker of the non-minimum phase type. Small variations are of minimum phase, including the dip at 9.4 kHz, but inspection of the phase identifies the fact that the speaker has a frequency-dependent spatial location. The acoustic position of the farthest forward equivalent speaker as determined by the offset frequency in the spectrometer is approximately 0.75 in. behind the position of the phase plug. This is in general agreement with the delay one would expect with a 10 kHz cutoff. From 2 kHz to 5.5 kHz the phase slope with independence of amplitude indicates a position which might be called Position 1. From 5.5 kHz to around 9.4 kHz the acoustic position

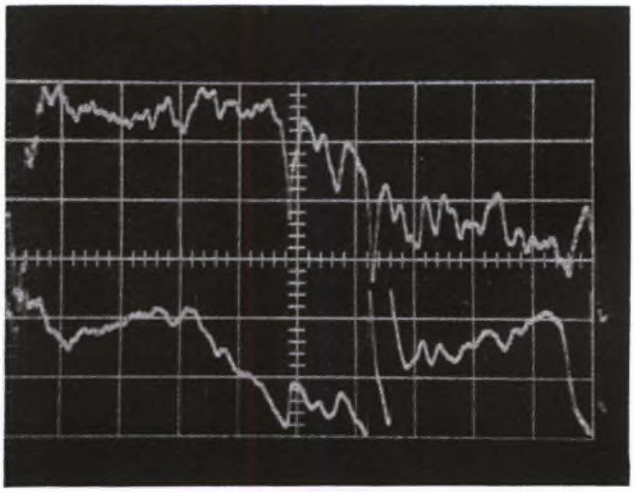


Fig. 5. DC to 20 kHz amplitude and phase response of the speaker of Fig. 4. Amplitude is 10 dB per division and phase 120° per division.

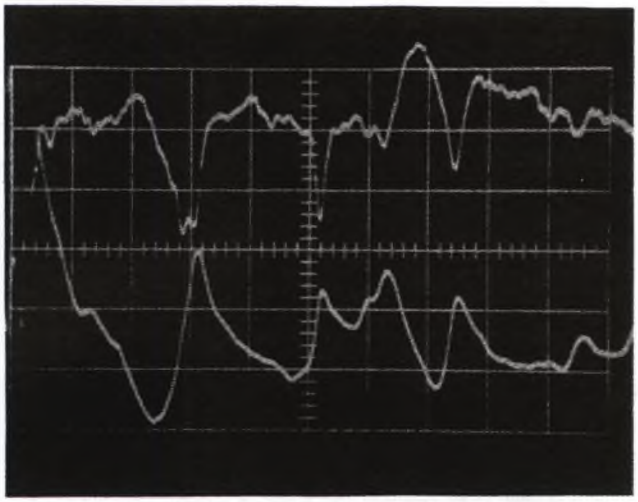


Fig. 6. Amplitude (**upper**) and phase (**lower**) response of 8 in. loudspeaker mounted in a reflex cabinet. Frequency scale is from dc to 2 kHz at 200 Hz per division. Amplitude is 5 dB per division and phase 30° per division.

can be seen to be behind Position 1 by almost 2 in. This is because the phase slope difference between the two regions differs by about 60° per kHz, which yields a time difference of 1/6 msec. The phase behavior between 1.1 kHz and 2 kHz in conjunction with the broad peak indicates a minimum phase resonance and not necessarily a change in location. This is further bolstered by observing that the phase slope from 800 Hz to 1.1 kHz is that of 2 kHz to 5.5 kHz. The high negative phase slope at the cutoff and the phase peak at 500 Hz is the minimum-phase behavior to be expected of the rapid drop in amplitude. The interpretation of the region around 1.5 kHz is that of a multi-pole resonance similar to a band-pass filter.

Figure 5 is the response from dc to 20 kHz for the speaker of Fig. 4. The behavior pattern shows definite non-minimum phase between 12 kHz and 14 kHz, with minimum phase elsewhere.

Figure 6 shows the on-axis behavior from dc to 2 kHz, at 200 Hz per cm, of an 8 in. speaker mounted in a reflex cabinet. The dip at 600 Hz coincides with the undamped back-wall reflection, and from the phase characteristic can be seen to be of the minimum phase type. This oscillograph was made with a five times expansion of a 10 kHz sweep in order that the space-equivalent bandwidth could be made small enough for measurement in a small room. For this reason the smoothing bandwidth is 70 Hz and the phase plot will appear slightly to the left of its proper frequency due to smoothing. By making the appropriate correction it can be seen that each of the peaks and dips is of the minimum phase type shown in Fig. A-2. The strong local fluctuations, such as the dips at 600 Hz, 1100 Hz, and 1500 Hz and the peaks at 400 Hz and 1350 Hz, produce phase fluctuations which are superimposed on the overall phase characteristic due to the low-frequency cutoff. This phase variation due to low-frequency cutoff is characteristically a high negative phase slope in the cutoff region and a smooth extension into the region of normal response.

The low-frequency behavior, which is strong due to the system response zero at dc is shown also in Fig. 7. This is a response measurement at 1 kHz per cm made from −5 kHz to +5 kHz passing through zero. It is a 15° off-axis response of an unenclosed paper whizzer-cone loudspeaker of the type normally used for replacement purposes in automobile radios. The required even and odd frequency characteristic of amplitude and phase is quite pronounced. For this loudspeaker the phase change is 180° at zero frequency, and does not commence its transition until within 200 Hz of zero. The apparent phase breakup below 100 Hz is due to strong low-frequency disturbances in the measuring room which captured the limiter when the loudspeaker signal dropped below their spectral distribution. This oscillograph demonstrates that the phase at zero frequency for which there is no loudspeaker output, may be obtained by centering the time delay spectrometry display at zero frequency and observing the point of symmetry as one approaches zero from both directions. This loudspeaker may be seen to have a minimum phase dip at 3.5 kHz and a possible minimum phase peak at 2.2 kHz, while the remainder of the spectrum is non-minimum phase. This is not unexpected since the response was obtained off-axis and the diffraction and reflection around the whizzer are substantial. Note in particular the absorption dips around 3.5 kHz which do not have substantial phase variations.

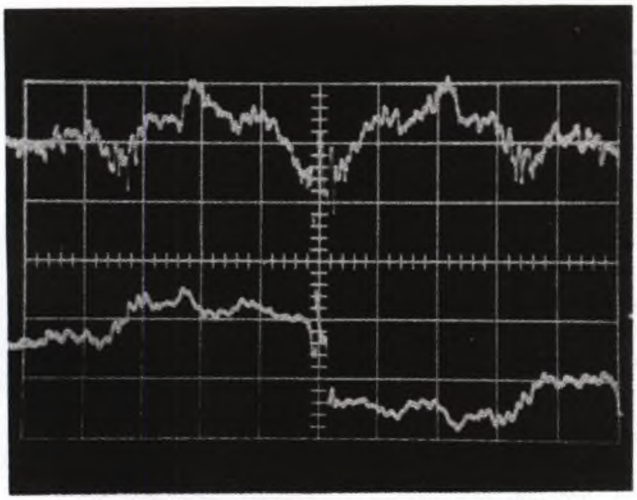


Fig. 7. Amplitude (**upper**) and phase (**lower**) response taken 15° off-axis on an unenclosed whizzer-cone speaker. Frequency scale is 1 kHz per division and extends from −5 kHz through dc to +5 kHz. The even frequency symmetry of amplitude and odd frequency symmetry of phase necessary for causality is readily discerned. Amplitude is 10 dB per division and phase is 120° per division.

Figure 8 shows the response of another horn-loaded compression tweeter. The scale factors for frequency, amplitude and phase are 1 kHz, 10 dB, and 30° per cm respectively. The spectrum encompasses dc to 10 kHz. With the possible exception of the region around 8.4 kHz the response is definitely minimum phase.

Figure 9 shows the response from dc to 20 kHz of a quality midrange electrostatic loudspeaker. This is also of minimum phase characteristic throughout the spectrum. The large apparent phase jump at 18 kHz is a phase detector transition through the equivalent 360° point.

An extremely difficult loudspeaker to measure, due to severe environmental dependence, is the full-range corner horn. The inclusion of the necessary walls and floor rather effectively nullifies a normally anechoic environ-

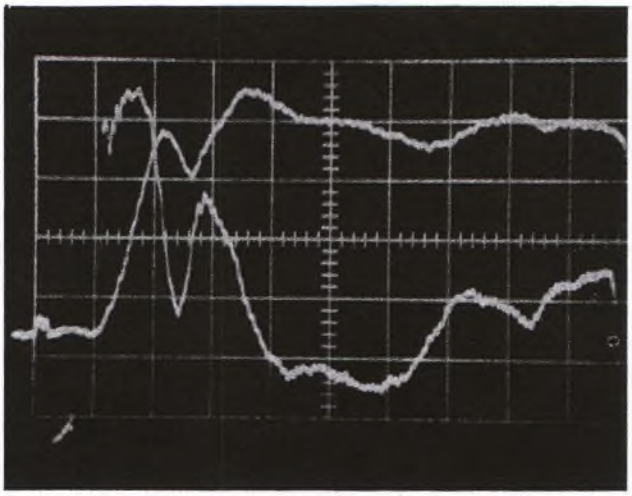


Fig. 8. Amplitude and phase response from dc to 10 kHz at 1 kHz per division of a moderately expensive horn-loaded compression tweeter. Amplitude is 10 dB per division and phase 30° per division.

ment. Figure 10a is for a measurement taken 3 ft on-axis in front of a medium-size corner horn enclosure. The frequency range covered is dc to 1 kHz at 100 Hz per cm. The amplitude scale is 10 dB per cm while the phase scale is 60° per cm. The response dip at 260 Hz was determined to be a genuine loudspeaker aberration and not a chance room reflection by the simple expedient of moving the microphone physically around and noting response. The basic response is seen to be reasonably uniform from about 70 Hz to well beyond 1 kHz with the exception of a strong minimum phase dip at 260 Hz. Figures 4, 8, and 10a compared near cutoff reveal a rather similar behavior pattern. According to the analysis of Appendix A the response dip at 260 Hz is minimum phase and removable. Figure 10b is the result of a simple inductance-capacitance peaking circuit placed between the power amplifier and loudspeaker terminals. The scale factors and magnitudes of Fig. 10b are identical to those of Fig. 10a to show a direct comparison before and after removing the minimum phase response dip. Because the network represents a loss, investigation of Fig. 10b reveals that the overall transfer gain was reduced by about 20 dB with an additional 60° of incurred midrange phase lag. However, the response dip of Fig. 10a estimated at close to 20 dB has been effectively removed. The remaining response dip at 150 Hz may be seen to be an independent response aberration by comparing with Fig. 10a.

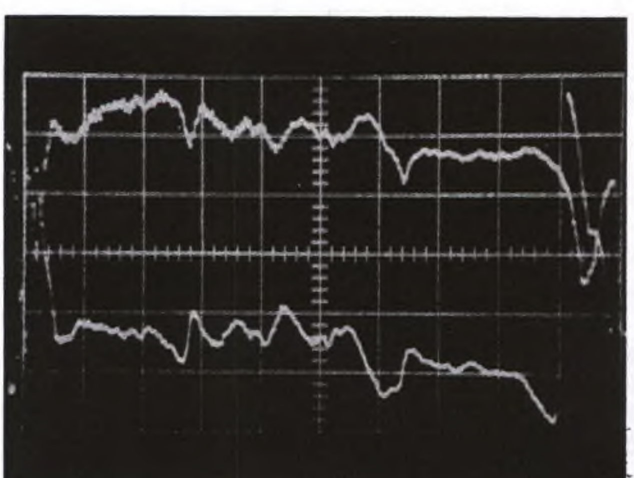


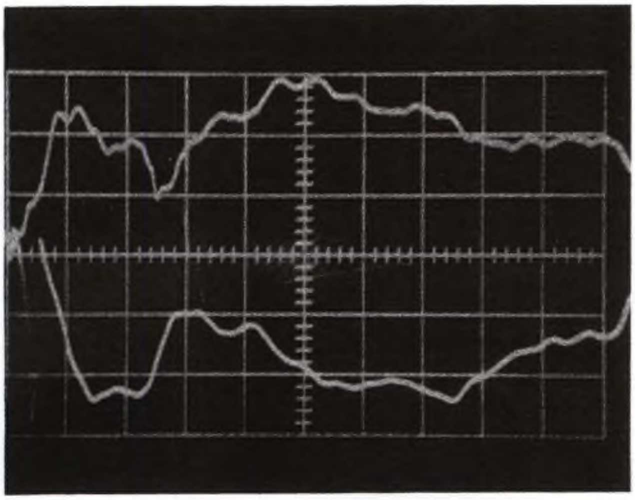
Fig. 9. DC to 20 kHz response at 2 kHz per division of a midrange electrostatic loudspeaker. Amplitude is 10 dB per division and phase is 60° per division.

Investigation of the equalized response of Fig. 10b and other portions not illustrated revealed a smooth amplitude and phase response at frequencies which were harmonically related to 260 Hz. Thus, if room reflections did not constitute substantial energy at the microphone location, the squarewave response with a 260 Hz fundamental should have been improved. Figure 10c is the response to such a squarewave of the configuration of Figs. 10a and 10b. The upper trace is the response after equalization. The lower trace, made as a second exposure with the equalizer removed and system drive reduced accordingly, shows the unequalized response of Fig. 10a. There is no question that the transient behavior is improved. The squarewave frequency of Fig. 10c was chosen to show the improvement due to the response null removed, and was not deliberately modified for a more pleasing waveform after equalization. In fact, a substantial range of squarewave frequencies from about 70 Hz to 300 Hz show a distinct waveform improvement for the equalized speaker even with obvious room resonances.

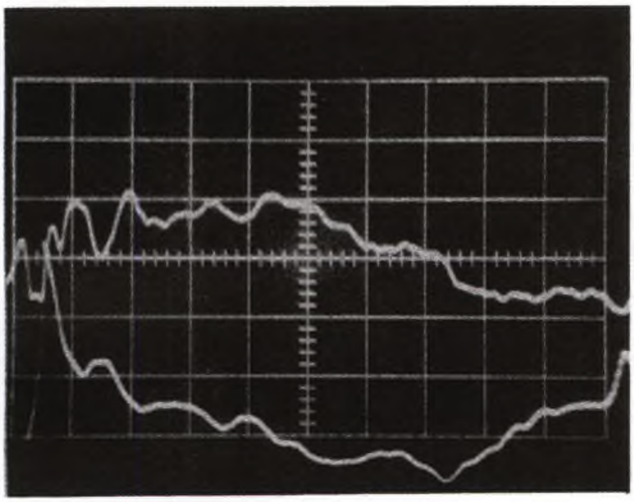
CONCLUSION

This paper represents a preliminary report on loudspeaker frequency response measurement. An attempt has been made to provide a more rigorous approach to understanding the role that the neglected partner, phase, plays in the resultant performance of a loudspeaker. The measurement of phase as a spectral distribution and its correlation with response in the time domain has not received the attention which has been devoted to amplitude. There is, in fact, a substantial void in open literature discussion of phase distributions, which it is hoped will be partially filled by this paper. While there is rather complete agreement about the effect of peaks and dips on the response of a loudspeaker, one can usually expect animated discussion on the subject of phase variations. Part of the reason is the difficulty in instrumenting a phase measurement since phase is intimately related to time of occurrence. Thus measurements on magnetic recorders, disk recorders, loudspeakers and microphones are normally restricted to amplitude characteristics. Where lack of reverberation allows definite measurement of a source, measurements in the time domain by impulse testing or otherwise are used as a supplement to amplitude response measurements in the frequency domain [20]. By utilizing a different measurement technique it has been demonstrated here that loudspeaker phase response measurements may be made with the same facility and validity as amplitude response ones.

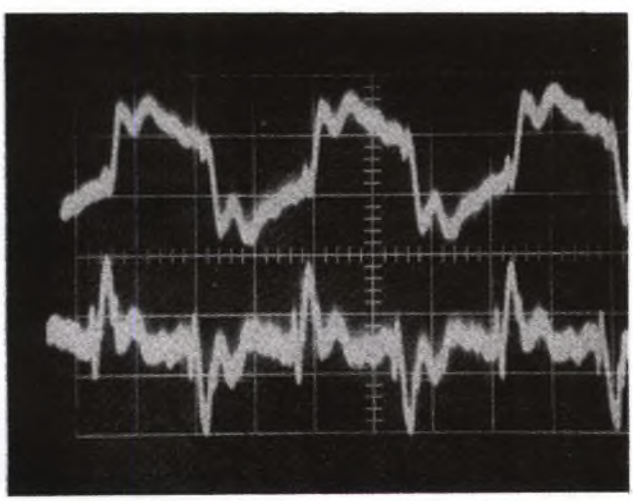
Having thus secured the capability of simultaneous measurement of amplitude and phase spectra, it is necessary to demonstrate a reasonable need for this capability. Looked at another way, any prior analysis which dictated a valid need for phase information would most certainly have precipitated a measurement technique. It



a.



b.



c.

Fig. 10. **a.** Three foot on-axis measurement of a corner horn loudspeaker system. Frequency scale is 100 Hz per division and covers dc to 1 kHz. Amplitude **(upper)** is 10 dB per division and phase **(lower)** is 60° per division. **b.** Result of equalization of the single major response dip of Fig. 10 a. **c. Upper** trace is the acoustic response to a square-wave of the equalized loudspeaker of Fig. 10b. **Lower** trace is the acoustic response of the unequalized loudspeaker of Fig. 10a. Vertical scale is uncalibrated pressure response and horizontal scale is 1 msec per division.

was required, then, to analyze the conditions under which it is necessary to have both amplitude and phase characteristics before one can say one knows everything about the frequency behavior. This led to electronic circuit analysis, where both spectra are normally considered, and a consideration of that class of networks known as minimum phase. The answer was not found completely in circuit analysis, since the very complex behavior of a loudspeaker quite frequently overtaxes the simplified concept of circuit time delay. The appendices to this paper help to bridge the gap in using circuit concepts for analysis of loudspeakers when one has both amplitude and phase spectra. Appendix A derives a simple graphical relationship between amplitude and phase which allows one to state whether a network under analysis is of the minimum phase type. This derivation was necessary since circuit analysis normally proceeds from a knowledge of the equations governing behavior, whereas loudspeaker characteristics are a product of a measurement on a system for which one has incomplete analysis.

Appendix B is a derivation of the inter-domain Fourier transformation relationship in the case of a dispersive medium without absorption. This is, of course, precisely what one finds with a loudspeaker if all minimum phase aberrations are removed. Surprisingly, this is also a common delay situation in many other branches of wave mechanics, yet the transform of a fixed delay of time or frequency (a special case of a non-dispersive medium) is the only form found in widely respected literature.

A head-on attack on the concept of time delay in a dispersive medium with absorption, which is a general characterization of loudspeakers, seems required. The concept of time delay in such a medium does not lack publication; however, such analysis is generally so specific that an engineer quite understandably hesitates in applying it to a general problem. An attack on this problem, discussed in a paper originally planned as a third Appendix to the present one [26], proceeds from the premise that a real-world system is causal; the output follows logically from the input and cannot predict the input. In pursuing this premise we were willing to accept a multiple-valued time behavior of spectral components. As a result, we found that there is indeed a concept of time delay of a system which is causal and makes engineering sense. Furthermore, there is a natural analogy between frequency and time which allows an engineer to draw from knowledge in one domain to add to comprehension in the other domain, and thus to relate modulation theory to time delay incurred by a complex frequency transfer function. Although a unique concept of time delay, the derivation involved is shown to be consistent with the work of Rayleigh [15], Brillouin, [16] and MacColl [9]. An important byproduct of this analysis is an explanation for the confusion created by attempts at using group delay in an absorptive medium. It is shown that the classic concept of group delay is not applicable to a minimum phase medium, and hence to any causal medium with absorption.

Thus, it is possible to identify the effect of amplitude and phase variations as equivalent to what would exist if the actual loudspeaker were replaced by a large number of perfect loudspeakers spread out in space in a frequency-dependent manner. With this approach to the interrelation of amplitude and phase, one can state from measured performance whether a given loudspeaker may be improved by electrical or mechanical means. Since it makes no difference whether the response characteristic is traceable to the loudspeaker, enclosure, or immediate environment, if the resultant behavior is minimum phase one knows that it can be improved.

Finally, some amplitude and phase spectra of typical loudspeakers have been included. The product of any mathematical analysis of phase and time delay would be questionable if it could not be applied to the practical physical problems of an audio engineer. It has been demonstrated by example that a loudspeaker may not be minimum phase. In the case where a loudspeaker is found to be minimum phase, a simple example was included to demonstrate that such minimum-phase response deficiencies can be removed by simple networks.

As more familiarity is gained with the phase response of loudspeakers it is likely that such measurement will become more commonplace. By relating the amplitude and phase characteristics to a seldom considered form of distortion, time delay, and by providing a visual means of determining whether this time delay distortion is removable by relatively simple means, it is hoped that a tool has been provided which will lead to improved quality of sound reproduction.

APPENDIX A

Relations Between Amplitude and Phase for Minimum Phase Network

Assume the transfer function of a network is

$$H(s) = A(s)e^{i\phi(s)} = e^{\alpha(s)+i\phi(s)} \qquad \text{(A1)}$$

where $\alpha(s) = \ln A(s)$ and $s = \sigma + i\omega$. This will be defined as the transfer function of a minimum phase network if $\alpha(s)$ and $\phi(s)$ are uniquely related each to the other. If either $\alpha(s)$ or $\phi(s)$ are known, then everything is known about the network. This implies that there are no zeros or poles of the transfer function in the right-half s plane, since the derivatives must not only exist along the $i\omega$ axis for all values of the frequency ω but are related by the Cauchy-Riemann equations within and on the boundary of the right-half s plane [2,6]:

$$\frac{\partial\alpha(s)}{\partial\sigma} = \frac{\partial\phi(s)}{\partial\omega}, \quad \frac{\partial\alpha(s)}{\partial\omega} = -\frac{\partial\phi(s)}{\partial\sigma}. \qquad \text{(A2)}$$

Taking further derivatives,

$$\frac{\partial}{\partial\sigma}\left(\frac{\partial\alpha(s)}{\partial\sigma}\right) = \frac{\partial}{\partial\sigma}\left(\frac{\partial\phi(s)}{\partial\omega}\right), \qquad \text{(A3)}$$

$$\frac{\partial}{\partial\omega}\left(\frac{\partial\alpha(s)}{\partial\omega}\right) = -\frac{\partial}{\partial\omega}\left(\frac{\partial\phi(s)}{\partial\sigma}\right) \qquad \text{(A4)}$$

$$= -\frac{\partial}{\partial\sigma}\left(\frac{\partial\phi(s)}{\partial\omega}\right),$$

from which it follows that

$$\frac{\partial^2\alpha(s)}{\partial\omega^2} = -\frac{\partial^2\alpha(s)}{\partial\sigma^2} \qquad \text{(A5)}$$

and

$$\frac{\partial^2\phi(s)}{\partial\omega^2} = -\frac{\partial^2\phi(s)}{\partial\sigma^2}. \qquad \text{(A6)}$$

If one assumes that the complex s plane is a topological map with $\alpha(s)$ as the elevation, the situation is as plotted in Fig. A-1 for the simple one-pole one-zero

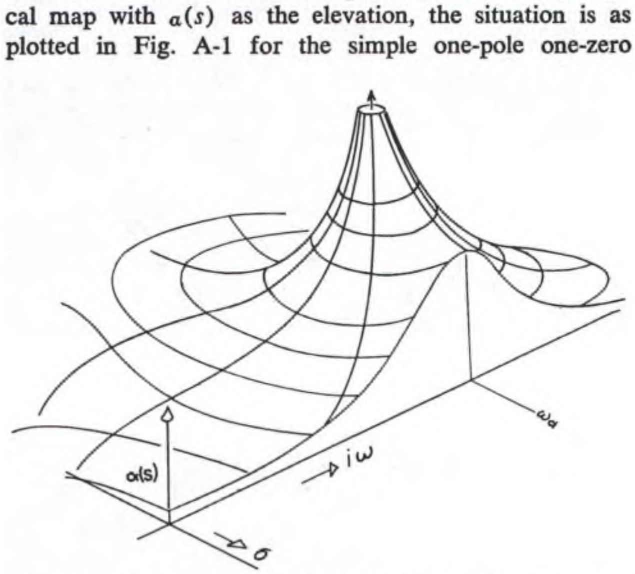


Fig. A-1. Representation of the topological plot of amplitude response in the complex s-plane. A section is taken along the $i\omega$ axis which corresponds to the curve normally considered as the amplitude response. The behavior near a pole in the left half s-plane is characterized and lines of steepest descent and of equipotential are shown.

function. For such a plot of $\alpha(s)$ the lines of steepest descent for $\alpha(s)$ are equipotential lines for $\phi(s)$. The lines of descent for $\alpha(s)$ can originate only from those points at which $\alpha(s)$ is positive or negative infinite, since no finite maxima or minima occur and there can only be saddle points where

$$dH(s)/ds = 0. \qquad \text{(A7)}$$

Because there are no singularities in the right half-plane one can then state that for a point on the imaginary axis (where $\sigma = 0$ and the real-world concept of transfer function exists) when at or near a frequency of closest approach to a singularity such as ω_a in Fig. A-1,

$$\partial\alpha(s)/\partial\sigma \text{ is a maximum negative value.} \qquad \text{(A8)}$$

From Eq. A2, then, at this frequency

$$\partial\phi(s)/\partial\omega \text{ is a maximum negative value,} \qquad \text{(A9)}$$

or, a point of inflection exists for $\phi(s)$,

$$\partial^2\phi(s)/\partial\omega^2 = 0. \qquad \text{(A10)}$$

At this point any penetration into the s plane along the direction of the σ axis must be along an equipotential line of phase and consequently a maximum rate of change of phase with respect to σ, since it follows from Eqs. A6 and A10 that

$$\partial^2\phi(s)/\partial\sigma^2 = 0. \qquad \text{(A11)}$$

This means that

$$-\frac{\partial}{\partial\sigma}\left(\frac{\partial\phi(s)}{\partial\omega}\right) = \frac{\partial^2\alpha(s)}{\partial\omega^2} \qquad \text{(A12)}$$

$$= \text{maximum negative.}$$

This leads to the very simple rule for a minimum phase network: At a local maximum or minimum in the transfer function, a frequency of maximum curvature of amplitude corresponds to a point of inflection of phase.

It may also be seen that the skew symmetry of Eqs. A2 allow a similar rule stating that a maximum curvature of phase corresponds to a point of inflection of amplitude.

Also, from inspecting the polarity of the functions it is evident that if one considers the direction of increasing frequency, when $\partial^2\alpha(\omega)/\partial\omega^2$ is a maximum positive then $\partial\phi(\omega)/\partial\omega$ is a maximum positive, while when $\partial^2\phi(\omega)/\partial\omega^2$ is a maximum positive then $\partial\alpha(\omega)/\partial\omega$ is a maximum negative.

Note that these relationships should be held precisely only if the frequency under analysis is that which is closest to the singularity (pole or zero in transfer function) which produces the high rate of change of curvature. Perturbations on either side of the frequency of the singularity will adhere less to this pattern the further one proceeds from the singularity. Figure A-2 shows several simple cases of amplitude and phase functions which one can use to identify minimum-phase networks. Points of maximum curvature are shown by the dashed circle around which the curve might be thought of as bent. Points of inflection are shown by horizontal dashes. Note that the frequency scale is linear, amplitude is plotted logarithmically in dB with increasing gain as a positive quantity, and phase is plotted as an angle with the standard convention of phase lag as a negative angle.

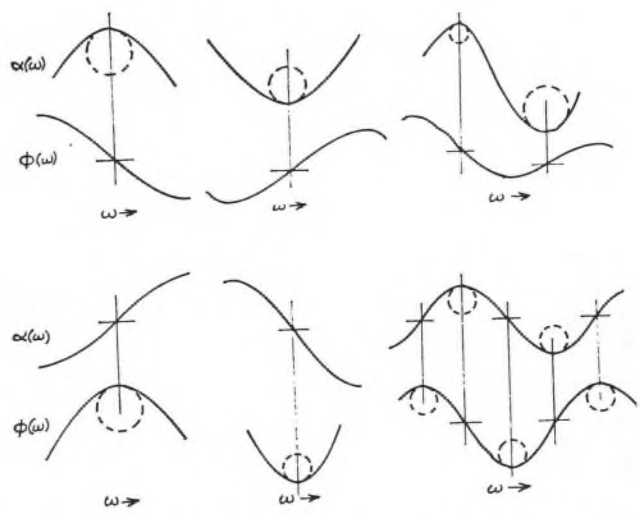


Fig. A-2. Characteristic amplitude and phase plots which may be used to identify minimum phase behavior. Points of maximum curvature are shown by **circles** around which the curve may be considered as bent, and associate points of inflection indicated by **horizontal lines.**

APPENDIX B

Fourier Transform of Frequency-Dependent Delay

A medium in which a time-dependent disturbance propagates is said to be dispersive if the velocity in the medium, or time of traverse, varies with the frequency of the time dependence. If there are many frequency constituents in the initial form of the disturbance, then some time later these constituents are dispersed. The observed waveform at a fixed "down-stream" location will no longer be identical to the initial disturbance, and the resultant waveform will be a distortion of the original waveform even if no reduction in amplitude of these constituents has occurred.

Assume that the time delay for each frequency ω has been determined to be

$$T(\omega) = T_0 + \tau(\omega) \tag{B1}$$

where T_0 is a fixed delay in seconds and $\tau(\omega)$ is a dispersive time delay. Since we are interested in frequency dependent factors, the frequency phase dependence becomes

$$\Theta(\omega) = \int_0^\omega T(\omega)\,d\omega = T_0\omega + \int_0^\omega \tau(\omega)\,d\omega. \tag{B2}$$

This is the phase dependence of each frequency component comprising the dispersed output waveform. The expansion of all such components in a Fourier integral will yield the resultant time function at the output. Thus, for the case of an otherwise perfect transmission system, an input signal with a time dependence $g(t)$ and a frequency transform

$$G(\omega) = \int_{-\infty}^{\infty} g(t)e^{-i\omega t}dt \tag{B3}$$

will produce an output from the dispersive medium of $f(t)$, where

$$f(t) = (1/2\pi)\int_{-\infty}^{\infty} G(\omega)e^{-i\Theta(\omega)}e^{i\omega t}d\omega. \tag{B4}$$

The negative sign is used for time delay. This leads to the important finding that:

If $G(\omega)$ is the transform of $g(t)$, then

$$G(\omega)\cdot e^{\pm i\int_0^\omega T(\omega)d\omega}$$

is the transform of

$$g[t \pm T(\omega)].$$

By the same reasoning, if $g(t)$ is the transform of $G(\omega)$ then

$$g(t)\cdot e^{\pm i\int_0^t \Omega(t)dt}$$

is the transform of

$$G[\omega \mp \Omega(t)].$$

Note that the commonly encountered transform relations for a fixed delay constitute the special case of non-dispersive time delay.

REFERENCES

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

2. C. H. Page, *Physical Mathematics* (D. Van Nostrand Co., Princeton, 1955), p. 220.

3. S. J. Mason and H. J. Zimmerman, *Electronic Circuits, Signals, and Systems* (Wiley, New York, 1960), p. 350.

4. H. W. Bode, *Network Analysis and Feedback Amplifier Design* (D. Van Nostrand Co., New York, 1955), p. 312.

5. Bateman Manuscript Project, *Tables of Integral Transforms* (McGraw Hill Book Co., New York, 1954), p. 239.

6. E. A. Guillemin, *The Mathematics of Circuit Analysis* (M.I.T. Press, Cambridge, 1965); p. 330.
7. W. R. Bennett and J. R. Davey, *Data Transmission* (McGraw Hill Book Co., New York, 1965), p. 84.
8. R. V. L. Hartley, "Steady State Delay as Related to Aperiodic Signals," *B.S.T.J.* **20**, 222 (1941).
9. S. Goldman, *Frequency Analysis Modulation and Noise* (McGraw Hill Book Co., New York, 1948), p. 102.
10. G. A. Campbell and R. M. Foster, *Fourier Integrals* (D. Van Nostrand Co., Princeton, 1961), p. 7.
11. M. J. Lighthill, "Group Velocity," *J. Inst. Math. Applics. I* (Academic Press, London, 1965), p. 1.
12. S. P. Mead, "Phase Distortion and Phase Distortion Correction," *B.S.T.J.* **7**, 195 (1928).
13. P. R. Geffee, *Simplified Modern Filter Design* (J. F. Rider, New York, 1963), p. 73.
14. S. Hellerstein, "Synthesis of All-Pass Delay Equalizers," *IRE Trans. on Circuit Theory* **PGCT-8**, No. 3, 215 (1961).
15. Rayleigh, *Theory of Sound* (Dover Press, London, 1945), Vol. 1, p. 301.
16. L. Brillouin, *Wave Propagation and Group Velocity* (Academic Press, London, 1960).
17. H. F. Olson, *Acoustical Engineering* (D. Van Nostrand Co., Princeton, 1960).
18. A. Papoulis, *The Fourier Integral and Its Applications* (McGraw-Hill Book Co., New York, 1962), p. 148.
19. D. Slepian and H. O. Pollak, "Prolate Spheroidal Wave Functions, Fourier Analysis and Uncertainty," *B.S.T.J.* **40**, 353 (1961).
20. E. D. Sunde, "Pulse Transmission by A. M., F. M., and P.M. In The Presence of Phase Distortion," *B.S.T.J* **40**, 353 (1961).
21. G. N. Watson, "The Limits of Applicability of the Principle of Stationary Phase," *Proc. Cambridge Phil. Soc.* **19**, 49 (1918).
22. C. Eckart, "The Approximate Solution of One-Dimensional Wave Equations," *Rev. Mod. Phys.* **20**, 399 (1948).
23. C. A. Ewaskio and O. K. Mawardi, "Electroacoustic Phase Shift in Loudspeakers," *J. Acoust. Soc. Am.* **22**, 444 (1950).
24. F. M. Weiner, "Phase Distortion in Electroacoustic Systems," *J. Acoust. Soc. Am.* **13**, 115 (1941).
25. F. M. Weiner, "Phase Characteristics of Condenser Microphones," *J. Acoust. Soc. Am.* **20**, L707 (1948).
26. R. C. Heyser, "Group Delay, Excess Delay, and Overall Time Delay," to be published.

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. 17, no. 2, pp. 130–137, 1969 April.

Loudspeaker Phase Characteristics and Time Delay Distortion: Part 2

RICHARD C. HEYSER

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

Fourier Integral concepts are explored for the relation existing between a function in the frequency domain and its time domain counterpart. A derivation is obtained for the effect of a loudspeaker's imperfect frequency response as a specific type of time delay distortion of the reproduced signal.

INTRODUCTION In an earlier paper [1] the definition of loudspeaker frequency response was expanded to include the phase of the pressure wave produced by an electrical stimulus as well as the conventionally measured amplitude. A technique of measurement was introduced which allowed a measurement to be made of this more complete response, and some examples were included of the response of common types of loudspeaker. Since the proper role of a loudspeaker is the acoustic reproduction of a time-dependent signal, the measurement of even the more complete frequency response is academic unless some inference can be obtained from this measurement as to whether the loudspeaker does its job well. Accordingly a presentation without proof was made of a means of visualizing the effect of imperfect loudspeaker frequency response as producing a time delay distortion equivalent to a frequency-dependent spatial distribution of otherwise perfect loudspeakers. It is the purpose of the present work to investigate the determination of temporal response from the more complete frequency response and develop this acoustic model.

In considering time response it must be remembered that engineers work in a causal world where cause distinctly precedes effect and time advances in its own inexorable fashion. No analysis performed on a network as complicated as a loudspeaker may be considered valid if it violates causality and allows the clock to run backward. Because of considerable mathematical complexity, the subject of time delay in a dispersive medium with absorption is generally avoided in most written material. The reader of such material is left instead with some simplified relations using the frequency phase spectrum, which for most systems yield time delay answers close to observed behavior. Those systems for which the answer violates a prior physical premise are considered anomalous. When all that is available on the frequency response of a loudspeaker is the pressure amplitude spectrum one cannot utilize the simplified temporal relations, and hence no questions arise. With the introduction of a means for measuring the complete frequency response one runs into immediate difficulty with application of the simplified concepts of time behavior because in many cases it is found that causality cannot be maintained.

In order to understand the distortion which a loudspeaker may impart to a time-dependent signal because of its imperfect frequency response, it becomes necessary to look more closely at the concept of frequency-dependent time delay and generate revisions required to present an understandable acoustic equivalent for an actual loudspeaker. This paper proceeds by first demonstrating why the common concept of group delay is not applicable to minimum-phase systems with absorption. Then a substitute for group delay is developed and is shown to provide the proper solution for some systems commonly considered to have anomalous behavior. Finally, a network concept is introduced which leads to an appropriate acoustic model for a loudspeaker.

GROUP DELAY, EXCESS DELAY, AND OVERALL TIME DELAY

Historically the concept of time delay in a dispersive medium was recognized as early as 1839 by Hamilton, but the distinction between phase delay and group delay seems to have been put on a firm foundation by Lord Rayleigh in publications in 1877 [2]. Rayleigh considered that group velocity represented the actual velocity of propagation of groups of energy in a medium. Group delay is defined to be the time delay in traversing a fixed distance at this group velocity [3]. To understand group delay one need only consider that the transformation from the analysis of a problem in the frequency domain to the solution in the time domain involves a Fourier Integral of the form

$$f(t) = (1/2\pi) \int_{-\infty}^{\infty} G(\omega)\, e^{a(\omega)}\, e^{i[\omega t + \phi(\omega)]}\, d\omega \qquad (1)$$

This may fall into a class of integral equations of the type

$$f(t) = \int F(s) e^{tg(s)} ds \qquad (2)$$

where $s = \sigma + i\omega$, $g(s) = x + iy$ is an analytic function, t is large, positive, and real, and $F(s)$ varies slowly compared with the exponential factor [4, 5].

Lord Kelvin's method of stationary phase evaluates integrals of this type by deforming the path of integration where possible through saddle points where x is constant and

$$\partial x/\partial\sigma = \partial x/\partial\omega = 0 \text{ and } \partial y/\partial\sigma = \partial y/\partial\omega = 0 \qquad (3)$$

For this path the modulus of $\exp[t \cdot g(s)]$ is constant while the phase varies. When all of these conditions are met, not only may an asymptotic solution be achieved but what is more important, Eq. (3) shows that the major contribution to the integral takes place where the phase is stationary and

$$dg(\omega)/d\omega = 0. \qquad (4)$$

When these conditions are applicable the major contribution to the solution of Eq. (1) occurs at a time t such that

$$t = -d\phi(\omega)/d\omega. \qquad (5)$$

Since time commences in the analysis at initiation of input stimulus, this means that the time delay of the signal through the network is this value of t, called group delay.

Since the principle of stationary phase is a commonly used derivation of the network theory concept of group delay, it is of utmost importance to note the restrictions on the use of this derivation. The most important restriction is that the modulus remain a slowly varying function of frequency in that region of the frequency domain where the phase is changing the least. This means that when working with a network element this condition may be met by solutions which involve very long time delays, such as transmission lines, or when applied to networks that have no amplitude variation with frequency, such as all-pass lattices where $a(\omega)$ in Eq. (1) is always a constant.

In setting up relations for a network with absorption and short overall time delays, one gets an equation deceptively similar to Eq. (2) but with a substantial real as well as imaginary term in the exponent. The time function of Eq. (1) is an inversion integral evaluated along a path which is the entire imaginary axis from $-\infty$ through the origin to $+\infty$, closed to the left with a semicircle of infinite radius and the origin as center. This is done so as to encircle all singularities of the integrand for time greater than zero. The path of integration is restricted to the $i\omega$ axis when the expression of Eq. (1) is used and the real and imaginary parts of the exponent are related by the Cauchy-Riemann differential relations. Thus, even if $a(\omega)$ is generally a slowly varying function, just at that point on the $i\omega$ axis where the phase is stationary, $a(\omega)$ varies rapidly and one *cannot* use the principle of stationary phase. If the time delay of the network is small relative to several periods of the frequency under analysis, which is a condition commonly found in loudspeakers, then this inapplicability of stationary phase can lead to solutions for time delay which are absurd. Consider for example the circuit of Fig. 1. This network is certainly well behaved, yet the group delay is negative from zero frequency to the geometric mean of the transfer-function break points. Since obviously the output cannot predict the input, the only logical solution would be that the time delay of this network is not represented by group delay. There will of course exist a proper solution for time delay, but this requires a careful evaluation of the inversion integral through the saddle points of Eq. (3) where one may either use Kelvin's method of stationary phase with a path through the saddle points with x constant, or the method of steepest descent which chooses a path of integration so as to concentrate the large values of x in the shortest possible interval with y constant. The two methods are nearly equivalent, since the paths cross the same saddle points and can be deformed one into the other provided contributions from any singularities crossed are taken into account.

The minimum phase transfer function is the function with the minimum accumulation of phase lag (negative phase shift) as ω proceeds from dc to infinity. Because of this the accumulation of phase lag in a minimum phase

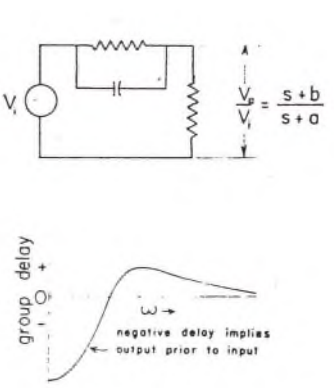


Fig. 1. A simple minimum phase circuit and its group delay, illustrating the inapplicability of group delay to such a causal circuit.

network may be negative as well as positive, as in the case of the circuit of Fig. 1. Since the Cauchy-Riemann equations actually define a minimum phase network and are necessary conditions that a circuit cannot predict the occurrence of a signal, the behavior of phase accumulation means that the group delay of Eq. (5) does *not* provide a measure of time lag in a minimum phase network.

While this might appear to destroy the concept of group delay for minimum phase networks, consider now the special non-minimum phase network called the allpass or flat network, with a constant amplitude of response [6, 7, 8]. For this network there is accumulated phase lag at a rate which is not negative at any frequency, yielding a group delay which is never negative. For this network, since $a(\omega)$ is constant, the principle of stationary phase is valid on the imaginary axis. The time delay thus calculated according to Eq. (5) is everywhere meaningful; this time delay of an allpass network with the transfer function

$$H(\omega) = e^{-i\theta(\omega)} \tag{6}$$

will be defined as excess delay

$$t_{\text{excess}} = d\theta(\omega)/d\omega. \tag{7}$$

One must be careful to observe that the excess delay is the time elapsed from the injection of a transient to the major contribution of the output waveshape. There may be minor ripples, or forerunners to use a phrase of Brillouin [9], which precede this major change as well as the latecomers which provide the effect commonly called ringing, but nonetheless the major change will occur at the time which was called excess delay.

If a network is minimum phase, there exists a unique relationship between amplitude and phase which allows a complete determination of phase from amplitude. If a network is non-minimum phase with a transfer function $H(\omega)$, there will exist a unique minimum phase network $G(\omega)$ with the same amplitude response, and an allpass network with a phase response $\theta(\omega)$ in cascade such that [10]

$$H(\omega) = G(\omega)e^{-i\theta(\omega)} = A(\omega)e^{-i\phi(\omega)}e^{-i\theta(\omega)}. \tag{8}$$

If the time delay characteristics of minimum phase networks and allpass lattices are considered, one can reconstruct the time behavior of any arbitrary physically realizable network. There will exist some total time delay of the network $H(\omega)$ which will be called t_{overall}. There will also exist some time delay for the minimum phase network $G(\omega)$ which will be called $t_{\text{min. phase}}$. The relation between these delays is

$$\begin{aligned} t_{\text{overall}} &= t_{\text{min. phase}} + t_{\text{excess}} \\ &= t_{\text{min. phase}} + [\partial\theta(\omega)/\partial\omega]. \end{aligned} \tag{9}$$

The commonly used group delay is the frequency slope of the total measured phase of $H(\omega)$, or from Eq. (5)

$$t_{\text{group}} = [\partial\phi(\omega)/\partial\omega] + [\partial\theta(\omega)/\partial\omega] \tag{10}$$

which may be expressed as

$$t_{\text{group}} = t_{\text{overall}} + \{[\partial\phi(\omega)/\partial\omega] - t_{\text{min. phase}}\}. \tag{11}$$

Consequently the group delay will be quite close to the overall time delay of the network if

$$t_{\text{overall}} >> \{[\partial\phi(\omega)/\partial\omega] - t_{\text{min. phase}}\}. \tag{12}$$

This is another verification that if a network has a sufficiently large overall time delay, then group delay may be considered a satisfactory substitute provided that the group delay of the equivalent minimum phase network is reasonably well behaved.

TIME DELAY AS A DISTRIBUTION

Turn now to a consideration of time delay in a general network. Attempts at a direct derivation of time delay do not seem particularly fruitful, since the classic definition requires that one make a sudden change in some parameter and see how long it takes before this change appears in the output; however, the moment a discontinuity is created in a time derivative of an electrical parameter, one no longer has that parameter but a large set of sideband frequencies which interfere with the measurement. Thus one is led to look for another solution which involves the relationship existing between frequency and time.

The relationship existing between a function in the time domain $f(t)$ and the same function in the frequency domain $F(\omega)$, is given by the Fourier integrals

$$f(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} F(\omega)e^{i\omega t}d\omega \tag{13}$$

and

$$F(\omega) = \int_{-\infty}^{\infty} f(t)e^{-i\omega t}dt. \tag{14}$$

There is an obvious symmetry which analytically lets a function in time commute with a function in frequency. Indeed, if a function were given in a dummy parameter and one did not know whether it was of time or frequency, there would be no way of ascertaining the proper domain. A remarkable fact would arise if one blindly inserted this function into the wrong equation: if the function were as well behaved as any related to a real world containing dissipation, the answer would be correct in form. This is because functions may be transferred in the Fourier integral if the sign of one of the parameters is reversed [11]. The implications of this are enormous, as many facts laboriously proven in one domain may automatically be transferred to the other domain. For example, as pointed out in a previous paper [12], if one terminates a time series there exists a frequency overshoot analogous to Gibbs' phenomenon.

The commutation of parameters, then, gives the remarkable simplification that the analysis of a distribution in time due to a complex transfer function is isomorphic with the frequency distribution due to complex modulation in time. This isomorphism considerably frees our imagination when trying to cope with the concept of the time delay of a frequency. If one imagines that the variation of amplitude with frequency of a frequency transfer function is analogous to the variation of amplitude with time of a time transfer function, one can imagine that there are "time sidebands" analogous to the frequency sidebands of modulation theory. In the case of frequency, all values from $-\infty$ through zero to $+\infty$ are allowed and we conveniently identify negative frequency as a phase reversal of a positive frequency. For the parameter time it is conventional to start analysis for a value of zero and assume no activity prior to this.

This merely requires that the time function have an even and odd component which cancel each other for all times less than zero. For such a time function which does not allow prediction, this means that the frequency function similarly has an even (amplitude) and odd (phase) component, although these will not necessarily cancel out at any frequency. This physical realizability criterion also means that the frequency transfer function must have complex conjugate poles and zeros in order to satisfy the even-odd requirement.

The concept of time delay of a frequency component is not complete, since the functions discussed so far are voltages in terms of either frequency or time. Consider, however, a frequency function which has a distribution that is forming as we observe in real time. This is called the running transform $F_t(\omega)$ [13, 14]

$$F_t(\omega) = \int_{-\infty}^{t} f(t)e^{-i\omega t}dt. \qquad (15)$$

In this case there is a distinct relation between the distribution of sideband energy and time. There will exist a spectral distribution of frequencies corresponding to an instant in time which may be single-valued, multiple-valued, or a continuous distribution. By interchanging time for frequency one may infer that the time delay of a network for a given frequency may also be a distribution. This goes a long way toward clarifying the confusion created by investigators who attempt to come up with a single-valued number for the delay of a network. In those regions in which the actual delay distribution is small or single-valued, the simple group delay scores very well, but in regions of moderate to large dispersion group delay falls down completely and even yields absurdities.

By observing the conjugate behavior of time and frequency it should be apparent to anyone familiar with modulation theory that a network frequency transfer function

$$F(\omega) = A(\omega)e^{-i\phi(\omega)} = e^{a(\omega)}e^{-i\phi(\omega)} \qquad (16)$$

represents a distribution of time delayed functions around the value

$$t_{\text{group}} = [d\phi(\omega)/d\omega]. \qquad (17)$$

Furthermore, the group delay will represent the absolute delay of each component only if

$$a(\omega) = \text{constant}. \qquad (18)$$

The distribution around the group delay in Eq. (17) is certainly consonant with the paired echo concept of Wheeler and MacColl [15] which treats the effect of minor deviations from the ideal transfer function by expanding the time function around these deviations.

DELAY IN MINIMUM PHASE NETWORK

Having recognized that the true network time delay of Eq. (19) may not necessarily be single-valued and may even be a finite distribution, we turn out attention to deriving the form of a minimum phase time $t_{\text{min. phase}}$ for several simple expressions.

As shown earlier, the group delay of a network with constant gain is the proper delay. Consider the single

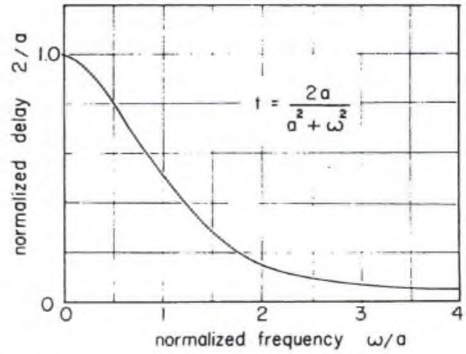


Fig. 2. Normalized plot of excess delay for a first-order allpass lattice.

pole allpass lattice function

$$L(\omega) = (s-a)/(s+a) = (i\omega - a)/(i\omega + a) \text{ for } \sigma = 0. \qquad (19)$$

This is a constant gain function with a group delay

$$t_{\text{group}} = 2a/(a^2+\omega^2). \qquad (20)$$

This is shown in Fig. 2. The time delay is maximum at zero frequency, and there is no delay at infinite frequency. There is also the very useful fact that single pole functions can be expressed as combinations of this lattice, for example,

$$\frac{1}{s+a} = \frac{1}{2a}\left(1 - \frac{s-a}{s+a}\right) \qquad (21)$$

and

$$\frac{s+b}{s+a} = \frac{1}{2a}\left[(a+b) - (b-a)\frac{s-a}{s+a}\right]. \qquad (22)$$

Equation (21) is that of a simple lowpass filter, and Eq. (22) describes the circuit of Fig. 1 if a is greater than b. The lefthand side of these equations is the commonly encountered system transfer function $H(s)$, consisting of a frequency-dependent amplitude and phase function. The system transfer function is the frequency transform of the time response to an impulse of voltage $h(t)$; thus,

$$H(s) = \int_{-\infty}^{\infty} h(t)e^{-st}dt. \qquad (23)$$

Normally we think of the system transfer function as the quotient of output to input signal and use this concept to generate the common form expressed by the lefthand side by using a sinewave signal. This concept, however, is only valid if a sinewave is used, since there must in general be a time delay in a network; since Eq. (23) does not contain an explicit time dependence it is apparent that this time discrepancy is absorbed in the complex frequency spectrum and thus locked up so that we cannot readily predict time behavior without mathematical manipulation. The righthand sides of Eqs. (21) and (22) show alternate forms of the system transfer function, obtained purely from a special class of transfer functions which represents a known frequency-dependent time delay without a frequency-dependent amplitude. (Using this form allowed us to unlock the time behavior.)

Examining Eq. (21) it is apparent that the simple lowpass filter can be considered to consist of two parallel constant-amplitude delay functions, one with no delay and the other the delay of Eq. (20). At very high frequencies these two delay signals cancel each other since they arrive at the same time and are of opposing polarity, while at low frequencies there is not a simultaneous output and hence no complete cancellation. A similar interpretation can be placed on Eq. (22).

Thus, the search for a meaningful concept of time delay in a circuit has revealed that there are simple allpass functions which possess a frequency-dependent time delay that fits out intuitive concept of delay; furthermore, a simple minimum-phase network for which the concept of group delay is invalid is now seen to be represented as a combination of allpass delay functions. Figure 3 shows the minimum phase time delay and group

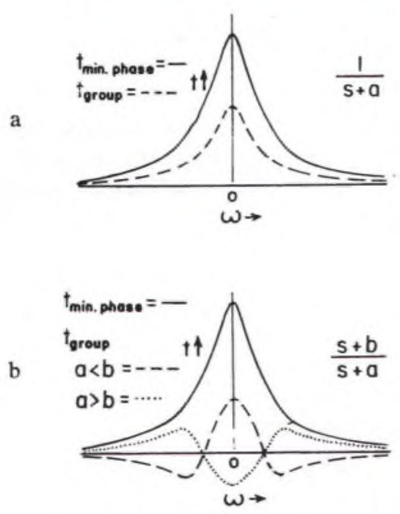


Fig. 3. Minimum phase delay and group delay. The actual minimum phase delay is double-valued and composed of a straight-line zero delay and a bell-shaped delay of the form of Fig. 2. The group delay is single-valued. **a.** Single-pole lowpass circuit. **b.** Single-zero single-pole transfer function.

delay for the single-pole functions of Eqs. (21) and (22). The minimum phase delay is seen to be double-valued for these single-pole functions. The strength of these delayed signals is obtained from the coefficients of Eqs. (21) and (22). It is immediately apparent that group delay is quite misleading for the function of Fig. 1, since this goes to negative time over a substantial portion of the frequency spectrum. The actual delay, as can be seen, never goes negative. Similarly, the group delay of Fig. 3a, although never negative, is nonetheless improper.

A NETWORK CONCEPT

The single-pole single-zero allpass lattice function of Eq. (19) is a primitive function which can be used as a building block for more complicated delays. Two lattices in cascade may, like relations (21) and (22), be composed of combinations of the constituent lattices; for example, if $a \neq b$,

$$\frac{s-a}{s+a}\cdot\frac{s-b}{s+b}=1+\frac{a+b}{a-b}\cdot\frac{s-a}{s+a}-\frac{a+b}{a-b}\cdot\frac{s-b}{s+b} \quad (24)$$

Similarly, one can expand other products of lattices as linear combinations of the individual lattices.

At first glance this would appear to invalidate the conclusion that the time delay of any allpass network is the frequency derivative of the phase function, as the latter is single-valued whereas Eq. (24) shows an expansion which is definitely multiple-valued. Reconciliation may be obtained by remembering that the principle of stationary phase yields the time at which the largest contribution will occur for the integral in Eq. (1). This time will be that of Eq. (7). We might expect that there will be prior contributions and these are discerned in the expansion on the right hand side of Eq. (24). If a sufficiently complicated network of such allpass functions were generated and an oscilloscope used to view the network output with a sudden input transient, the output waveform would be observed to have forerunners preceding the main signal transition. The only condition under which no forerunners would be observed is when the individual lattice sections are identical, in which case there can be no expansion such as Eq. (24). In other words, there is no linear combination for an iterated lattice,

$$[(s-a)/(s+a)]^N \quad (25)$$

and in this case the delay of Eq. (7) is the only delay. In this special case, if the frequency parameter a is very high, approaching infinity as rapidly as the number of identical sections n, then in the limit as n becomes large without limit this relation becomes the transfer function of ideal delay, [10]

$$e^{-T_0 s}=e^{-iT_0\omega} \text{ for } \sigma=0. \quad (26)$$

For all other iterated lattices the delay distribution will be a summation of the constituent delays and in the limit for such a dispersive network will be an integral expression (derived in an earlier paper [1]). The magnitude of terms on the righthand side of Eq. (24) and any such expansion is such that no single term contributes appreciably to the resultant output prior to the time indicated by Eq. (7). Instead each term is effectively nullified by a term representing a prior or later delay, and nullification is not substantially removed until the time of Eq. (7).

From the preceding discussion of forerunners it is quite easy to see how it is possible for a network with the transfer function and time delay of Fig. 3b to be cascaded with a complementing network to produce a constant-gain zero-delay output; thus,

$$(s+b)/(s+a)\cdot(s+a)/(s+b)=1. \quad (27)$$

While there is a finite delay component in Eq. (22), there is no necessity to envision a negative time delay to cancel the term of the form (24) which occurs in the cascaded combination, since each and every forerunner except a unity-gain zero-delay forerunner is cancelled completely. Some remarkable facts may now be deduced from the preceeding observations about network transfer functions which have all poles and zeros on the real σ axis.

1. Any network with simple poles and zeros restricted to the real σ axis may be considered as equivalent to a *parallel* combination of first-order allpass lattices. There will be one allpass lattice for each pole of the network transfer function. The pole, and hence time delay distribution, of each lattice will be determined by the asso-

ciate transfer function pole, while the strength and polarity of each lattice will be determined by the joint distribution of zeros and poles.

2. Higher order poles in the transfer function will yield series combinations of the associate lattices, with the number of lattice sections determined by the order of the pole.

3. Series combinations of networks may be considered as parallel combinations of the constituent lattices of each network.

Because of the associative property of the Fourier transform, the foregoing conclusions concerning the distribution of equivalent networks mean that since each lattice has a frequency-dependent time delay, the time delay of the network output is not single-valued but a multiple-valued combination of the primitive delays. Fig. 2 is a time-delay frequency distribution for the simple one-pole function. Any other minimum phase network which can be expressed as a rational function factorable to the form

$$\frac{(s+a)(s+b)\ldots}{(s+\alpha)(s+\beta)\ldots} \tag{28}$$

will have a time delay frequency function expressible as a sum of delays of the form of Eq. (20) and will have a graphical plot of delay *vs* frequency such as Fig. 4. To

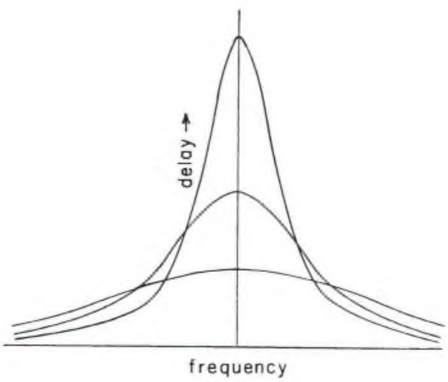


Fig. 4. The multivalued delays to be anticipated for a transfer function with a multiplicity of simple poles at the same frequency.

consider the time delay behavior of such a network, we may thus draw the equivalent network of Fig. 5, where each lattice is considered a frequency-dependent delay line with the delay of Eq. (20). A zero delay may be assumed due to a lattice with a pole at zero frequency. The gain and polarity of each delay line channel is assumed to be determined by a summing amplifier, for the sake of illustration only.

When dealing with a physical process which involves propagation with a frequency-independent velocity, such as sound in air, an equivalent interpretation of Fig. 5 would be that there is a distribution of otherwise perfect sources which assume a frequency-dependent position in space such that the delay due to the additional distance travelled at the velocity of propagation is identical to that of the equivalent delay line.

The allpass lattice of Eq. (19) has a single-pole and single-zero configuration on the real axis. This, as was seen, is quite satisfactory for discussing the time delay of

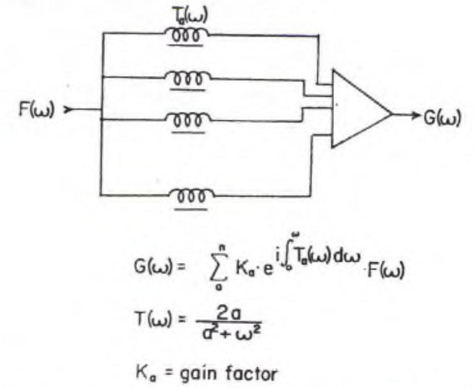


Fig. 5. Symbolic representation of a network with a transfer function expressible as a rational product of terms with poles and zeros. This network may be interpreted as a parallel combination of delay lines with constant amplitude transfer function but a frequency-dependent delay as shown. An input signal with spectral distribution $F(\omega)$ will produce the output $G(\omega)$.

any minimum-phase network with poles on the real axis, i.e., with the terms of Eq. (28) which do not have an imaginary component. A loudspeaker, however, generally has poles with an imaginary component, which leads to peaks and dips in the frequency response and damped ringing in the time response. For this case there exists one type of allpass lattice which, like Eq. (19) on the real axis, can be used to represent the time delay of any network with imaginary poles. This is the second-order lattice with conjugate complex poles and zeros and with the transfer function

$$\frac{(s-a+ib)}{(s+a+ib)} \cdot \frac{(s-a-ib)}{(s+a-ib)}. \tag{29}$$

There does not exist a simple one-parameter delay such as represented by Eq. (20); instead the delay relation now depends upon the position of a and b. The form of delay may be ascertained by allowing the expansion of Eq. (29) to be considered as two cascaded sections of the type of Eq. (19) with appropriate shift in complex frequency. Since the transfer function is now a sum of phase shifts, the time delay from Eq. (20), is [7]

$$t = \frac{2a}{a^2+(\omega-b)^2} + \frac{2a}{a^2+(\omega+b)^2}. \tag{30}$$

Obviously, if the term b approaches zero this becomes the transfer function of Eq. (25) with $n = 2$, so that the delay becomes twice that of Eq. (20). On the other hand, if for a given value of b the term a approaches zero, the phase shift in the vicinity of the frequency of b

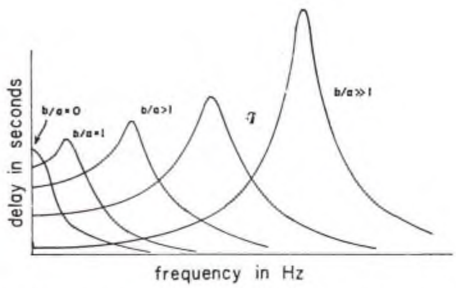
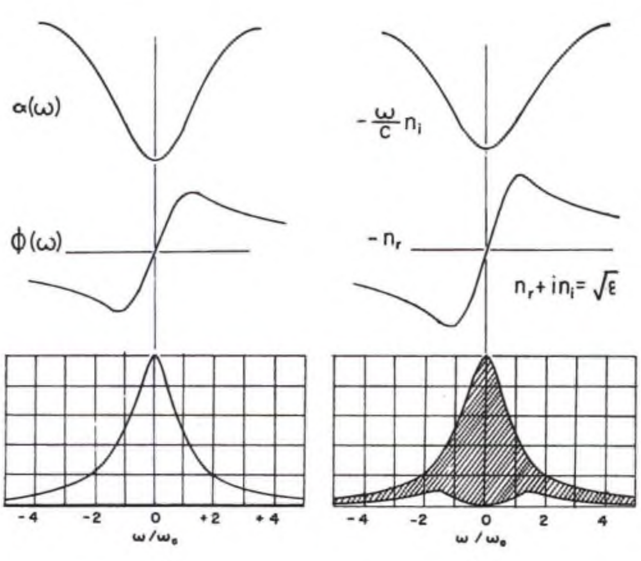


Fig. 6. Representation of the form of the excess delay of a second order allpass lattice.

becomes very large for a small change in frequency. In this case the time delay becomes large without limit. The nature of the delay time for various positions of the poles is shown in Fig. 6.

The form of the delay for the case where b is very much greater than a is the same as in Fig. 2, with the contribution of excess delay occurring at the frequency of b. This leads to the considerable simplification that as long as one is considering local variations in a loudspeaker response, one may consider all activity centered at the frequency of this variation and use the simple expression of Eq. (20). Because local loudspeaker fluctuations in phase and amplitude are usually significant, the equivalent delay and consequently the effective acoustic position relocation may be significant for the frequency of strong local fluctuation. A physical interpretation of this may be secured by observing what would happen if the loudspeaker were fed a transient signal which had in its spectrum this frequency of unusual delay. The pressure wave output would have all frequencies except this component, since for a short time this component will not have arrived. It is a calculable fact that removal of a component is tantamount to adding a cancelling-out of the phase-equivalent component to the original signal. Consequently, the output pressure transient will be perceived to have a "ringing" component at the frequency that is removed. Within some period of time the component frequency will arrive, gracefully one might add, since it is really a distribution of the form of Fig. 2, and the interpretation is that the ringing has now subsided. If the signal is removed from the loudspeaker terminals, the delayed component must persist for some time and the interpretation of this waveform would be that there is a ringing of the output with polarity reversed from the start-up transient.



a b

Fig. 7. **a.** Complete plot of amplitude, phase, and time delay (double valued) for the circuit of Fig. 1 with the frequency of maximum absorption at dc. **b.** Equivalent amplitude and phase characteristic of the transfer function of an electromagnetic wave passing through a single resonance dielectric medium exhibiting anomalous dispersion in which the group velocity by calculation can exceed the velocity of light in vacuum. The frequency dependent time delay (after Brillouin [9]) which has been normalized to the same center frequency of Fig. 7a is a continuum within the shaded region.

ANOMALOUS DISPERSION

A particularly significant distribution of amplitude and phase when discussing group delay is afforded by the transfer function for a real passive dielectric medium with a single simple resonance. The group velocity of a wave propagating in this medium could exceed the velocity of light and gave rise to the term "anomalous dispersion". The concept of group velocity established by Lord Rayleigh was so firmly entrenched that this solution posed a serious challenge to the theory of relativity. So great was this discrepancy that an exceedingly complicated solution was worked by Sommerfeld and Brillouin. Figure 7b is a plot of amplitude, phase, and time delay as worked out by Brillouin [9]. He observed that there was no unique delay, but depending upon sensitivity of apparatus there was a distribution of delays in the shaded region. For comparison with the solution above, Fig. 7a is a similar display for the function of Eq. (22) when a is greater than b. The agreement is quite satisfactory when one realizes that the index of refraction which plays the role of the network transfer function involves a square root of a function of the form of Eq. (22) and hence does not have a simple pole and zero but branch points. The branch points lead to the continuous distribution, whereas simple poles and zeros yield singular functions for time delay.

SUMMARY

A loudspeaker, when considered as a transducer of electrical signals to acoustic pressure, has a transfer function which has a frequency-dependent amplitude and phase response. The effect of these amplitude and phase variations may be considered to be the introduction of a time delay distortion in the reproduced pressure response. The response of an actual loudspeaker will be identical to the response one would have from an ensemble of perfect loudspeakers each one of which assumes a frequency-dependent position in space behind the actual loudspeaker. The number of equivalent loudspeakers, and hence the measure of time delay smearing, will increase with the complexity of the amplitude and phase spectrum. In those portions of the frequency spectrum where the actual loudspeaker is of minimum phase type, it is always possible to modify the response by mechanical or electrical means such that all equivalent loudspeakers merge into one position in space. When this is done there is no frequency-dependent time delay distortion, and the pressure response may be made essentially perfect. Attempts at minimum phase equalization of those portions of the frequency spectrum where the actual loudspeaker is non-minimum phase will not coalesce the equivalent loudspeakers but will leave a spatial distribution which is equivalent to a single perfect loudspeaker with a frequency-dependent position behind the actual loudspeaker.

REFERENCES

1. Richard C. Heyser, "Loudspeaker Phase Characteristics and Time Delay Distortion," *J. Audio Eng. Soc.* **17**, 30 (1969).
2. Rayleigh, *Theory of Sound* (Dover Press, London, 1945), Vol. 1, p. 301.
3. M. J. Lighthill, "Group Velocity," *J. Inst. Math. Applics. I* (Academic Press, London, 1965), p. 1.
4. G. N. Watson, "The Limits of Applicability of the

Principle of Stationary Phase," *Proc. Cambridge Phil. Soc.* **19**, 49 (1918).
5. C. Eckart, "The Approximate Solution of One-Dimensional Wave Equations," *Rev. Mod. Phys.* **20**, 399 (1948).
6. S. P. Mead, "Phase Distortion and Phase Distortion Correction," *B. S. T. J.* **7**, 195 (1928).
7. P. R. Geffee, *Simplified Modern Filter Design* (J. F. Rider, New York, 1963), p. 73.
8. S. Hellerstein, "Synthesis of All-Pass Delay Equalizers," *IRE Trans. on Circuit Theory* **PGCT-8**, No. 3, 215 (1961).
9. L. Brillouin, *Wave Propagation and Group Velocity* (Academic Press, London, 1960).
10. S. J. Mason and H. J. Zimmerman, *Electronic Circuits, Signals, and Systems* (Wiley, New York, 1960), p. 350.
11. G. A. Campbell and R. M. Foster, *Fourier Integrals* (D. Van Nostrand Co., Princeton, 1961), p. 7.
12. R. C. Heyser, "Acoustical Measurements by Time Delay Spectrometry," *J. Audio Eng. Soc.* **15**, 370 (1967).
13. A. Papoulis, *The Fourier Integral and Its Applications* (McGraw Hill Book Co., New York, 1962), p. 148.
14. D. Slepian and H. O. Pollak, "Prolate Spheroidal Wave Functions, Fourier Analysis and Uncertainty," *B. S. T. J.* **40**, 353 (1961).
15. S. Goldman, *Frequency Analysis Modulation and Noise* (McGraw Hill Book Co., New York, 1948), p. 102.

Mr. Heyser's biography appeared on page 41 of the January, 1969 issue of the Journal.

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.

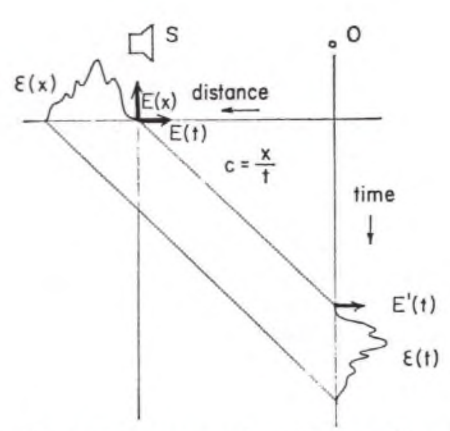


Fig. 1. Symbolic representation of time–distance world line for observer O perceiving energy from loudspeaker S.

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 nonminimum 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

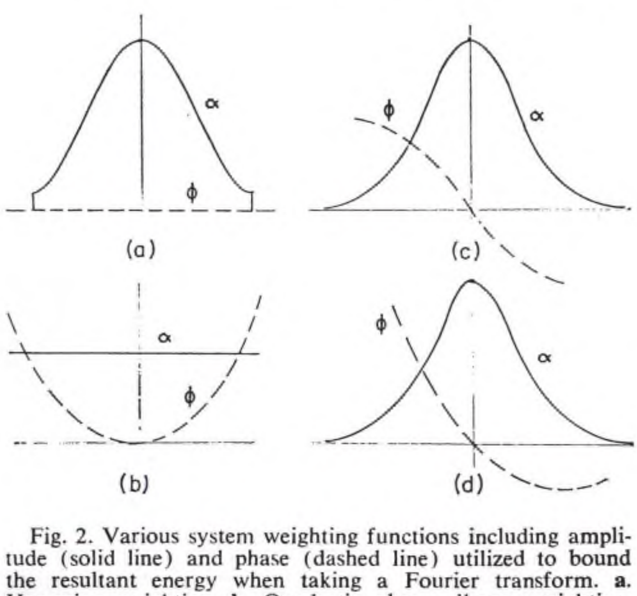


Fig. 2. Various system weighting functions including amplitude (solid line) and phase (dashed line) utilized to bound the resultant energy when taking a Fourier transform. **a.** Hamming weighting. **b.** Quadratic phase all-pass weighting. **c.** Passive simple resonance weighting. **d.** Simple TDS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

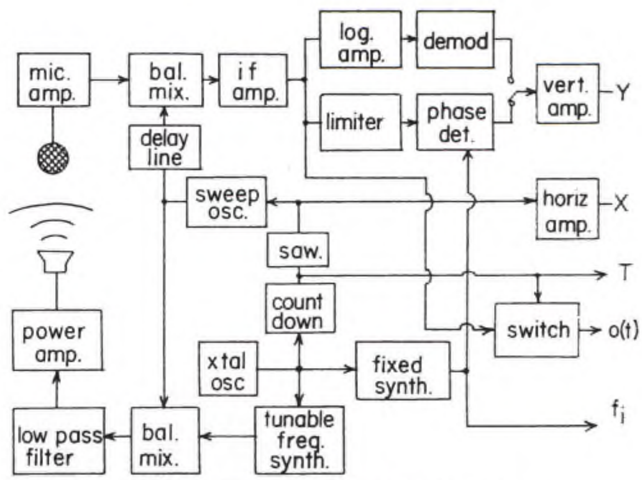


Fig. 3. Block diagram of simple TDS.

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

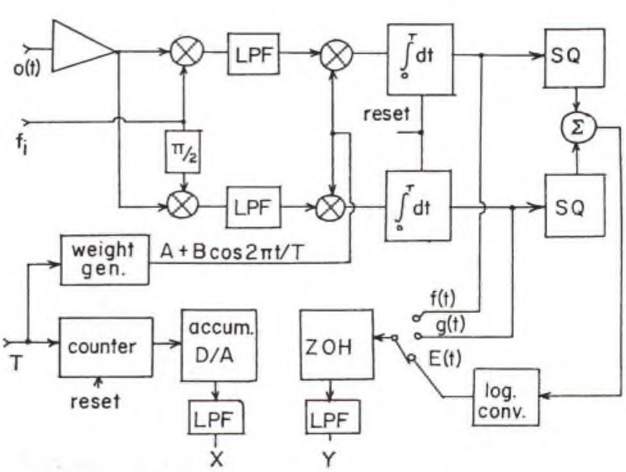


Fig. 4. Block diagram of Fourier transformation equipment attaching to a TDS capable of displaying time-domain plots. **a.** Impulse response $f(t)$. **b.** Doublet response $g(t)$. **c.** Total energy density $E(t)$.

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.

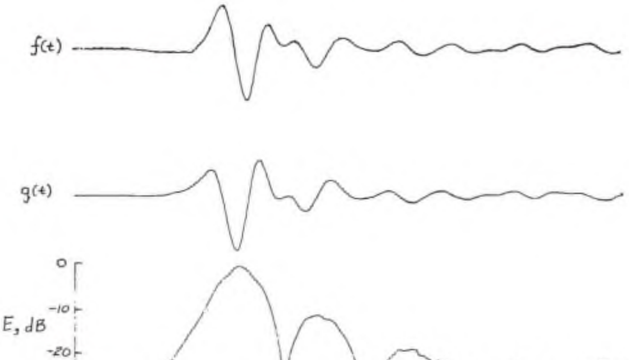


Fig. 5. Measured plots of impulse response $f(t)$, doublet response $g(t)$, and total energy density E for a midrange horn loudspeaker for spectral components from dc to 10 kHz.

* 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*.

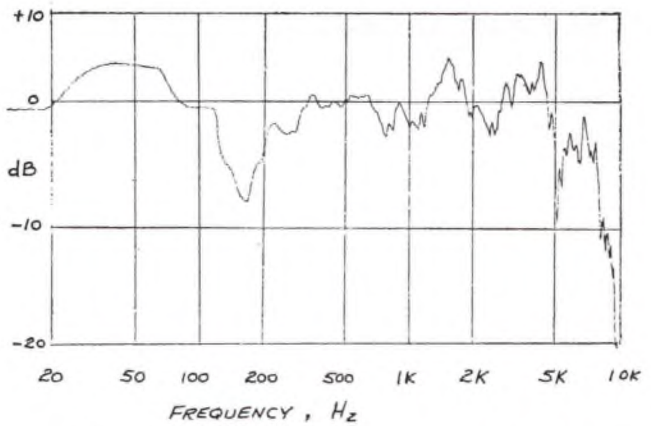


Fig. 6. TDS plot of frequency response (amplitude only) of an eight-inch open-back cabinet mounted loudspeaker with microphone to cone air path spacing of three feet.

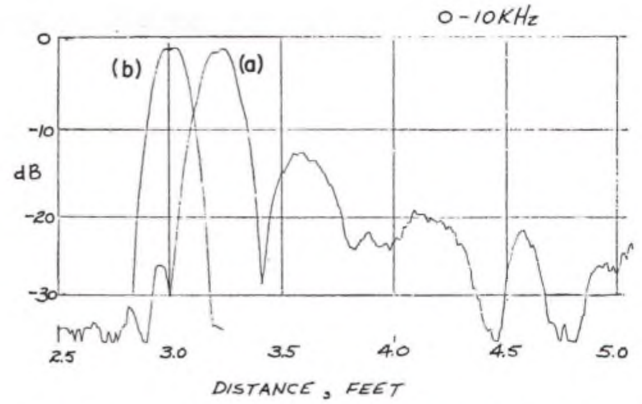


Fig. 7. Curve (a)—Energy-time arrival for loudspeaker of Fig. 6, taking all frequency components from dc to 10 kHz. Curve (b)—Superimposed measured curve to be expected if loudspeaker did not have time-delay distortion.

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.

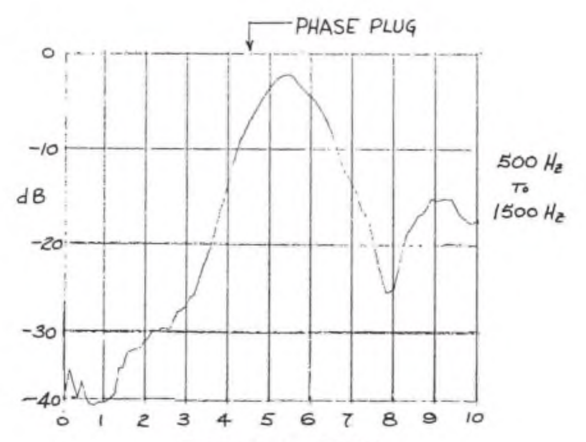


Fig. 8. Energy-time arrival for high-quality midrange horn loudspeaker for all components from 500 Hz to 1500 Hz. Measured position of phase plug is shown.

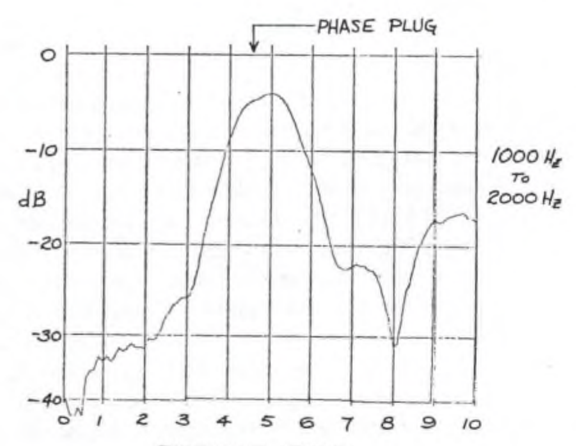


Fig. 9. Same loudspeaker as Fig. 8, but excitation is from 1000 Hz to 2000 Hz.

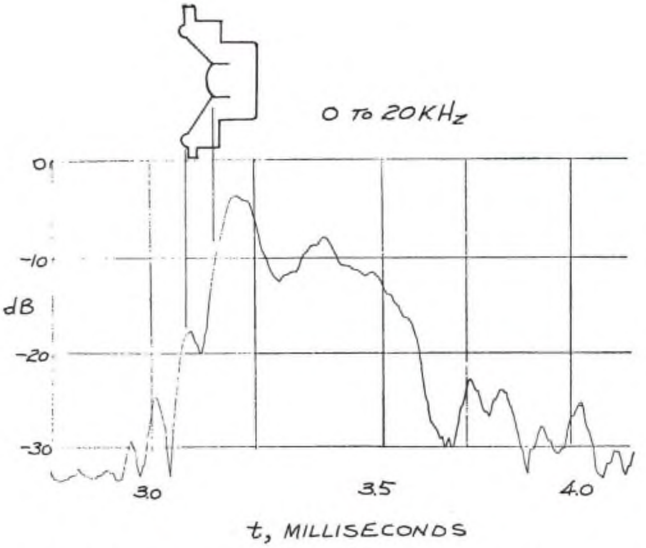


Fig. 10. Energy-time arrivals for unbaffled high-quality eight-inch loudspeaker. Frequency band is dc to 20 kHz and phantom sketch of loudspeaker physical location is included for identification of amount of time-delay distortion relative to speaker dimensions.

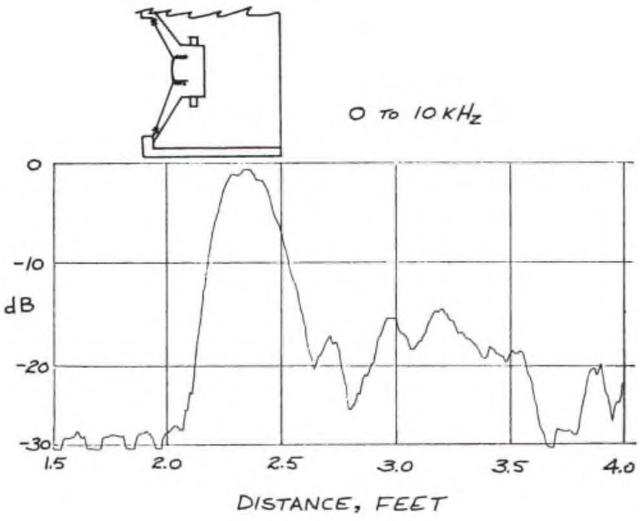


Fig. 11. Energy-time arrivals for open-back cabinet mounted medium-quality eight-inch loudspeaker. Frequency band is dc to 10 kHz and loudspeaker position shown to approximate scale.

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

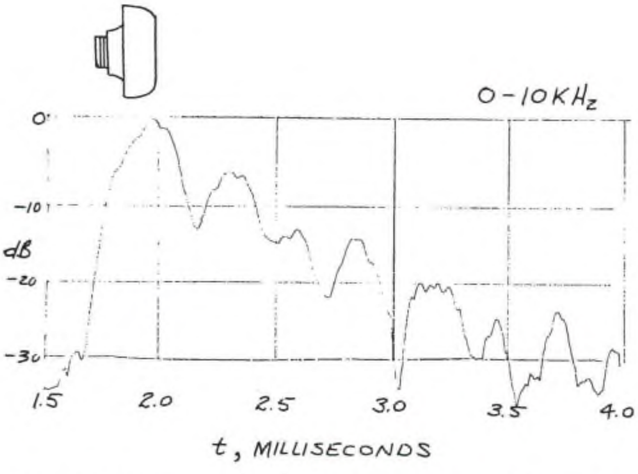


Fig. 12. Energy-time arrivals of unterminated low-quality midrange driver unit. Frequency range is dc to 10 kHz and physical position of driver shown.

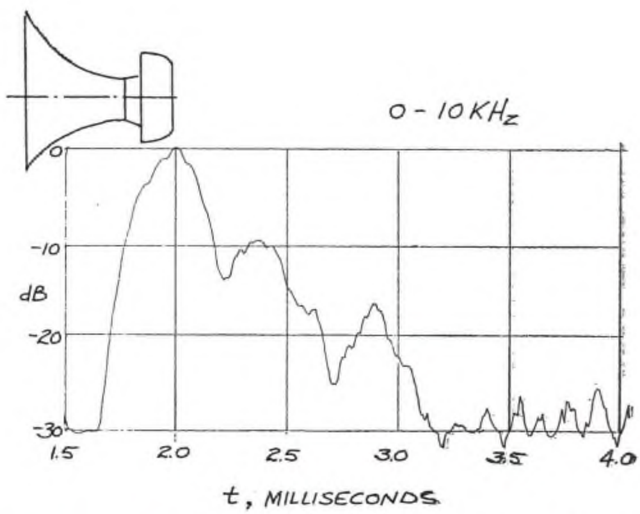


Fig. 13. Energy-time arrivals for improperly terminated driver unit of Fig. 12.

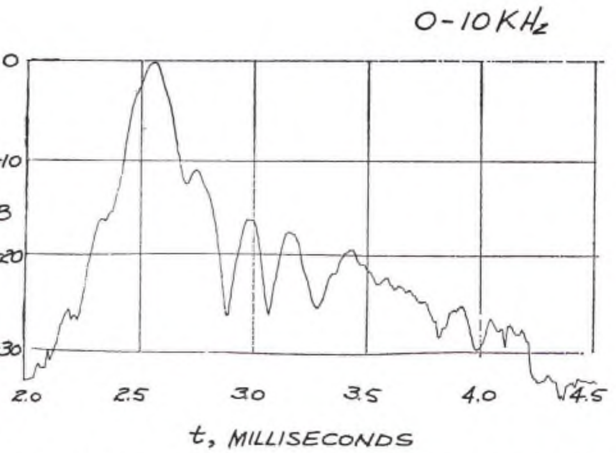


Fig. 14. Dc to 10 kHz energy-time arrivals for midrange horn loaded loudspeaker exhibiting distinct nonminimum phase frequency response.

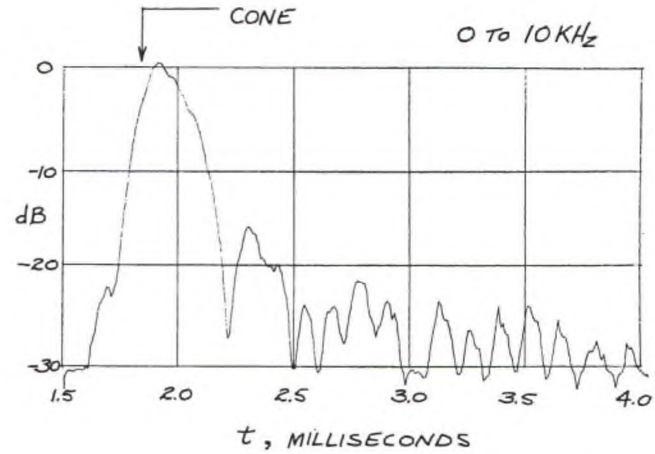


Fig. 15 Dc to 10 kHz energy-time arrivals of paper cone tweeter exhibiting distinct reverberation characteristic.

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,

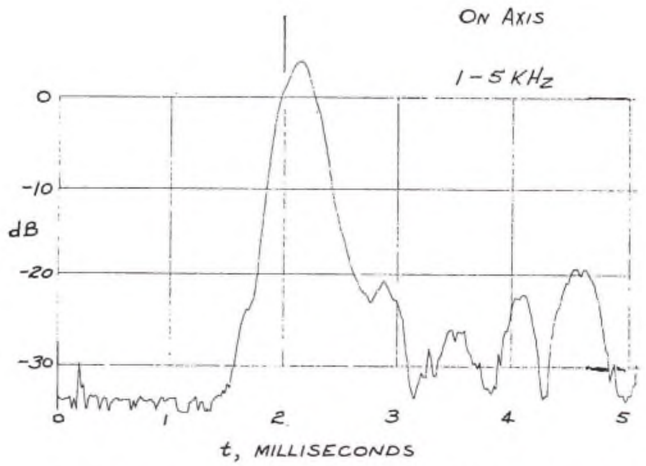


Fig. 16. On-axis energy-time response of high-quality electrostatic multi-panel midrange speaker with position of closest panel equivalent to air path delay of 2 milliseconds and 1–5-kHz excitation.

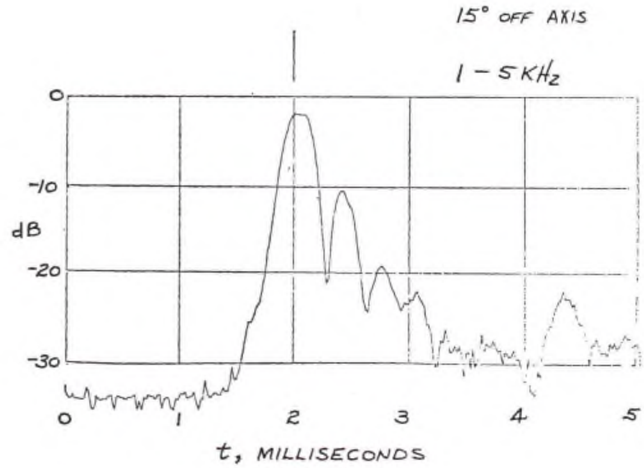


Fig. 17. 15-degree off-axis energy-time response of electrostatic loudspeaker of Fig. 16.

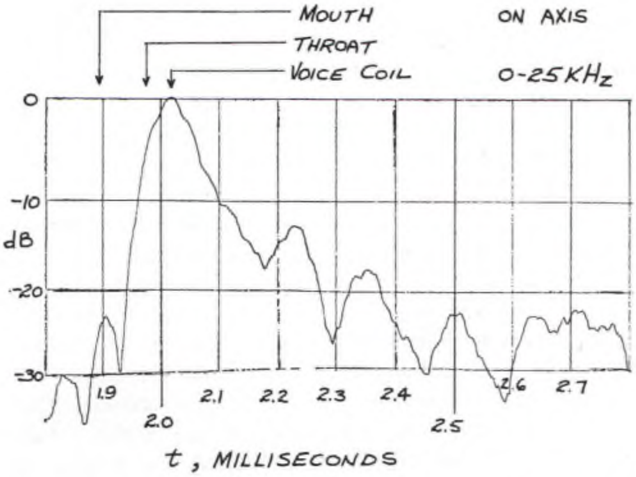


Fig. 18. On-axis energy-time response of a high-quality horn loaded compression tweeter. Frequency excitation is dc to 25 kHz and positions of mouth, throat, and voice coil shown.

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

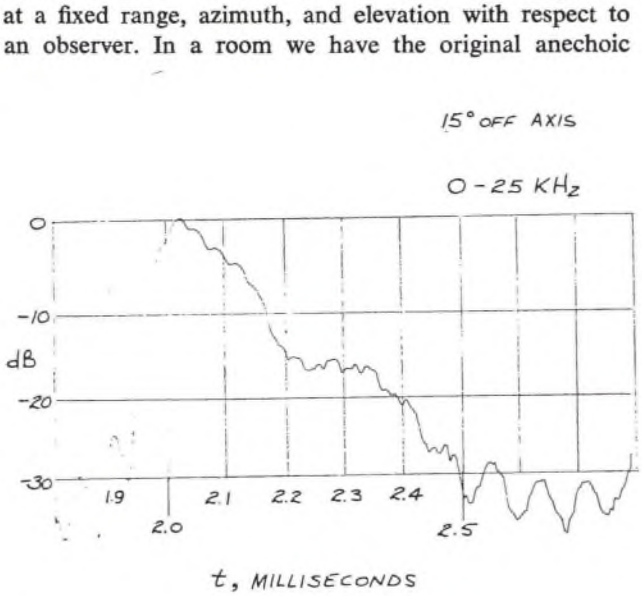


Fig. 19. 15-degree off-axis energy-time response of speaker of Fig. 18.

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

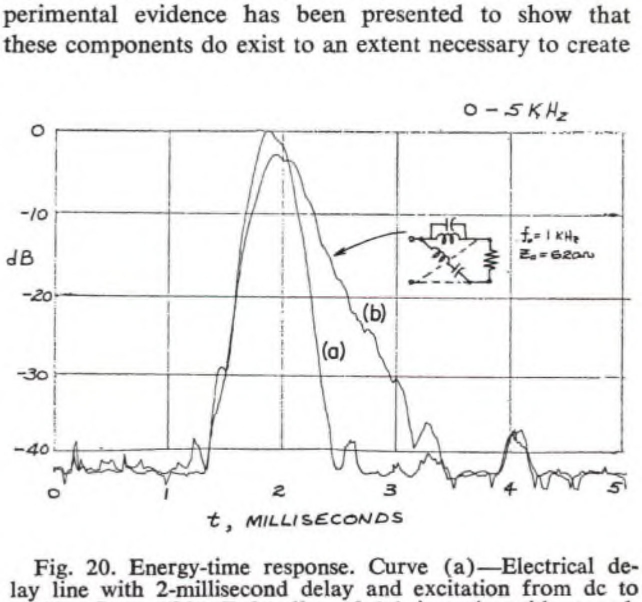


Fig. 20. Energy-time response. Curve (a)—Electrical delay line with 2-millisecond delay and excitation from dc to 5 kHz. Curve (b)—Delay line of (a) in series with second-order all-pass lattice which exhibits severe impulse response distortion due to rapid phase shift at 1 kHz.

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.

* 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*.

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 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

$$\sqrt{T} / \sqrt{V} = \tan \theta. \tag{26}$$

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

$$\sqrt{E} = \sqrt{V} + i \sqrt{T}. \tag{27}$$

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

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

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,

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

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

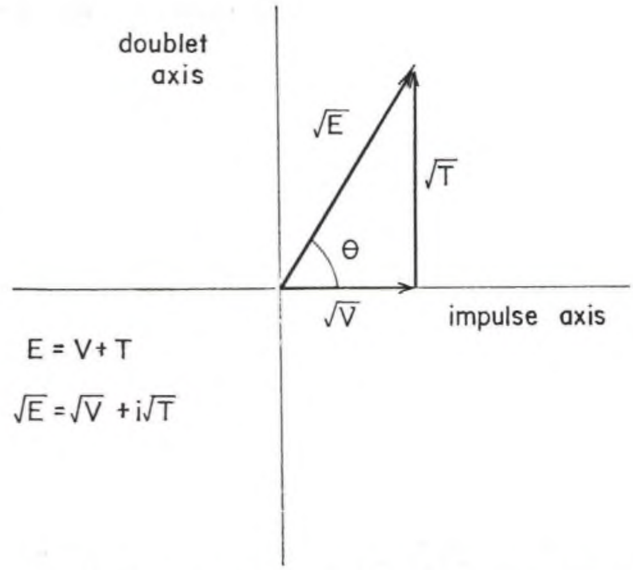


Fig. A-1. Root energy density plane defined such that one axis is system impulse response while quadrature (Hilbert) axis is doublet response.

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),

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

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

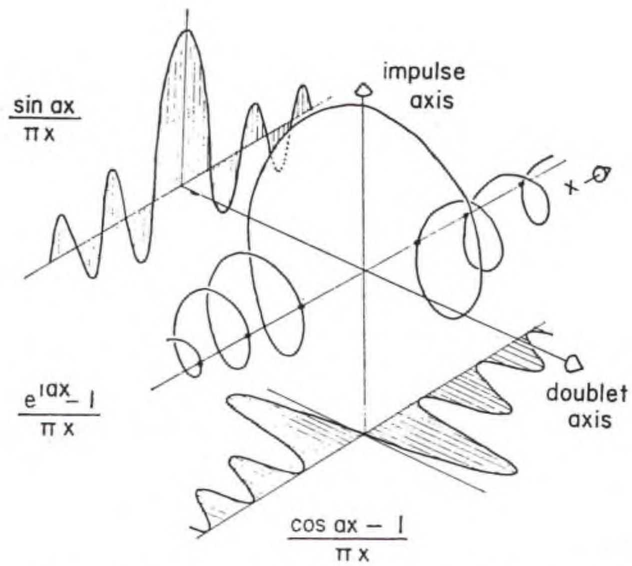


Fig. A-2. Sketch of defined complex energy vector prior to allowing the parameter a to become large without limit. Impulse (7) and doublet (8) are shown as orthogonal projections from this vector.

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

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

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

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

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

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

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

$$\epsilon(y) = \delta(y) + id(y) \quad (34)$$

where by definition

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

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

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])

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

In the limit, utilizing (33),

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

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

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

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.

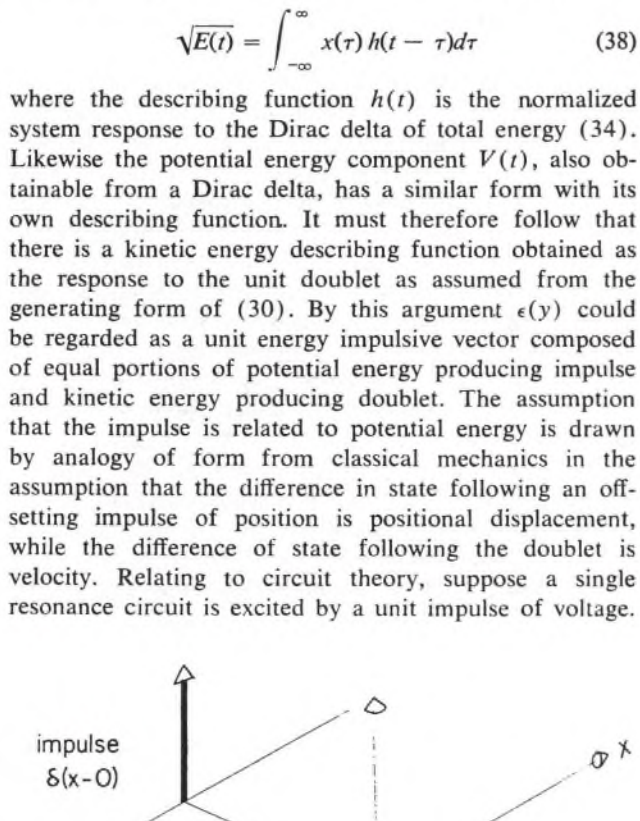


Fig. A-3. Sketch of limiting form assumed by components of Fig. A-2 when a is allowed to go to the limit. Note that the defining envelope of both impulse and doublet even as the limit is approached is proportional to the reciprocal of coordinate x.

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

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

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),

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

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

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

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

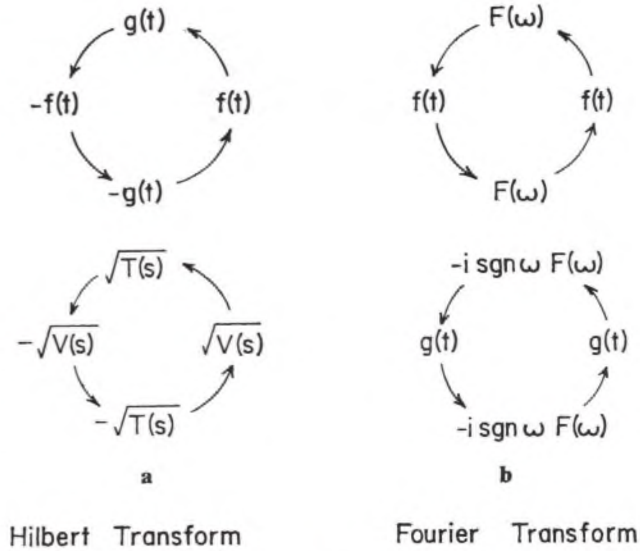


Fig. A-4. Symbolic representation of functional changes brought about by successive applications of **a.** Hilbert transformation to conjugate functions of same dimensional parameter; **b.** Fourier transformation to functions of reciprocal dimensional parameter.

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.

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.

Reprinted from *JAES*, vol. 21, no. 9, pp. 690–701, 1973 November.

The Delay Plane, Objective Analysis of Subjective Properties*: Part I

RICHARD C. HEYSER

Jet Propulsion Laboratory, Pasadena, Calif.

Any audio system can be completely measured by impulse response, steady-state frequency response, or selected variations of these such as square wave, tone burst, or shaped pulse. The measurements will unfortunately always remain unintelligible to the nontechnical user of audio systems. The difficulty lies not with the user but with the equations and method of test. These do not use the proper coordinates of description for human identification. Heretofore the mathematical validity of Fourier transform methods has inhibited a serious search for alternatives. This paper presents the results of a search which shows that the time domain and frequency domain are only two of an indefinitely large number of valid function space representations for audio-system analysis. A particular function space is presented which uses entities of description that may be related to the human identifiable sound attributes of pitch, spectral distribution, intensity, and temporal or spatial spread of an effective source. A meaningful dialogue may thus be envisioned between the engineer performing objective measurement and the nontechnical listener utilizing subjective terminology. An orthonormal set of functions is established which allows an expansion of transfer functions of wave propagation (seismic, sonic, electromagnetic) in terms of relative time delay and spectral content of signal. An algorithm is thus established for mapping between time and frequency without the use of a Fourier transform, and characterization of multipath structures may be made indefinitely accurate without the normal limitations of the uncertainty principle. A closed-form solution is made possible for propagation through dispersive absorptive media.

INTRODUCTION: For a very long time the evaluation of audio or acoustic systems has been hampered by the existence of two schools of thought. One, the objective view, insists that mathematically precise and repeatable measurements are necessary. The other, the subjective view, places no reliance on plots, curves, or equations, but attempts to use attributes which can be related to those evoked sensations of sound which can be verbalized. Attempts at reconciling these two views have been frustrating to say the least since there did not seem to be a way of mathematically describing subjective terminology.

Previous work in the objective evaluation of loudspeakers [1], [2], [4] has shown that the spatial–temporal extent of a signal source can be inferred from the objective measurements. This appeared to point the way to a possible dialogue between objective and subjective interpretation. In particular, a network theory concept presented in this *Journal* [3] looked most promising for that purpose. Accordingly a program of research was instituted with the intent of formalizing some method of objective description which might be subjectively acceptable.

This paper is a preliminary report on the results of that research. The initial goal appears to have been achieved—an analytical tool for audio evaluation which may be interpreted in either subjective or objective terms.

In the approach to be outlined herein the major burden will be placed on the engineer who would be using a new way of writing his equations. However, this also places a requirement on the subjective interpretation such that those descriptions based on evoked images of spatial, temporal, and spectral program content be employed.

The analysis which follows will develop a mathematical description which may be used for problems involving network transfer functions. This is not too restrictive since the results will be applicable to microphones, loudspeakers, tape and disc systems, equalization networks,

* Presented May 16, 1973, at the 45th Convention of the Audio Engineering Society, Los Angeles.

auditoriums, and control room acoustics, in short anything which represents a linear processing of a signal from one point in space and time to another.

PROBLEM OVERVIEW

One of the goals we are trying to reach is the very practical one of meaningful subjective description of the modification of the evoked sound image due to imperfect signal processing. No use of mathematics is or should be made for this purpose. The other goal is a method of meaningful objective analysis using terminology which can be translated into that language of subjective imagery. Not only is mathematics required for such objective analysis, but as will be shown, the complexity is greater than that used for contemporary audio analysis. In a way this is understandable from the observation that if the present tools of analysis had been sufficient, the bringing together of objective and subjective analysis would have been accomplished long ago.

A reasonably short technical paper could be written by presenting the results in a concise form under the hypothesis that the reader is conversant in the mathematics of normed linear spaces and finds the reason for that approach self-evident. In this author's opinion that is not only a disservice to the professional who deserves to know more about new tools of his trade, but is almost certainly assured to inhibit its use by the very person to whom this paper is directed, the practical audio engineer.

The paper will start therefore with an overview of the principles that are to be employed and use heuristic terminology wherever possible. The intent is to present the practical engineer with some of the background of the development of a delay space analysis where sound can be described in terms of a spectral modification combined with an arrival time spread.

The role of mathematics in analysis is so obvious that it is rarely expressed, yet its misinterpretation is the cause of many serious errors. Mathematics is an abstraction concerning the interrelationship among conceptual entities. The language of expression involves signs and symbols generally presented as equations of relationship. Mathematics exists quite apart from physical processes but can occasionally be employed to act as a simulacrum for such processes when an analogy of form can be expressed between the physical observation and a mathematical abstraction. We know, for example, that a loudspeaker does not have to solve an equation in order to function, but this in no way prevents us from predicting loudspeaker performance from our equations because the equations behave analogously to the physical process when we properly identify the conditions of analogy.

The weakness in this lies in the extent to which an analogy exists between the real and the abstract. We must constantly be alert to the possibility that an analogy considered quite good enough to allow prediction of preliminary effects diverges sufficiently in detail that errors of interpretation can occur. That is the situation addressed by this paper.

There is an extremely valuable chapter of mathematics, called Fourier transforms, governing the independent expression of functions of dimensionally reciprocal coordinates. It has long been known in acoustics that one of these coordinates can be brought into close analogy with what we think of as time. The other coordinate appeared closely analogous to an entity of the pitch of a pure tone which had been called frequency by Young near the end of the 18th century. Consequently Rayleigh among others called this Fourier reciprocal time coordinate "frequency" and identified it with a subjective property of hearing [5]. This is the view generally held today.

The analogy is quite acceptable as a coarse measure of pitch, but serious problems emerge upon detail examination as have been pointed out by Carson [6], Gabor [7], and Page [8], for example. An enormous amount of effort has been expended by mathematicians and scientists in bending these two independent coordinate descriptions into a form in which a physical analogy is more closely found.

In this paper we are searching for a meaningful expression of subjective sound images. In order to avoid the quagmire of psychoacoustics, an assumption is made that whatever the process of hearing is, the effects may be verbalized as modifications to the evoked images of sound. Those verbal descriptions are the subjective terms we seek since they constitute a usable language. An investigation of those subjective effects of improper sound processing produces the multiple coordinates of expression to be presented shortly. After removing all coordinates relating to nonstationarity of processing and angle of arrival, two subjective coordinates remain, pitch and time delay. In other words, it does not seem possible to cram the apparent two-dimensional verbal description of subjective impression into the one-dimensional mathematically exact equations of analysis.

In this paper no attempt is made to deform the coordinates as is conventional. Instead another chapter of mathematics in topology and normed linear spaces is investigated to see whether or not other coordinate expressions exist which are more closely analogous to the human observables of sound. This quest has proved successful.

From a purely mathematical point what is found is that the one-dimensional spaces of presently used objective expression are only two of an infinite number of function spaces of representation. All of these belong to what are called Hilbert spaces, and one can find such a space for any number of dimensions one chooses to use.

Recalling that mathematics is used as a simulacrum, we have an answer to a very old dilemma: pitch is not frequency. When we draw an analogy between "wall-clock time" and a coordinate of single dimension, the other reciprocal coordinate, frequency, has a physical analog, but it is not that of pitch because in detail it breaks down.

A description in one dimension, such as impulse response or steady-state frequency response, may be mapped to and from any of the higher dimensional spaces with absolutely no loss of detail so long as the resultant space is a proper Hilbert space. The famed Heisenberg uncertainty relationship [9, p. 119] is a rule of interspace mapping and is not applicable within a valid Hilbert space. In fact, the presently used uncertainty relationship must be a special case of a more general interspace mapping rule among spaces of the same and different cardinality.

In this paper we introduce a two-dimensional space of complex functions of two scalar variables (class C^2) onto which one can map the existing equations of complex functions of one scalar variable (class C^1). So far this is purely a mathematical abstraction. But when we inquire into the analog of these coordinates in human interpretation (when one of the C^1 coordinates is analogous to clock time) we find that one of them is very close to the intuitive concept of pitch and the other is very close to the intuitive concept of time delay. Both of these are the very subjective coordinates below which we appear unable to simplify subjective verbal descriptions.

The outline of this paper then directly follows from that observation. We first find out how to rewrite our equations in this two-dimensional form, which will be defined as a delay plane; then, develop some sort of graphical tool which can be employed to express practical audio situations on this plane. Since the coordinates of this plane are interpretable in subjective terms, we then attempt to tie certain known situations with a delay plane predicted subjective effect to see if it makes sense.

The delay plane is two dimensional, but we can use it to handle azimuth, elevation, and intensity by superimposing solutions. We can thus solve a problem in parts, then assemble the whole for a more complete solution.

As a final practical observation, delay plane techniques are too complicated for analysis of simple networks, unless one wants to know the subjective equivalent. They can, however, be employed to obtain a genuine simplified solution for those situations for which existing methods are hopelessly complicated, such as room acoustics. The fact that the solution is also reducible to subjective terminology lends credence to the point that one should place strong reliance on the observations of a trained human observer, even if the existing mathematical tools do not appear to bear him out. The chances are that he is correct.

COORDINATES OF SUBJECTIVE DESCRIPTION

There are two problems which must be solved before a common language may be envisioned for a dialogue between the objective engineer and the subjective listener. First, a description must be obtained for subjective impressions of sound which may be expressed in mathematical terms. Second, the existing mathematical structure must be transformed to accommodate the terms of this description. Coordinates for a model useful for subjective descriptions will now be considered.

Conceptual Image Space

Hearing, along with sight and touch, are the major sources of information concerning the geometrical order and configuration of the world about us. It will be assumed that each of us has a mental image of a conceptual geometrical framework erected about us and associated with the sources of sound which we hear. This conceptual image space is assembled so as to be consistent with the geometrical spaces of vision and touch. Each source of sound is associated with an apparent object size and position in this space, and the total effect of the distribution of sound sources constitutes the ambience associated with our acoustic environment.

Time Delay

We presume an ordered measure of sound sequence which may be termed relative time for a continuing progression of events or simply time delay for the relative measure of sequenced events. Because we are accustomed to the perception of sound with a fixed propagation velocity, time delay may consciously or unconsciously be associated with the geometry of the conceptual image space. Sound which obviously has the same program content but different time delays and angular arrivals may contribute to the geometrical sound image of an enclosure with reflecting boundaries. Our subconscious ability to localize an apparent source for such an array of sound based on first arrival is well demonstrated as a precedence effect.

Pitch

The sound which we hear will be presumed to evoke a sensation which may be related to pitch. Pitch will be defined as the ordered measure of the property of the purest tone of moderate constant total energy density which can be perceived. The distribution at any moment of the perceived sound in terms of pitch and intensity will be called the spectral distribution of the sound. A subjective measure of spectral distribution will be timbre. Pitch is defined only for a pure tone, but it will be understood that the subjective pitch of a complex tone will refer to the ordered measure of a pure tone of the same loudness judged to have the same subjective basis.

Intensity

Intensity will be defined as related to the total energy density of the sound. The loudness of a sound will be assumed to be the relative auditory sensation of strength of a sound and is related to sound intensity by a pitch-dependent transfer characteristic of hearing. Similarly the auditory sensation of pitch may, in the nonlinear process of hearing, be dependent upon intensity.

These terms of conceptual image space, pitch, intensity, and time delay are definitions assumed for the purpose of this paper and are such as to be defined in either ob-

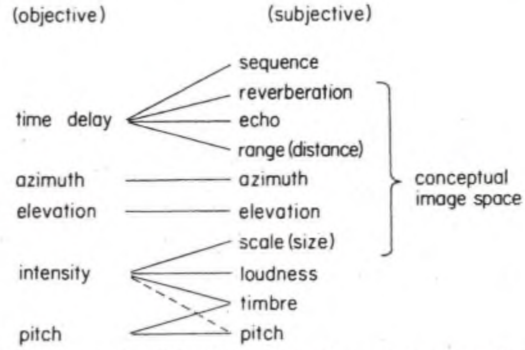


Table I. Several subjective effects in sound perception and their proposed objective counterpart.

jective terms or mathematical symbols. In describing a sound subjectively they may be identified as where the sound is located and how it is spread out, what its pitch or timbre is, how loud it is, and what its time sequence is. Table I is a tentative list of these subjective effects and a breakout of terms which might serve as coordinates for objective description.

Some subjective verbal descriptions now in use and their possible objective interpretation in these coordinates are as follows:

1) Change of Timbre: Bright, shelved, peaked, thin, boomy;

2) Change of Spatial Configuration: Transparent, projects, smears, spreads, reverberant, echoey;

3) Change of Size: Prominent, weak, bigger than life;

4) Combined Spectral–Spatial: Wander, shift, smeared, overlap, indistinct location;

5) Warping of Conceptual Image Space: Improper location, foreshortening, expansive.

A reasonable job of description can be performed by breaking each effective sound source out in terms of its azimuth and elevation angles and treating intensity as a fixed "gain factor" for stationary networks. This still leaves pitch and time delay as network coordinates.

THE CONCEPT OF FUNCTION SPACES

Historically there have been two alternative and independent sets of mathematical descriptions which can be employed for any given audio process, time-dependent and frequency-dependent description [4, p. 735]. Both of these are satisfactory as objective descriptions because instrumental observation analogs exist for the mathematical entities. Neither of these is satisfactory for the subjective human impression of sound which is proposed here, because there are not enough coordinates available, even when one includes azimuth and elevation angle of the sound source.

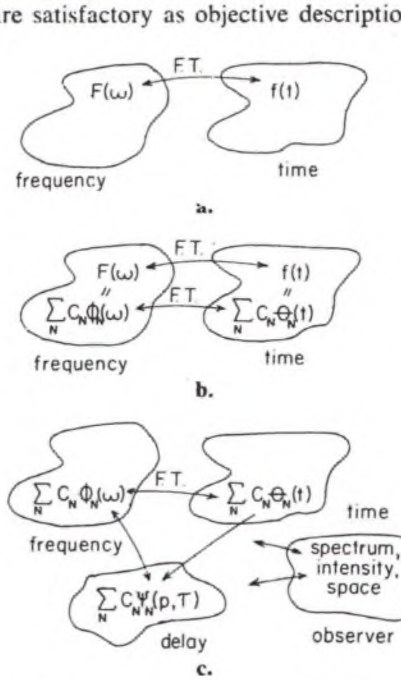


Fig. 1. Symbolic representation of function spaces used in audio analysis. **a**. Steady-state frequency analysis which yields the spectral response $F(\omega)$, and impulse response analysis yielding $f(t)$ can be mapped from one form to the other by Fourier transformation (FT). **b**. An orthonormal set of functions exists within each space such that the steady-state frequency response and impulse response can always be expressed as a linear sum of terms. **c**. The orthonormal terms are a key which allows mapping equations from either of the one-parameter time or frequency expressions to a two-parameter delay space and from that to higher order dimensionality. The multi-parameter descriptions of a human observer may be mathematically characterized in some space and major terms then transformed to a delay space for analysis. Human observation tends to lose meaning and identification for the one-parameter frequency and time descriptions, even though accurate, because the coordinates are not those of human experience.

The solution to this involves a reappraisal of our mathematics of audio engineering. A symbolic representation of the presently used processes involved is shown in Fig. 1**a**. The amorphous blobs symbolize function spaces in which equations may be written to describe an audio process. One of these function spaces will have equations of the form we know as steady-state frequency response $F(\omega)$, and this will be called the frequency space. Another space, which will be called the time space, may be used to describe the same network in terms of the impulse response $f(t)$. We may map a description from one of these two spaces to the other under the process known as Fourier transformation and symbolized by the labeled double arrow. An equation written in either function space can properly and completely describe an audio process.

These particular function spaces have functions defined on the line, that is, in terms of a linearly continuous single coordinate. Therefore the dimensions must be reciprocal, seconds and reciprocal seconds or hertz [4]. Dimensionally reciprocal spaces require all values of coordinate in one space to be used in order for a function to be mapped from that space to a single coordinate in the other. Incomplete use of the coordinates of one will result in a distribution about the single value of coordinate of the other. That is why all possible time must transpire before we can express the true frequency value of a function. This lies at the very heart of the problem of attempting to identify the strictly mathematical concept of frequency with the pitch of a tone.

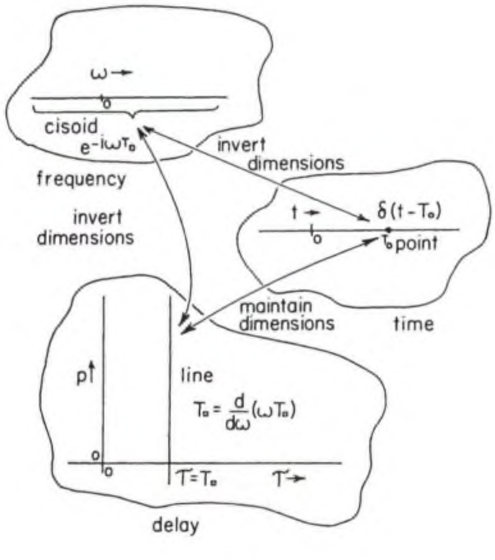
It will be shown first that any network which has a known $F(\omega)$ or $f(t)$ may be expressed as a linear sum of simple terms of an orthonormal set $\{\phi\}$ or $\{\theta\}$, respectively. This is symbolized in Fig. 1**b** for the simplest expansion.

Unlike a Fourier transformation between $F(\omega)$ and $f(t)$ which is the result of an integral process on the entire function, the transformation from the set $\{\phi\}$ to the set $\{\theta\}$ is known term by term and is a linear superposition which can be written down by inspection once the coefficients of each term are known. Table II is included for determining the coefficients from the steady-state frequency response. These results should be of interest to the general audio analyst since they provide an algorithm for interdomain mapping which avoids the normal complexity of Fourier transformation.

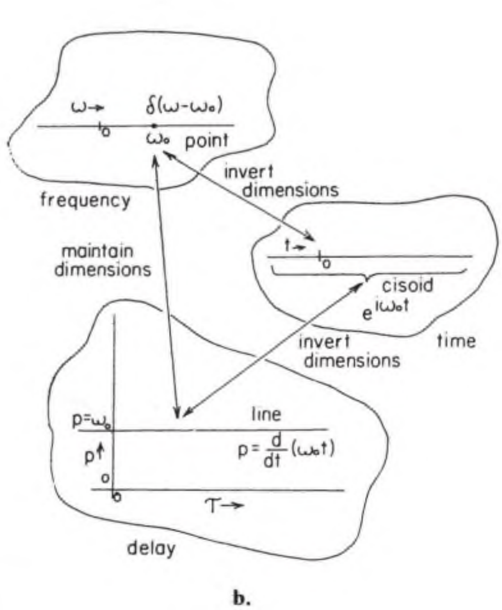
The next step in developing subjectively interpretable mathematical descriptions is symbolized in Fig. 1**c**. It will be shown that there are many function spaces which may be used, not just two. The spaces may have a higher dimensionality than those with which we are familiar. The most significant function space for the purpose of writing equations which have terms with an analog in human experience is the linear delay space. The delay space C^5 has coordinates interpretable as time delay, pitch, azimuth, elevation, and intensity. Space and time invariant propagation systems, common to audio engineering, allow a simple reduction to C^2 for linear analysis with coordinates of time delay and pitch. This is the set $\{\Psi\}$ and it will be referred to in this paper as two dimensional, whereas time and frequency spaces are C^1 and considered one dimensional. Each term of the set $\{\phi\}$ may be mapped to a term of a set $\{\Psi\}$ in the delay space, and, as in the case of conversion from frequency to time, the expansion on the delay space may be written by inspection. A graphical representation of the $\{\Psi\}$ expansions for easier manipulation by engineers will be presented.

These function spaces are, of course, mathematical abstractions and are not the process of sound. They may be used to codify and predict the process of sound if an analogy can be demonstrated between a given functional dependence and a sound-induced effect. The function space labeled "observer" is therefore based upon such an assumed mathematical description for the subjective interpretation of the effect of a network transfer function on program content. If an observer verbalizes the sound as "pinched" or "bright" or certain instrumental voices as "smeared" or "indistinctly located," then a major presumption is made that this can be broken down into effects which are related to the human observables of spectral content, intensity, and relative time or spatial spread. These can then form coordinates of description in some observer function space.

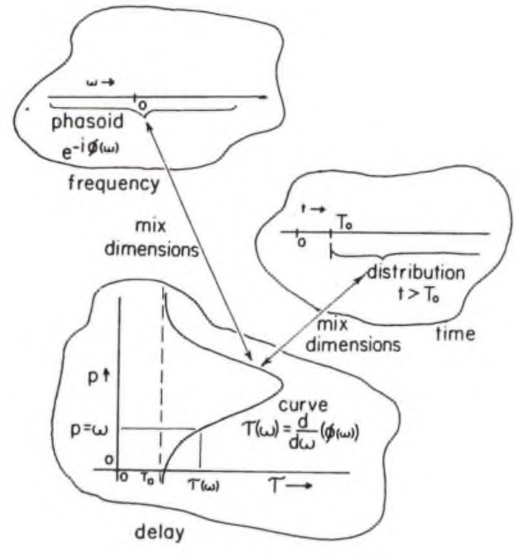
If any one of the mathematical function spaces can be found which uses coordinates uniquely relatable to those of the observer's function space, then the two function spaces can be made isomorphic. In simpler but less precise terminology, an engineer who can measure in or transform his measurements to the proper function space will speak the same language as the observer who uses the proper subjective terminology. The appropriate delay space meets these requirements.



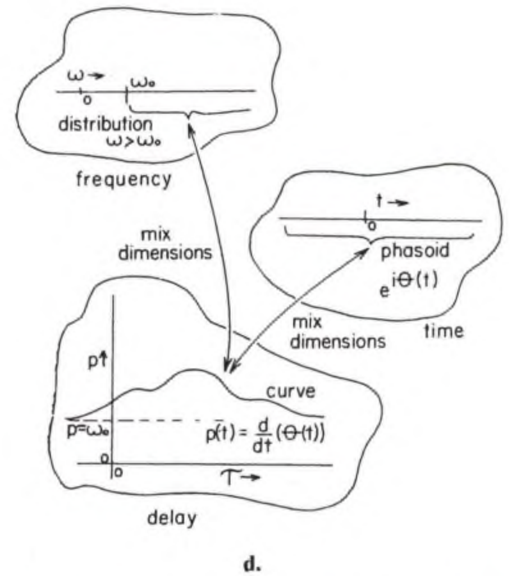
a.



b.



c.



d.

Fig. 2. Symbolic description of the mapping between the one-parameter time and frequency space and the two-parameter delay space. The coordinates in the delay space are equivalent to the human identifiable parameters of pitch, dimensioned in reciprocal seconds and labeled p, and time delay, dimensioned in seconds. **a.** An impulse in time maps to a cisoid in frequency and a coordinate line at a given value of time delay in the delay space. **b.** An impulse in frequency maps to a cisoid in time and a delay space pitch coordinate. **c.** A phasoid, a general phase dependence, maps to a curve in the delay space. The cisoid is a special limiting form of phasoid. Each term of the orthonormal expansion in Fig. 1c is a phasoid with complex multiplier. Any frequency description of the type used in audio network analysis can be expressed as a set of delay-space curves of the type shown in this figure, each one of which has a complex gain. **d.** A time phasoid maps to a curve of pitch dependence as a function of delay.

MAPPING THE LINE TO A PLANE

It is normally considered regressive to take simple one-dimensional functions and rewrite them in terms of two or more dimensions. However, this is necessary in order to join subjective and objective analysis for dynamic program material, and there is nothing canonical about a single coordinate expression which does not work. There appear to be few guidelines in such an endeavor, so the following analysis, which is by no means a proof, may give some heuristic interpretation to this process as used in this paper.

There are simple functions which uniquely tie one function space to another. The cisoid, which is the analytic signal form of the sinusoid, is such a function as shown symbolically in Fig. 2**a** and **b**. The cisoid is used to map the frequency line to a single point on the time axis and conversely. The delay space is represented here as a plane, which will be called the delay plane. The coordinates of the delay plane are time delay τ, in seconds, and pitch p, in reciprocal seconds. This latter coordinate will also be called periodicity. A point in the time space maps to a line in the delay space and the relationships of mapping and dimensionality are shown in Fig. 2**a**. Similarly a point in the frequency space maps to an orthogonal line in the delay space as shown in Fig. 2**b**.

The cisoid is a unit-amplitude linear phase-dependent function which allows a general mapping from the frequency to the time space and conversely. The cisoid is not adequate for general mapping onto the delay space since it may be used only for a straight line which amounts to a coordinate shift. For the purpose of this paper a unit-amplitude phase-variable function will be defined as a phasoid. The cisoid is a special phasoid. As shown in Fig 2**c** and **d**, a phasoid maps into a single curve in the delay space.

The basis for naming the delay plane axes may now be seen. If we experimentally construct a signal which is a replica of a time phasoid and ask for that attribute of sound represented by the time rate of change of phase, it is that subjective property we call pitch. If we double the rate, the subjective effect is a doubling of pitch. What we hear is a flute tone of definable pitch at any moment.

There is an apparent one-to-one relation between that parameter which becomes one axis of the delay plane and the subjective impression of pitch so long as we consider linear effects. Subjective effects are known to be nonlinear at higher sound levels, so we cannot expect a linear space to be a perfect simulacrum. Although not perfect, it is still a great deal closer than what we have been using.

The other axis is labeled time delay for the same reason, a physical network model with that property bears a one-to-one relationship with delay. The isomorphism of L^2 spaces [10, p. 217], [11, p. 43] means that frequency is not pitch by the same measure that time is not delay. Pitch can have meaning at a place in time as time delay can have meaning at a place in frequency.

The frequency space is defined on the line. An analytic simplification of engineering problems, particularly those which are transient problems with a switched start, is made by defining the frequency function on a complex plane. The mapping from this plane to the time space is then performed by a Laplace transformation. Depending on one's view, this complex frequency plane either makes the physically interpretable Fourier transform a special case of the Laplace transform [12] or makes the tools of complex integration available to solve Fourier integrals [13]. The expressions developed in this paper will use the complex frequency $s = \sigma + i\omega$ wherever possible so that either Fourier or Laplace methods may be used by the engineer.

NETWORK EXPANSIONS

Define the generalized all-pass function

$$\phi(\omega,\gamma) \triangleq \phi(\omega) = \frac{i\omega + \gamma^*}{i\omega - \gamma} = \frac{i\omega + (\alpha - i\beta)}{i\omega - (\alpha + i\beta)} \tag{1}$$

where the asterisk means complex conjugation. This is a special phasoid with one pole and one zero, and these are shown in Fig. 3 on the complex frequency plane.

The pole at γ has two degrees of freedom and any such generalized all-pass function may be transformed to any other by a combined frequency coordinate shift and expansion.

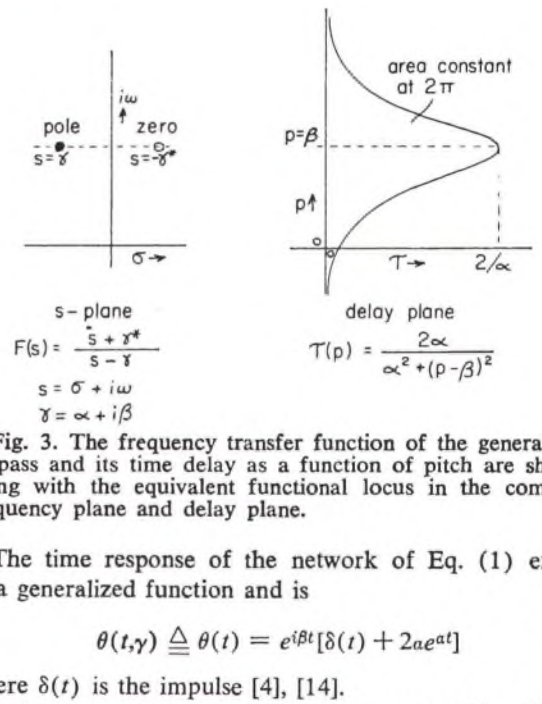


Fig. 3. The frequency transfer function of the generalized all-pass and its time delay as a function of pitch are shown along with the equivalent functional locus in the complex frequency plane and delay plane.

The time response of the network of Eq. (1) exists as a generalized function and is

$$\theta(t,\gamma) \triangleq \theta(t) = e^{i\beta t}[\delta(t) + 2\alpha e^{\alpha t}] \tag{2}$$

where $\delta(t)$ is the impulse [4], [14].

It is a remarkable fact that any reasonably well-behaved frequency transfer function may be expanded as a linear sum of generalized all-pass functions. These functions in fact form an orthonormal set. For network transfer functions which have N simple poles the expression is

$$F(\omega) = \sum_{1}^{N} C_K \phi(\omega,\gamma_K) \tag{3}$$

where

$$C_K = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} F(\omega)\, \phi^*(\omega,\gamma_K)\, d\omega .$$

Usually the frequency transfer function is known as a product of factors involving poles and zeros. In that case there will be a $\phi(\omega)$ for each pole. For example, the conventional resistance–capacitance low-pass filter of Fig. 4 may be expanded as

$$F(\omega)=\frac{1}{i\omega+a}=\frac{1}{2a}-\frac{1}{2a}\left(\frac{i\omega-a}{i\omega+a}\right). \tag{4}$$

The pole $-a$ is real so that the complex conjugation does not change any sign.

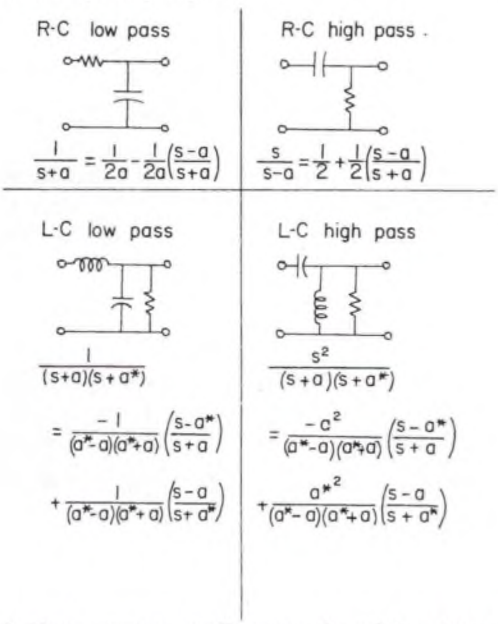


Fig. 4. Four common audio networks with s-plane transfer function are equated to a form demonstrating their equivalence as a sum of generalized all-pass functions. Refer to entries 1, 2, 6, and 10 in Table II.

Other common audio networks shown in Fig. 4 may similarly be expressed as follows:

RC high pass:

$$F(\omega)=\frac{i\omega}{i\omega+a}=\frac{1}{2}+\frac{1}{2}\left(\frac{i\omega-a}{i\omega+a}\right). \tag{5}$$

LC low pass:

$$F(\omega)=\frac{1}{(\alpha^2+\beta^2)+i2\alpha\omega-\omega^2}=\frac{1}{(i\omega-\gamma)(i\omega-\gamma^*)}$$
$$=\frac{i}{4\alpha\beta}\left(\frac{i\omega+\gamma^*}{i\omega-\gamma}\right)-\frac{i}{4\alpha\beta}\left(\frac{i\omega+\gamma}{i\omega-\gamma^*}\right). \tag{6}$$

LC high pass:

$$F(\omega)=\frac{-\omega^2}{(\alpha^2+\beta^2)+i2\alpha\omega-\omega^2}$$
$$=\left\{\frac{1}{2}+i\frac{\alpha^2-\beta^2}{4\alpha\beta}\right\}\left(\frac{i\omega+\gamma^*}{i\omega-\gamma}\right)$$
$$+\left\{\frac{1}{2}-i\frac{\alpha^2-\beta^2}{4\alpha\beta}\right\}\left(\frac{i\omega+\gamma}{i\omega-\gamma^*}\right). \tag{7}$$

For physically realizable networks such as these the poles and zeros will either be real or occur in conjugate complex pairs [3]. The coefficients of such paired terms must therefore be complex conjugate.

Table II is provided to enable the ready determination of coefficients for any network of the type normally encountered in audio engineering and expressed as a rational polynomial fraction factored into poles and zeros. The advantage of the expressions such as Eqs. (4)–(7) is that the time response is a simple sum of the responses of each term. It is not necessary to perform a complicated Fourier transform if the program response is known to only one generalized all-pass network. If one knows the response of only one generalized all-pass network to any program he can calculate the response of any audio network to that program. This is summarized in Fig. 5.

$$F(s)=C_0+C_1\frac{(s-a)}{(s+a^*)}+C_2\frac{(s-a^*)}{(s+a)}+C_3\frac{(s-b)}{(s+b)}$$

	$F(s)$	C_0	C_1	C_2	C_3
1	$\frac{1}{(s+a)}$	$\frac{1}{2a}$	$-\frac{1}{2a}$		
2	$\frac{(s+b)}{(s+a)}$	$\frac{(a+b)}{2a}$	$\frac{(a-b)}{2a}$		
3	$\frac{(s+c)}{(s+a)(s+b)}$	$\frac{c}{2ab}$	$\frac{(a-c)}{2a(b-a)}$		$\frac{(c-b)}{2b(b-a)}$
4	$\frac{(s-a)(s-b)}{(s+a)(s+b)}$	1	$\frac{(a+b)}{(a-b)}$		$-\frac{(a+b)}{(a-b)}$
5	$\frac{(s-a)}{(s+a)(s+b)}$	$-\frac{1}{2b}$	$\frac{1}{(b-a)}$		$-\frac{(a+b)}{2b(b-a)}$
6	$\frac{1}{(s+a)(s+a^*)}$		$\frac{1}{(a^*-a)(a^*+a)}$	$-\frac{1}{(a^*-a)(a^*+a)}$	
7	$\frac{(s-a)(s-a^*)}{(s+a)(s+a^*)}$	-1	$\frac{2a^*}{(a^*-a)}$	$-\frac{2a}{(a^*-a)}$	
8	$\frac{(s+b)(s+b^*)}{(s+a)(s+a^*)}$	$\frac{(b^*+b)}{(a^*+a)}$	$\frac{(a^*-b^*)(a^*-b)}{(a^*-a)(a^*+a)}$	$-\frac{(a-b)(a-b^*)}{(a^*-a)(a^*+a)}$	
9	$\frac{(s+b)}{(s+a)(s+a^*)}$	$\frac{1}{(a^*+a)}$	$-\frac{(a^*-b)}{(a^*-a)(a^*+a)}$	$\frac{(a-b)}{(a^*-a)(a^*+a)}$	
10	$\frac{s^2-b^2}{(s+a)(s+a^*)}$		$\frac{(a^*-b)(a^*+b)}{(a^*-a)(a^*+a)}$	$-\frac{(a-b)(a+b)}{(a^*-a)(a^*+a)}$	
11	$\frac{1}{(s+a^*)}$	$\frac{1}{(a^*+a)}$	$-\frac{1}{(a^*+a)}$		
12	$\frac{(s-a)}{(s+a^*)(s+b)}$	$\frac{(a-b)}{2b(b-a^*)}$	$\frac{1}{(b-a^*)}$		$-\frac{(a+b)}{2b(b-a^*)}$
13	$\frac{(s-a)(s+c)}{(s+a^*)(s+b)}$	$(c-b)\frac{a^*(a+b)+a(a-b)}{2b^2(a^*+a)}$	$\frac{a^*b+ac}{b(a^*+a)}$		$-\frac{(c-b)(a+b)}{2b^2}$

Table II. Basic table for determining the coefficients for transforming a frequency transfer function $F(s)$ into a linear sum of generalized all-pass functions. Any order polynomial $F(s)$ may be expanded using this table. Entries 1–5 are for resistance–capacitance networks, entries 6–10 are for inductance–capacitance terms, and entries 11–13 are auxiliary expressions needed to expand higher order polynomials than those listed.

Because it is an orthonormal expansion, any combination of series and parallel networks will have the same form and properties. Thus while the simple networks such as Eqs. (4)–(7) show little benefit from such expansions, a genuine simplification occurs for the more complicated audio problems, particularly those with multipath time delay as we shall now show.

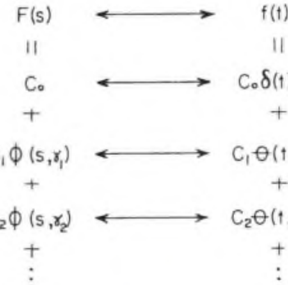


Fig. 5. Because the time response θ of a generalized all-pass function ϕ is always known, the impulse response of any network may be written as a linear sum in the manner shown (Refer to Fig. 1**b**). The response of any network $F(s)$ to any arbitrary input may be written as the linear sum of the response of the known functions ϕ to that input.

DELAY PLANE EXPANSIONS

The generalized all-pass function is the key which lets us map from the frequency space to the delay space. Each phasoid, Eq. (1), maps into a special type of curve in the delay plane as shown in Fig. 3. This curve is such that the time delay τ for each signal component of pitch p is [3]

$$\tau(p) = \frac{2a}{a^2 + (p-\beta)^2}. \tag{8}$$

Each curve in the delay plane has the same characteristic bell shape shown in Fig. 3. The area between this curve and its effective zero delay, to which the curve becomes asymptotic at large periodicities, is constant at 2π. Thus a frequency space pole with a low damping will lead to a large delay and a narrowing of the delay dispersion. A constant multiplier, or intensity factor, is assigned to each curve. Hence pitch, time delay, and intensity are effectively handled.

The use of the delay plane as a graphical representation of complicated audio and acoustic structures may be understood from Fig. 6. The originating signal is most conveniently thought of as curve (a), that is, some program which has a dynamic spectrum spread of pitch components and the existence of which defines a null time delay reference.

If the observation is made with some fixed delay T_0, relative to the actual program, then the curve (b) is the delay plane locus of the observer's program in terms of the original program. This could, for example, be the delay due to sound traveling a distance X at velocity c.

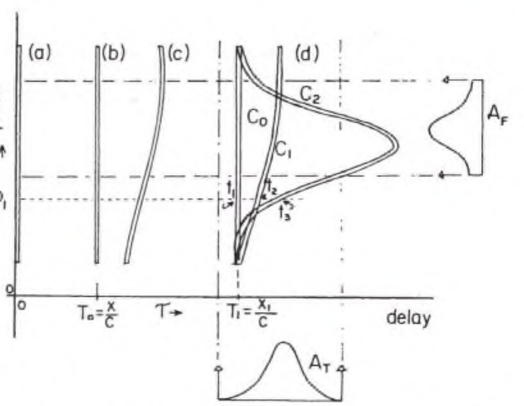


Fig. 6. Delay-plane representation of networks exhibiting the properties of delay, dispersion, and absorption.

Dispersive propagation with pitch-dependent velocity is readily handled as curve (c) if no amplitude change occurs. It is seen that curve (c) could be mapped by a unitary operation onto another delay space such that in that space the curve (c) is straight. This is one manifestation of the indefinitely large number of function spaces available for analysis.

The curves (d) are those due to the generalized all-pass function, and the set represents some network when coefficients C are assigned. The basic time reference T_1 depends upon the nature of the observation. Curves (*d*) may represent the pitch–time spread due to reproduction of program (a) by a loudspeaker and heard at a distance X_1. As far as the listener is concerned, the reference time starts at the first available sound and what is labeled T_1 could be called 0 in his reference system.

As one example of this graphical representation of the delay plane the expression for the low-pass network of Eq. (6) consists of two curves and is drawn to scale in Fig. 7 for representation of a loudspeaker crossover network with a damping of 0.707 and a cutoff of 500 Hz. The peak delay is

$$\tau_{\text{peak}} = \frac{2}{a} = \frac{2}{\zeta\omega_0} = 0.9003 \text{ ms}$$

which will occur at a pitch component of

$$p = \frac{\beta}{2\pi} = 354 \text{ seconds}^{-1}.$$

The two curves of Fig. 7 simply mean that an observer listening to a signal passed through this network hears a distortion that is identical to that which would be

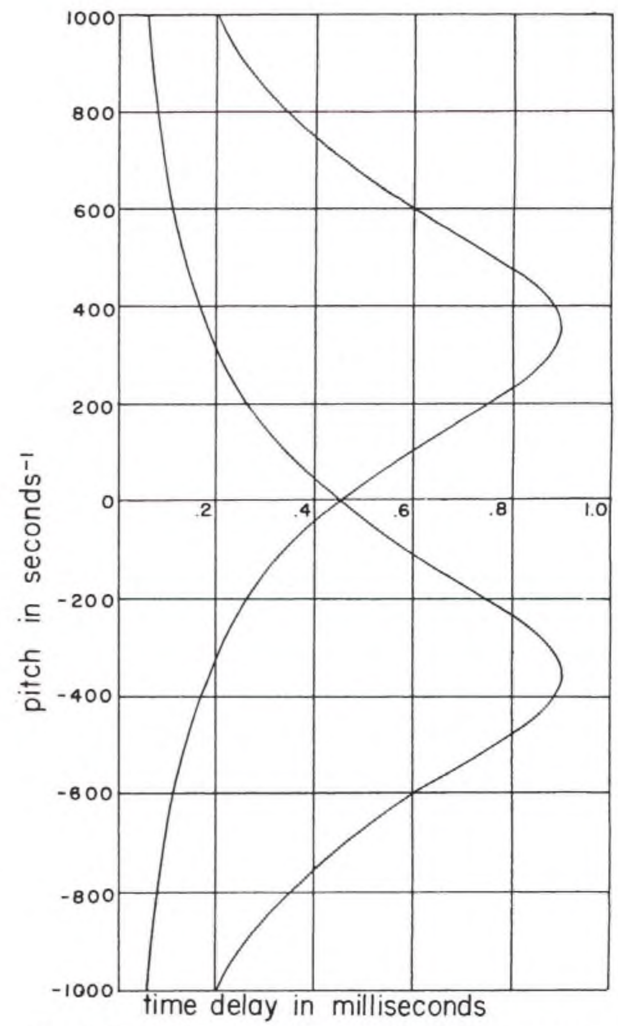


Fig. 7. Delay-plane expansion of the *L–C* low-pass network of Fig. 4. Parameters are dimensioned to correspond to a loudspeaker crossover network with a damping of 0.7 and a cutoff of 500 Hz.

caused by two nearly simultaneous arrivals of the signal. For this network one arrival advances the phase of each signal component by 90 degrees and delays the signal by 0.77 ms at a pitch component equal to a frequency of 500 Hz. The other arrival is a 90-degree phase retardation with 0.13 ms of delay at 500 Hz. At low-pitch components signals reinforce each other; at high-pitch components they cancel, yielding the "low-pass" characteristic of sound associated with this network.

DELAY SPACE INTERPRETATION

At this point we have established that any network transfer function may be expressed as a simple linear sum of basic all-pass network terms. The generalized all-pass network may not be itself physically realizable, but can in parallel combination always represent an actual network. We have further defined that this basic network represents a single valued pitch-dependent time delay. The linear summation means that we may also think of a network as composed of a set of weighted, parallel, pitch-dependent delay lines. This is not a transversal filter [15], [16], or Rake implementation [17], or Hardy filter [9, p. 170] but a parallel configuration. The delay plane, composed of perpendicular axes of program pitch content and perceived time delay, is a simple graphical tool for representing audio networks in terms of the parallel all-pass functions.

The delay plane may be used to place the results of the previous *Journal* papers in context with this work. The measurement of the set of functions constituting the delayed spectrum expansions of Fig. 6 is called time delay spectrometry [1]. The excitation signal used to map a frequency or time space representation of a system into a delay space may be any phasoid which defines a unitary operator. The simplest such phasoid, a linear sweep of pitch versus time, was used for simplicity in all previous discussions.

The orthonormal expansion in a delay space gives rise to an interpretation of the signal propagation as equivalent to a weighted multiple-valued delay to each pitch component. In Fig. 6, for example, a signal at pitch component p_1 will emerge from network (d) as three signals at times T_1, T_2, and T_3. The nature of this distortion imparted to a signal is such that this author named it "time delay distortion" [2]. In order that no misinterpretation occurs, this is not the same as the term "delay distortion" used in communication circuits [18]. The difference is that time-delay distortion relates to true clock time between input and output in a causal sense, and generally is a multiple-valued quantity. Delay distortion is a single-valued quantity derived from the frequency response and is the departure of the slope of the phase-versus-frequency curve from a constant value. As this author pointed out [3] this slope, called variously group delay or envelope delay, is never coincident with true time delay for any minimum phase network, and resulted from an unfortunate misunderstanding of Kelvin's principle of stationary phase as applied to evaluating the Fourier integral along the complex frequency axis. In fact the tools of analysis presented here can now be used to show that group delay bears the same relationship to time delay as "instantaneous frequency" bears [15, p. 81] to pitch.

The delay plane legitimately gives two degrees of freedom for describing network response. The delay plane is definitely not a graphical representation of the joint frequency–time delay properties of a network. The distinction is that the delay plane is a graphical representation of a particular Hilbert space and functions may thus be uniquely expanded in terms of an orthonormal expansion on this plane.

A strong intuitive need has long been felt by analysts for a description of joint frequency–time properties, and so it is common to assemble such a graphical construct for all manner of analysis. Because it is an artificial "gluing together" of two independent Hilbert space representations on the line, a frequency–time plane cannot possess orthonormal expansions.

The delay plane of Fig. 6 may be smeared into a plot of frequency–time delay by merely replacing the coordinate labeled p by one labeled ω. This is now no longer a Hilbert space but is what one would get from assembling an inpulse response and its Fourier transform on one piece of paper. The sharply defined bell-shaped delay plane curves now become smeared and overlap in accordance with the usual uncertainty relation [4]. There is no such uncertainty in the original delay plane simply because the Heisenberg concept is an interspace mapping relationship and does not exist within one space.

The extra degrees of freedom possessed by the delay space can give us the capability of subjective interpretation of the objective description of an audio system. This is because we can meaningfully state when a given pitch component should arrive at the position of a listener. Sound at constant velocity gives us the tradeoff between time delay and apparent distance to a source. There are of course azimuth and elevation considerations, but to introduce the concept we will first presume sound coming from one fixed direction.

There is a complex coefficient associated with each curve on the delay plane, even a fixed time delay. Curve (b) with a real positive coefficient constitutes perfect reproduction from the standpoint of spatial spread of a source relative to a listener. Depending on the magnitude of the coefficient, the source may appear the same, louder than normal, or softer than normal. Subjectively this may yield what has been called "scale distortion" [19], which relates to apparent physical magnitude of the source due to all factors being correct relating to range and pitch, but disproportionate in intensity.

The set of curves (d) with complex coefficients will give rise to a subjective impression involving both a change of timbre, since for each pitch component some curves will reinforce while others cancel, and a spatial spread due to the time-delay distortion. If the curves are so closely spaced that a conversion from pitch to frequency notation produces significant overlap, then the subjective effect may be describable in terminology relating to timbre change and scale distortion but little or no apparent radial smear.

Curves with increasing delay spread will give rise first to subjective terminology relating to reverberant properties, then toward terminology relating to echo and multiple sources as the delay becomes greater.

DELAY SPACE ANALYSIS

There is no more information in a delay-space analysis than in a frequency-space analysis. The tremendous usefulness of a delay space is its closer tie with the human coordinates. An illustration of how one may possibly use a delay plane in an actual situation with angular dependence of sound transmission may be obtained from Fig. 8.

Consider an observer O listening to a two-speaker monitor in a room. The usual situation of a high-frequency speaker S_H mounted forward of a low-frequency speaker S_L is assumed. One sound reflecting surface R is to be considered. The high-frequency speaker establishes the time-delay reference for the observer since the earliest sound comes from that source.

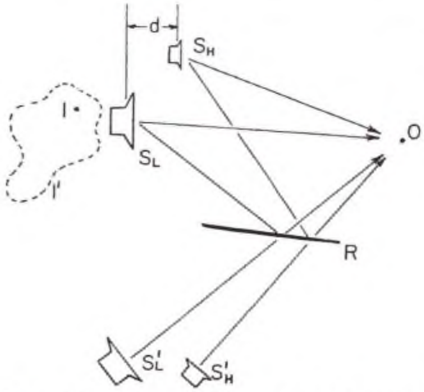


Fig. 8. Geometrical simplification of an observer O listening to the direct sound from a monitor speaker as well as a reflection off a hard surface R. The monitor is assumed to have low- and high-frequency speakers spaced a distance d apart. The sonic effect perceived by the observer is that any sound image point I is smeared to a region I'.

The direct sound from the low-frequency speaker is represented by the set of curves S_L in Fig. 9, and the image speakers S_H' and S_L' lead to the curves shown. If the reflecting surface R is specular, then to first order the curves for the image speakers will be similar to those of the direct source. If it is not specular, then the additional delay-plane curves due to its reflection properties must be added to those of the image speakers to account for this distortion. No matter how complicated the situation, one only adds or subtracts curves from the delay plane, never changes the form of those already there.

If each of the four sets of curves were drawn on a separate transparency and overlaid in accordance with the time delay of each set, the effect would be as shown in Fig. 9. Each curve within the set has an intensity factor associated with it so that not all curves are important for determining the gross character of sound.

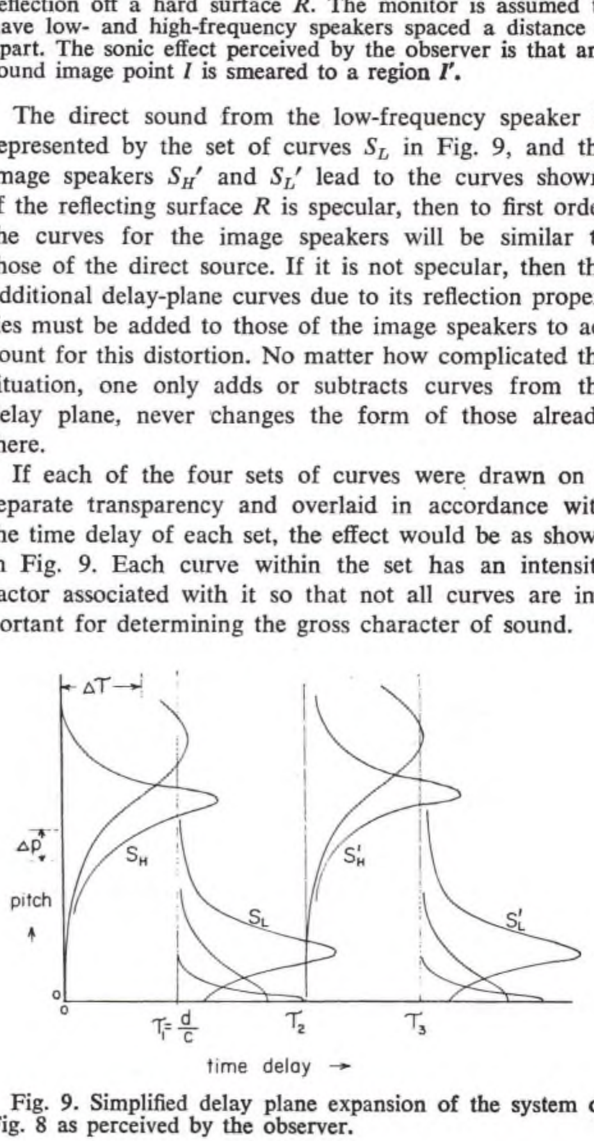


Fig. 9. Simplified delay plane expansion of the system of Fig. 8 as perceived by the observer.

What this tells us about the reproduction is as follows. Each delay-plane curve represents a separate path between source and listener. Each curve therefore amounts to a separate reproduction of the program material fed to the loudspeaker system. What a listener hears is an ensemble of programs, not just one. Each member of the ensemble has its own strength. The loudest members will dominate and produce the effect we call the "sound" of the program. All of the members of the low-frequency channel merge and cancel for high-pitch components, producing the characteristically dominant lower pitch program content. The converse is true for the high-frequency channel.

If one wished to know what the effect would be of an anechoic chamber rendition of a program, he could remove the S_H' and S_L' transparencies (and of course all other room reflection sets) leaving S_H and S_L. Then a change in spacing of low- and high-frequency speakers could be assessed by sliding S_L parallel to the time-delay axis. The effect of the room alone would be equivalent to replacing all room reflection curves and removing S_H and S_L. A change in the geometry of the room could be represented by sliding S_H' and S_L' relative to the direct-sound time delay to account for the appropriate path length delay. This graphical manipulation of curves can quickly lead to an intuitive feel for the subjective effects on the sound heard through the system.

If the electrical signal fed to the loudspeakers were such that a perfect loudspeaker would produce a point image I in the observer's conceptual image space, then the delay plane can be used to estimate the apparent deformation due to room and transducers. If the equivalent point I has pitch components in the range Δp, then S_H will spread the image in range by some $\Delta\tau$ and S_H' will similarly spread the image and pull it toward S_H' as perceived by O. If the observer were used to listening to a sound I in this room, then S_H' will not materially alter the angular position, but it will provide some angular smear. The effect of the imperfect reproduction will be the pitch-dependent spreading of the point I into the region I'.

The effective spread of the lower pitch terms of S_L will be larger than those of S_H for most loudspeakers, since the poles are closer to the frequency axis for a given cutoff rate in decibels per octave. This is the basis for the general rule presented in [2] that the mean average position of a sound source behind a loudspeaker, based on total energy density, is roughly inversely proportional to the high-frequency cutoff. Even if the observer O cannot perceive the individual delay terms, he will experience an apparent pitch-dependent spreading.

What the delay plane curves tell us about the loudspeaker spacing d is that there is a minimum critical spacing below which subjective quality will not improve due to the existing time spread of the individual speakers. The minimum allowable spacing for perceptible distortion will decrease with improvement in transducers. Because program content distortions add more delay-plane curves to those of the speakers, the critical spacing will depend on program material and content as well. Some program material may be so badly time-delay distorted that it makes little difference where the speakers are positioned relative to each other.

SUBJECTIVE INTERPRETATION

Subjective interpretation of defects in sound reproduction is an intensely personal experience which is not amenable to engineering analysis by means of a response to a prescribed stimulus. It is better understood in terms of illustrative examples which have a counterpart in the common experience of sound. A hypothetical experiment for illustration of subjective sound perception may be envisioned as follows. Assume one were listening to a sound which was created by a natural process in a natural environment. For simplicity, assume it to be the voice of a friend. We have an acoustic as well as visual spatial image of the person talking. We could, if need be, point to the place in the room where the sound of his voice placed his conceptual image. If we were to close our eyes, his conceptual spatial image would not change position (a ventriloquist is well aware of this in providing visual as well as acoustic clues for altering the apparent physical location of a voice or sound). Assume somehow that while our eyes are closed, the person were spirited away and replaced by a sound reproduction of his voice. If the reproduction were perfect we would be unaware of the deception and could, upon request, point to a position in space where his conceptual image appeared to us. The question to which we address ourselves is, what would be the subjective effect of an imperfect reproduction?

The foregoing delay space analysis of imperfect reproduction indicates that the voice will be altered in spectral content or timbre and will change in size and be spread in the conceptual image space. Even if the total sound intensity is unaltered, the image corresponding to the voice may appear unusually large with a width and depth which is dependent upon the spectral content of the words being spoken. Even if the steady-state frequency spectrum of the reproducing device has good high-frequency response, the "liveness" of breath sound may be lost due to a spatial smear. Because the lower frequency components tend to greater relative spatial spread, his voice may subjectively appear stronger in bass components than warranted by frequency response.

If the entire system from microphone through loudspeaker is perfect, and if the listening environment is such that the human listener can unconsciously compensate for its deficiency by assigning a modification in conceptual space to account for the sound as it should be perceived in that environment, then the perceived sound will be to all intents a perfect reproduction. A nonuniform transmission will be identified with a warping of the conceptual image space. In highly symbolic terms the conceptual image space is warped in a manner analogous to the visual equivalent of looking at a scene through a multiplicity of panes of glass each one of which has a color-dependent thickness.

Because one effect of an imperfect response is a time smear, a subjective impairment of program acoustics should be evident in depth perception. A single instrumental voice will be unnaturally spread out in depth from a nominal position in the conceptual image space. Multiple voices which should occupy different apparent distances from the observer may tend toward closer apparent spacing and in extreme cases appear to merge at a poorly defined position. The reason for this would be the system delay terms which treat all program sources alike, unlike a natural acoustic environment in which complexity of transmission tends to increase with apparent distance from us.

All of this is purely conjectural and open to criticism since it is based on interpretation of delay-plane expansions as an equivalent acoustic effect. It does appear to this author to more closely correlate with subjective impressions than impulse or steady-state frequency response, even though they are known to be completely valid. It also appears to explain why certain defects of reproduction can have extremely bad impulse or square-wave responses, such as all-pass filter transmission, yet be quite difficult to hear [20], while other defects which barely show instrumental evidence, such as print-through, can be extremely objectionable.

CONCLUSION

In the process of formalizing audio system descriptions in terms with subjective interpretability it would appear that some of the results of this paper have been common knowledge for a long time. Thus the concept that an imperfect sound reproduction produces a spatial diffusion as well as spectral modification is not startling, nor should it surprise anyone that the pitch of a tone can change with time such as a glissando or that a tone of defined pitch can exist for a small time, then be silenced.

What is different about this is that these subjective effects result from a new analytical approach, and such descriptions could not result from the mathematics normally used for audio systems. Their significance thus rests on the fact that they are the result of an objective analysis rather than assumptions made prior to analysis.

It is apparent that a mathematical pedigree has been established for something which was common knowledge all along, namely, that the subjective interpretation of sound is not only describable but involves several independent variables. It has been shown that the conventional one-dimensional objective descriptions of time and frequency are only two of an indefinitely large number of possible ways of describing network transfer functions. Furthermore, these may use two or more independent variables. None of these function spaces is more complete than any other, and the reason for complicating one's mathematics by using different systems is that some of them may be closer to subjective concepts than those which we now use. It is thus perfectly plausible to expect that a system which has a "better" frequency response may in fact sound worse simply because the coordinates of that measurement are not those of subjective perception. One should not expect a one-dimensional audio measurement to be meaningful in portraying an image of sound any more than he could expect an art critic to be appreciative of a painting efficiently encoded and drawn on a string.

Audio engineering is a very practical field. The concept of Hilbert spaces and orthonormal functions might appear far removed from mixdown sessions or auditorium acoustics. In fact, when one begins to translate the mathematical language into audio terms he is struck with the observation that much of what results reinforces his common sense ideas. More importantly, ideas emerge which may not have otherwise resulted. This is what it is all about; not becoming conversant in mathematics, but developing better audio engineering.

The orthonormal expansions among other things mean that a reflecting surface adds terms in the sound field but does not change other terms not dependent upon that surface. It comes as no surprise then that a mixing console surface does not modify the direct sound from the monitor speakers but adds its reflection component to the sound perceived by the mixer. What is new is that the analysis shows that there is a way not only of isolating the console reflections, but of interpreting the net sound as a combined spectral–spatial modification of the sound image as perceived by the mixer.

We have by no means exhausted the mathematical tools which may be utilized, but have restricted the discussion to those necessary to establish a dialogue between objective and subjective interpretation. Throughout all of this one common-sense fact should be kept in mind, the electrical and acoustic manifestations of audio are what is real. Mathematics is a simulacrum we employ to model and predict our observations of the real world. We should not get so impressed with one set of equations that we assume the universe must also solve these equations in order to function. It does not.

The material presented in this paper is a preliminary report of research still in progress. One hazard of reporting work in progress is that no opportunity exists for a polished presentation with all loose ends tied and all conclusions tabulated. A number of concepts new to audio engineering have been introduced in order to present a basic principle of delay-space analysis. In order to make this material available to audio engineers in a paper of reasonable length, it has been necessary to eliminate many details of development. A substantial amount has been left out, so this paper should be considered only as an introduction to the concept.

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] R. C. Heyser, "Determination of Loudspeaker Signal Arrival Times" (in three parts), *J. Audio Eng. Soc.*, vol. 19, p. 734 (1971); p. 829 (1971); p. 902 (1971).

[5] Rayleigh, *Theory of Sound*, vol. 1. (Dover Press, London, 1945).

[6] J. R. Carson, "Notes on the Theory of Modulation," *Proc. Inst. Radio Eng.*, vol. 10, p. 57 (1922).

[7] D. Gabor, "Theory of Communication," *Proc. Inst. Elec. Eng.*, vol. 93, p. 431 (1946).

[8] C. H. Page, "Instantaneous Power Spectra," *J. Appl. Phys.*, pp. 103-106 (1952).

[9] H. Dym and H. P. McKean, *Fourier Series and Integrals* (Academic Press, New York, 1971).

[10] P. L. Butzer and R. J. Nessel, *Fourier Analysis and Approximation*, vol. 1 (Academic Press, New York, 1971).

[11] P. R. Halmos, *Introduction to Hilbert Space* (Chelsea, New York, 1957).

[12] R. Bracewell, *The Fourier Transform and its Applications* (McGraw-Hill, New York, 1965), p. 220.

[13] E. A. Guillemin, *The Mathematics of Circuit Analysis* (M.I.T. Press, Cambridge, Mass., 1965), p. 547.

[14] M. J. Lighthill, *Fourier Analysis and Generalized Functions* (Cambridge University Press, Cambridge, England, 1959).

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

[16] H. E. Kallmann, "Transversal Filters," *Proc. IRE*, vol. 38, pp. 302-311 (July 1940).

[17] M. Schwartz, W. R. Bennett, and S. Stein, *Communication Systems and Techniques* (McGraw-Hill, New York, 1966), p. 580.

[18] W. R. Bennett and J. R. Davey, *Data Transmission* (McGraw-Hill, New York, 1965), p. 84.

[19] A. McKee, personal communication.

[20] P. W. Klipsch, "Delay Effects in Loudspeakers," *J. Audio Eng. Soc.*, vol. 20, p. 634 (1972).

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. 21, no. 10, pp. 786–791, 1973 December.

The Delay Plane, Objective Analysis of Subjective Properties* Part II

RICHARD C. HEYSER

Jet Propulsion Laboratory, Pasadena, Calif.

Editor's Note: Part I of The Delay Plane, Objective Analysis of Subjective Properties appeared in the November issue.

INTRODUCTION: Part I demonstrated that objective analysis of subjective sound properties is possible only by expanding analysis beyond the conventional approach of time domain and frequency domain. The case was presented that time and frequency, rather than being the only valid coordinates of representation, are in fact only two valid functional representations out of an indefinitely large number. A useful working audio tool was presented in the form of a C^2 space which could be manipulated when presented as a two-dimensional chart, called the delay plane. In order to present the nonmathematically inclined audio professional a heuristic working concept in Part I, the necessary mathematical basis is presented as a separate section in Part II.

MATHEMATICAL ANALYSIS, PREFACE

This paper draws heavily from the mathematics of normed linear spaces. This important branch of mathematics is not covered in audio literature, even though the audio engineer is rapidly expanding his technology of multidimensional sound processing into areas where he will find these tools valuable. Unfortunately, mathematical literature is seldom written in audio identifiable terms, and no single reference addresses all the points the engineer should know. The first two sections on definition and Hilbert space summarize the salient engineering features of this branch of mathematics and provide a digest version for the needs of this paper as well as a minimum background for audio engineers. The particular references used for these two sections are [9]–[11], [15], [21]–[28]. Individual contributions from each reference will not be signified since in many cases this author had to combine the contributions of two or more in a manner and terminology felt most suitable to audio engineering. The remainder of the mathematical analysis is original work.

DEFINITION OF TERMS

The main premise of this paper is that there are alternative sets of elements for the characterization of a network transfer function. By element is meant an entity of description such as the response to a sine wave. By set is meant the total collection of such elements. Mapping is the rule by which elements of one set may be assigned to elements of another set. The process of mapping is also called a transformation, or it may be performed under the control of what is called an operator. The mapping produces what is commonly called a function if the coordinates of the set represent numerical

* Presented May 16, 1973, at the 45th Convention of the Audio Engineering Society, Los Angeles.

values. One example of a coordinate representing numerical values is the set of all real numbers R, called the real line, such as used for the time domain of representation. Another example is the set of all complex numbers C, called the complex plane, such as used for the complex frequency domain of representation of Laplace transform theory. A highly structured numerical framework of representation where each element can be considered a single entity or point is called a function space. For the present paper the set of all complex functions defined on a coordinate basis composed of an n-fold of real numbers and representing a linear space will be called an n-dimensional complex space C^n. It is the intuitive extension of complex functions to an n-dimensional space. The frequency response of an audio network is a complex function (with amplitude and phase) of the single coordinate of frequency, hence is C^1.

The set of all numbers defining a connected open region of the function space is called the domain of the function defined in that region of the space. The relative distance between elements in the framework comprising the function space is called a metric of that space. A set of elements together with a metric for that space is called a metric space. Two different metrics defined on the same elements form two different metric spaces. A particular metric of value is one with the property of determining the size of any element as its distance from the origin. This real positive number is called the norm of the element x, denoted by $||x||$. The double vertical bars may be thought of as symbolizing that a very special absolute value is implied for the element. Because we are dealing with physical networks which can both store and exchange energy as well as dissipate it as heat, we can expect a functional description of an audio process to be a vector quantity. We can geometrically think of a vector as a directed line segment. The norm is the length of the vector which starts from the origin. A geometrical measure of how much of one vector is composed of, or is embedded in, another vector is afforded by how nearly perpendicular they are to each other. A complex measure which reduces in the simpler geometries to a value proportional to the cosine of the angle between two vectors is the inner product, also called scalar or dot product. The inner product between two vectors x and y is symbolized by (x, y). If two vectors are perpendicular their inner product is zero and they are said to be orthogonal. Vectors which are orthogonal and have a unity norm are said to be orthonormal. If the norm is determined by an inner product relationship between pairs of elements in a space and if convergence of all elements is what is termed complete for this norm, then the space is called a Hilbert space.

A more formal definition results from the following. The set of all functions defined and measurable on a set T in n-dimensional space, and with the property that the square of the norm is finite and induced by an inner product,

$$||x||^2 = (x,x) = \int_T x(t)x^*(t)\,dt < \infty$$

is a separable Hilbert space, which is designated $L^2\ (T)$ when the inner product is taken to be

$$(x,y) = \int_T x(t)y^*(t)\,dt.$$

When this is established, it follows that there is for every x in the space a unique expression

$$x = \sum_1^\infty C_N e_N$$

where $\{e_N\}$ is a countable set of a complete orthonormal set and $C_N = (x, e_N)$.

The series representation is called the generalized Fourier series of x with respect to e_N. If, for example, T is the interval $[0,2\pi]$ of the real line, the expansion is the usual Fourier sine and cosine series. It should be observed that contrary to common engineering terminology the sine and cosine expansion is not *the* Fourier series but *a* Fourier series. There are many types of Fourier series.

What this means in audio terminology is as follows. If we have any situation involving exchange as well as dissipation of energy, and if the total available energy is limited, then the equations of analysis may be expressed in a form belonging to a Hilbert space. When this is done, we will find that any function we express in this space may be expanded as a linear sum of simple terms, and this will be a generalized Fourier series. The degree of correlation between any two functions in this space may be obtained from the inner product of the two functions and will generally be a complex number. To be consistent with concepts of statistical theory, the highest numerical value of correlation among elements is traditionally considered unity. Circuit theory correlation is therefore frequently expressed as the inner product divided by the norm.

The mathematically complete expressions we have used in audio engineering all along, such as room normal modes and frequency and impulse response, are each expressed in a Hilbert space. From the generalities of the concepts just given there is no reason at all to expect that only two representations, time and frequency, should be possible. We should be able to map our relationships onto other Hilbert space representations.

A set of finite measure in any number of dimensions may be mapped to and from a linear segment with preservation of measure. Audio and acoustic problems may be expressed on a line, a plane, a cube, or any number of dimensions. The coordinates of these expressions must generally be different since it is not proper to glue together two valid one-dimensional parameters to simulate two dimensions. The appropriate plane must be assembled from a mapping process, not gluing, to maintain validity under close scrutiny.

Once an engineer associates a physical process with any function space coordinate, the interpretation of other coordinates of measure in that and other function spaces is established. To an engineer, most of these other function space coordinates may seem unusual, but a meaningful characterization is assured with their use. It is not inconceivable that once an engineer becomes accustomed to two or more dimensions, the one-dimensional description of impulse response and steady-state frequency response will seem awkward and contrived.

HILBERT SPACE REPRESENTATIONS

The engineering significance of Hilbert space representations may be summarized as follows. First, all functional representations will be linear sums of simple terms of the same type. Furthermore, each term in this summation will belong to an orthonormal set which means that two separate functional representations will be completely separable, no matter how much they appear to overlap in the frequency or time domain from which they were mapped. Second, even if the true representation is a long series of terms we get the best root mean square approximation if we use a limited sum of terms so long as we pick out the largest values. This best approximation on a power basis is a result of a sphere-packing interpretation of a limit in the mean approximation of a truncated series. Third, the square of the norm is interpretable as proportional to total signal energy. In fact, the average energy is proportional to the square of the signal point from the origin of the space, regardless of the number of dimensions (number of degrees of freedom or of messages). Fourth, the inner product not only establishes the norm but provides a measure of the degree of correlation between any two functional representations in this space. Fifth, the establishment of a finite square integrable space, of class L^2, assures dealing with practical problems of bounded energy.

Classic Fourier transforms may be applied to spaces other than class L^2; however, certain significant benefits accrue from dealing in a Hilbert space. For example, a Fourier transform on L^2 is a one-to-one norm preserving linear transformation (Parseval equation) of L^2 onto L^2. It preserves inner products, thus is unitary. In this regard a Fourier transform defines an isomorphism of the Hilbert space onto itself. A very significant fact we shall use is that a function and its Fourier transform play exactly the same role in L^2. This property was also stated and used in an earlier paper [2] to infer the existence of time-delay distortion. This valuable symmetry of purpose is shared among select spaces in higher dimension. Its engineering use is that a solution obtained for any space may be applied in that same form to other spaces of the same number of dimension, thus saving a great deal of effort and sometimes providing deeper insight into a physical process so represented.

GENERALIZED FUNCTIONS

With the introduction of more general tools of analysis, we are in a position to clarify some mathematical terminology which engineers are now using. The integral form of the topological property of the inner product is of particular value in system analysis. A continuous linear functional (f, ϕ) in a linear space Φ formed from the functions $\phi(x)$ defined in some set R, is called a generalized function [29], [30]. For every particularly well-behaved test function $\phi(x)$ we can associate a number (f, ϕ). Generalized functions, also called distributions, are not functions in the usual sense but are defined in terms of an inner product relation with test functions.

Where the topological basis is not considered, it has become conventional terminology to call the continuous linear functional f the generalized function associated with the number (f, ϕ) which is defined in terms of the integral with infinite limits. In this sense any regular sequence of good functions f_N which tend to the same inner product for the same test function as $N \rightarrow \infty$ are said to define the same generalized function [14]. This is sometimes expressed as the definition that a generalized function is that class of all regular sequences of good functions which in the limit yield the same infinite integral for the same test function.

Generalized functions are valuable in audio analysis because under certain conditions the integral formulation, along with the operations of translation and convolution, provide reasonable analogies for relations involving system response to a stimulus. The impulse $\delta(x)$ and what this author called the doublet $d(x)$ [4] are generalized functions in the sense that a test function $f(x)$ can be found such that the "impulse" response of a system at any moment x_0 following the defined zero time may be obtained as

$$f(x_0) = (\delta(x_0 - x), f(x)) \text{ for } x_0 > 0.$$

The doublet is defined as producing the value

$$g(x_0) = (\mathrm{d}(x_0 - x), f(x)) \quad \text{for } x_0 > 0.$$

The reason for this author's particular choice of words in defining the impulse and doublet should now be evident in view of the inner product relation. The intent of providing two generalized function stimuli prior to the limiting process, rather than simply define the impulse response then obtain its Hilbert transform as might be conventional, was to introduce an energy functional for more complete analysis of loudspeaker signal arrival times, yet avoid a problem of noncausality of the Hilbert transform.

The function expressed as Eq. (8) in Part I belongs to that sequence which approaches the generalized function corresponding to the impulse as the parameter a is allowed to approach zero. Because this generalized function is the mapping to the delay space of a frequency space pole, the limiting process as a approaches zero is equivalent to the frequency space pole with positive damping approaching the complex frequency axis indefinitely closely and corresponds to a network manifestation of infinite Q. In the delay plane this limiting curve approaches a generalized function which can be considered to correspond to the full line $\tau = 0$ and the half-line $p = \omega_0$ for $\tau > 0$. The inner product expressed in the delay space then is such that any test function with pitch and time-delay dependence is operated on to produce the pitch output ω_0 for all positive time delay and no output prior to zero (causality). As one might expect, a frequency plane pole in the right half-plane with negative damping approaching the complex frequency axis corresponds to a noncausal behavior.

DERIVATION OF PITCH

In the subjective interpretation of sound, pitch is a "now" thing. There is strong reason to associate pitch with the intuitive concept of instantaneous frequency, but the conventional mathematical form cannot be used for the reasons which have been cited in this paper. It is necessary to propose a different approach such as symbolized in Fig. 10.

Define a functional subspace T_1 such that only events from the indefinite past up to, but not including, the immediate present are represented. This space is a museum of past events, and that portion of the negative real time line used in this space is a hallway down which one can view the past.

Define a second functional subspace T_2 which extends from the same indefinite past as T_1 but includes the immediate present. Subspace T_2 is slightly larger than T_1 by the difference between the events of the immediate present and the immediate past in the subjective observer's time frame. The excess left after the process of mapping T_1 into T_2 represents the functional change caused by an increment in time at the moment "now."

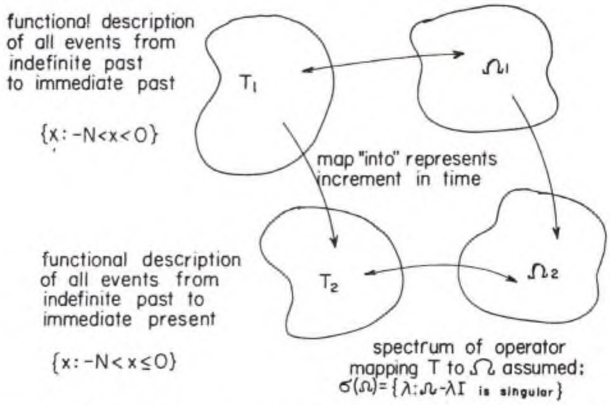


Fig. 10. Symbolic representation which allows an interpretation of the progressive flow of time as a continual process of mapping from one function subspace into a larger subspace.

Because the subspace T_1 stands complete, there is another conjugate subspace Ω_1 onto which the functional form in T_1 may be mapped by a process similar to Fourier transformation, and the same is true for T_2. In the limit as the indefinite past is greatly increased the process approaches Fourier transformation, as may be seen by considering the process from the standpoint of an observer who is always halfway back to the indefinite past and to whom our negative real line becomes his effective entire real line.

Mapping T_1 into T_2 to determine the functions in T_2 not found in T_1 is equivalent to the change brought about by an increment in time, if the functions defined in T_2 represent time functions. Mapping Ω_1 into Ω_2 may therefore yield the spectrum change brought about by such a time increment. If the function in T_1 and T_2 is a unit cisoid as shown in Fig. 11, then the distribution in Ω_1 and Ω_2 will be impulsive, that is, defined at a single coordinate. Remember, the subspaces are defined complete and the observer's future does not exist in them even as a null set. The observer's "now" is that subspace's eternity. Subtracting Ω_1 from Ω_2 to determine what has been added by an advance in time must be arbitrarily close to an impulse because if it were not, then Ω_2 would contain other coordinate components, which it does not.

If Ω_1 and Ω_2 have any arbitrarily, but identical, distribution for that contribution prior to the immediate past, these will nullify in subtracting Ω_1 from Ω_2. It is only the functional dependence in the finite interval between the immediate past and immediate present, Fig. 11, which determines the spectrum components of Ω_2 not found in Ω_1. No matter what functional form exists in T_1, if the additional component in T_2 has the form of a cisoid "spring" in this finite span, then the contribution in Ω_2 may be considered to be single valued at the component with that angular pitch value.

The single value of coordinate induced in Ω_2 by the increment in time between T_1 and T_2 will be called pitch and is given as follows.

If the functional time dependence is an analytic signal [4, p. 739],

$$h(t) = f(t) + ig(t) = e^{a(t)}e^{ib(t)}$$

then the pitch at any moment t_0 will be

$$p = \frac{d}{dt}[b(t)]\bigg|_{t=t_0}$$

if $a(t)$ remains constant for a finite interval prior to t_0.

It is important to note that the process of subspace mapping requires that the amplitude not only remain constant but do so for a finite prior interval in order for a single value of pitch to be defined in this manner. This is the major distinction between this entity and what has been called instantaneous frequency [15, p. 81].

The name pitch has been given to this coordinate because of its obvious tie with that subjective sound property. Also by a fortuitous accident of words it is a measure of the pitch of the cisoid when considered as a spring wound around the time axis. Another name for this parameter which is particularly fitting is "periodicity," since this conveys the recurrence of events per unit of the appropriate coordinate in much the same way as the term "frequency" is so identified. Pitch will be used when subjective descriptions are anticipated. No confusion should result from this.

The use of functional subspaces and the conception of mapping into another larger subspace in such a manner as to represent a progressive flow of time can be continued indefinitely. The manner in which the phasor of Fig. 11 changes as time advances can form an intuitively acceptable model of the sound process as perceived by an observer if it is recognized that the axes of the "immediate present" plane on which this is shown in Fig. 11 are analogous to entities derived from kinetic energy density and potential energy density [4, p. 739].

The length of the phasor may be made analogous to total energy density. The vector length will increase as the sound intensity increases. Pitch is then analogous to the rotational exchange rate between energy density partitioning when the intensity is constant. When the intensity varies Fourier's theorem can be used to show that this is equivalent to a summation of constant-length phasors of varying pitch.

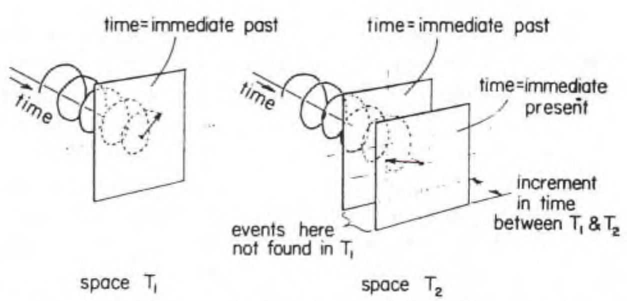


Fig. 11. Symbolic representation for describing pitch as a phasoid property dependent only upon events at the moment of perception.

It is a demonstrable fact that an oscilloscope presentation of the plane marked "immediate present" in Fig. 11 may be made by displaying the intercepted sound process on conjugate axes such that one axis is the microphone output and the other axis is its real-time Hilbert transform. The cathode-ray-tube spot will then trace out the signal that one would observe in Fig. 11 if he continuously moved along with the immediate present as the mapping process continued.

Thus an apparent mathematical abstraction can in fact be used as a displayable measurement tool for audio. Its engineering interpretation is beyond the scope of this paper since it represents the program signal rather than the network response.

GENERALIZED ALL-PASS FUNCTION RELATIONS

The following principles may be used to develop the orthonormal frequency-space expansions and, from them, the delay-space relations.

Principle 1

Any function expressible as a rational polynomial fraction

$$F(s)=\frac{N(s)}{D(s)}=\frac{N(s)}{(s-\gamma_1)^{m_1}(s-\gamma_2)^{m_2}\cdots(s-\gamma_n)^{m_n}}$$
$$=\frac{s^p+C_1s^{p-1}+\cdots+C_p}{s^q+D_1s^{q-1}+\cdots+D_q},\qquad p\leq q$$

may be expanded as a linear sum of generalized all-pass terms as

$$F(s)=C_{00}+\sum_{j=1}^{m_k}\sum_{k=1}^{n}C_{jk}\left(\frac{s+\gamma_k^*}{s-\gamma_k}\right)^{m_k-j+1}$$

where

$s=\sigma+i\omega$	complex frequency
$\gamma_k=\alpha_k+i\beta_k$	poles of $F(s)$
$\gamma_k^*=\alpha_k-i\beta_k$	complex conjugate of γ_k
n	number of poles
m_k	multiplicity of kth pole
C_{jk}	a frequency-independent coefficient and may be complex
C_{00}	a constant and may be complex.

Proof

$p<q$, $F(s)$ may be expanded into partial fractions

$$F(s)=C_0+\frac{C_1}{(s-\gamma_1)^{m_1}}+\frac{C_2}{(s-\gamma_1)^{m_1-1}}+\cdots$$
$$+\frac{C_3}{(s-\gamma_2)^{m_2}}+\frac{C_4}{(s-\gamma_2)^{m_2-1}}+\cdots.$$

Any partial fraction may be expressed in all-pass terms since, for first-order poles,

$$\frac{1}{(s-\gamma)}=\frac{1}{\gamma+\gamma^*}\left\{\frac{s+\gamma^*}{s-\gamma}-1\right\}.$$

Higher order poles may be similarly expressed in all-pass terms by expanding powers of both sides of this expression. Thus each partial fraction term may be expanded as a sum of all-pass terms in descending integral powers from the order of the pole yielding the partial fraction term.

$p=q$, $F(s)$ can be considered formed from some $G(s)$ with numerator coefficient $(q-1)$ as the product

$$F(s)=(s-\delta)\cdot G(s).$$

Since $G(s)$ can be expanded in partial fractions, a term-by-term multiplication by $(s-\delta)$, thus forming $F(s)$, may be expressed in all-pass terms since

$$\frac{(s-\delta)}{(s-\gamma)}=1+(\gamma-\delta)\frac{1}{(s-\gamma)}.$$

Principle 2

The generalized all-pass function

$$\phi(\omega,a)=\left(\frac{i\omega+a^*}{i\omega-a}\right)$$

can be made orthonormal under the conditions

$$\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T}\phi(\omega,a)\cdot\phi^*(\omega,b)\,d\omega=\begin{cases}0, & a\neq b\\ 1, & a=b\end{cases}$$

and

$$\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T}\{\phi^m(\omega,a)\}\cdot\{\phi^n(\omega,a)\}^*\,d\omega=\begin{cases}0, & m\neq n\\ 1, & m=n.\end{cases}$$

Proof

If $a=b$, then

$$\int_{-T}^{T}\phi\phi^*\,d\omega=\int_{-T}^{T}|\phi|^2\,d\omega=\int_{-T}^{T}d\omega$$

and

$$\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T}d\omega=1.$$

If $a\neq b$, then

$$\int_{-T}^{T}\left(\frac{i\omega+a^*}{i\omega-a}\right)\left(\frac{i\omega+b^*}{i\omega-b}\right)^*d\omega$$
$$=\int_{-T}^{T}\left(\frac{i\omega+a^*}{i\omega-a}\right)\left(\frac{-i\omega+b}{-i\omega-b^*}\right)d\omega$$
$$=\int_{-T}^{T}\left(\frac{i\omega+a^*}{i\omega-a}\right)\left(\frac{i\omega-b}{i\omega+b^*}\right)d\omega$$

which can be evaluated as a contour integral along the ω axis and closed with a semicircle of radius T. Each half-plane about the ω axis has a singularity with a finite residue. Hence the limit of the product of the finite contour integral and the vanishing term in T is zero.

For lattices of the same argument, but multiplicity greater than one,

$$\int_{-T}^{T}\left(\frac{i\omega+a^*}{i\omega-a}\right)^m\left(\frac{-i\omega+a}{-i\omega-a^*}\right)^n d\omega$$
$$=\int_{-T}^{T}\left(\frac{i\omega+a^*}{i\omega-a}\right)^{m-n}d\omega.$$

If $m=n$,

$$\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T}d\omega=1.$$

If $m \neq n$, one half-plane of the integrand is free of singularities; hence by Cauchy's theorem the contour integral is identically zero.

Principle 3

The generalized all-pass function is orthonormal under a relative time shift such that if

$$\Phi_\mu = \phi e^{-i\omega\mu}$$

where μ is a time delay relative to a fixed reference. Then,

$$\lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} \Phi_\mu \Phi_\xi^* \, d\omega = \begin{cases} 0, & \mu \neq \xi \\ 1, & \mu = \xi. \end{cases}$$

Proof

If $\mu \neq \xi$, the integral becomes

$$\lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} e^{i\omega(\xi-\mu)} \, d\omega = \lim_{T\to\infty} \frac{\sin T(\xi-\mu)}{T(\xi-\mu)} = 0.$$

If $\mu = \xi$,

$$\lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} e^{i\omega\cdot 0} \, d\omega = 1.$$

Note that since the generalized all-pass function can be used to make any all-pass phasoid, then Principles 2 and 3 will hold for any such phasoid.

REFERENCES

[21] G. F. Simmons, *Topology and Modern Analysis* (McGraw-Hill, New York, 1963)

[22] R. G. Douglas, *Banach Algebra Techniques in Operator Theory* (Academic Press, New York, 1972).

[23] E. C. Titchmarsh, *Introduction to the Theory of Fourier Integrals* (Oxford University Press, London, 1948).

[24] R. R. Goldberg, *Fourier Transforms* (Cambridge University Press, Cambridge, England, 1965).

[25] W. F. Donoghue, Jr., *Distributions and Fourier Transforms* (Academic Press, New York, 1969).

[26] N. Wiener, *The Fourier Integral and Certain of Its Applications* (Dover, New York, 1958).

[27] N. Wiener, *Extrapolation, Interpolation, and Smoothing of Stationary Time Series* (M.I.T. Press, Cambridge, Mass., 1949).

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

[29] I. M. Gel'fand and G. E. Shilov, *Generalized Functions*, vol. 1: *Properties and Operations* (Academic Press, New York, 1964).

[30] I. M. Gel'fand and G. E. Shilov, *Generalized Functions*, vol. 2: *Spaces of Fundamental Generalized Functions* (Academic Press, New York, 1968).

Reprinted from *JAES*, vol. 22, no. 2, pp. 104–106, 1974 March.

COMMENTS ON "THE DELAY PLANE, OBJECTIVE ANALYSIS OF SUBJECTIVE PROPERTIES: PART I"

MICHAEL A. GERZON

Mathematical Institute, University of Oxford, Oxford, England

In the following, we note some mathematical errors in Heyser's paper [1] and give references to important previous work in this field.

In summary, Heyser's paper [1] mathematically explicitly suggests the expression of arbitrary signals as linear combinations of "all-pass" signals, i.e., those whose Fourier transform has unity modulus, and in addition suggests that it is useful, when describing subjective properties, to draw the graphs of group delay versus frequency of these separate all-pass components, giving this graphical representation the name of "the delay plane." These suggestions may well prove very useful, but the mathematical means by which these representations were derived in [1] are largely faulty.

In particular, the "generalized all-pass function" $\phi(\omega,\gamma)$ of [1], defined by

$$\phi(\omega,\gamma) = (i\omega + \gamma^*)/(i\omega - \gamma)$$

and considered as a function of ω, does not satisfy the assertion preceding and including Heyser's eq. (3) [1] that the set of functions defined by different values of the complex parameter γ form an "orthonormal set," or that any reasonably well behaved function can be expanded as a linear combination of generalized all-pass functions.

The reasons for the falsity of these assertions are twofold. Firstly, the value of the "average over the real line"

$$\lim_{T\to\infty} \frac{1}{2T}\int_{-T}^{T} \phi(\omega,\gamma_1)\,\phi^*(\omega,\gamma_2)\,d\omega$$

is 1 for all parameters γ_1 and γ_2, and not 0 as it should be for an orthonormal set whenever $\gamma_1 \neq \gamma_2$ [2]. This may be proved simply by noting that the functions $\phi(\omega,\gamma)$ differ from 1 by less than ϵ, for any given small number ϵ, only over a finite range of values. The unity value of the above average contradicts Eq. (3) of [1].

The second reason for the $\phi(\omega,\gamma)$ not being an orthonormal set is that they are not members of an inner product space [2], since the axiom of strict positivity of inner product spaces is disobeyed by the counterexample

$$\lim_{T\to\infty} \frac{1}{2T}\int_{-T}^{T} |\,\phi(\omega,\gamma_1) - \phi(\omega,\gamma_2)\,|^2\,d\omega = 0.$$

Since the $\phi(\omega,\gamma)$ do not belong to a Hilbert space, and since this is the only new use of Hilbert spaces in [1], introduction of such concepts in [1] seems unmerited.

It should be noted that two quite different representations (Fig. 2*c* and *d*) in [1] are given the same name, i.e., "the delay plane," since identifying them might be very confusing.

Not only are arbitrary functions not expandable as a sum of generalized all-pass functions, but also some functions whose Fourier–Laplace transform is a polynomial rational, possessing a finite number of zeros and poles, are also not so expressible. This includes the effect of some LCR networks. If all poles are simple, by standard algebraic arguments there is a unique such expansion, but no such expansion is possible for functions with multiple poles. This may be seen by using Table II of [1] to derive the expansion in such cases; then the table gives infinite coefficients.

As observed in [1], the representation of a signal as a linear combination of "all-pass" signals is not unique in the sense that one signal can have many such representations. A representation of signals as the sum of Laguerre functions, which are indeed orthogonal "all-pass" signals, has been used since the early 1950s by Lee and Wiener [3]–[7] to analyze linear circuits and signals, and also to analyze general time-invariant nonhysteretic *nonlinear* processes, using the theory of Laguerre–Hermite expansions [5]–[7], which are easily instrumentable in terms of practical circuitry.

In [1] Heyser points to several unsatisfactory attempts to represent a signal on a domain parameterized simultaneously both by time and frequency. As has been pointed out by De Bruijn [8], such a representation has in fact been known since 1932 in quantum theory under the name of the Wigner distribution [9]–[12]. Translation into the language of communications theory is effected by renaming "position" time, "momentum" frequency, and (where necessary) putting Planck's constant $\hbar$ equal to 1. Unfortunately, the majority of useful results about Wigner distributions seem to be published only in the form of technical details within, or appendices to, papers assuming an extensive theoretical knowledge of specific technical areas within quantum theory, or in papers requiring a knowledge of functional analysis, e.g. [12].

The Wigner distribution of a signal $f(t)$ is defined as that function $W_f(t,\omega)$ of frequency ω and time t defined by

$$W_f(t,\omega) = (2\pi)^{-1/2}\int_{-\infty}^{\infty} f(t + \tfrac{1}{2}s)f^*(t - \tfrac{1}{2}s)e^{-i\omega s}\,ds.$$

This has many of the properties [8] to be expected of "the energy of the signal f at time t and frequency ω" except that unfortunately it is not positive. Closely related transforms which are always positive appear in the quantum literature, notably the Husimi transform [13]–[14] defined as the Wigner distribution of the signal f two-dimensionally convoluted with the Wigner distribution of a Gaussian function.

In general, the literature of quantum theory contains a great fund of analytic and computational techniques suitable for use in audio, as does the mathematical functional analysis literature. However, intending users of these techniques should check both that they are used correctly, and that the concrete system under consideration actually justifies the introduction of such techniques. There seems little point in using "advanced" techniques when elementary ones suffice.

REFERENCES

[1] R. C. Heyser, "The Delay Plane, Objective Analysis of Subjective Properties: Part I" *J. Audio Eng. Soc.*, vol. 21, pp. 690–701 (Nov. 1973).

[2] For example, see P. R. Halmos *An Introduction to Hilbert Space and the Theory of Spectral Multiplicity* (Chelsea, New York, 1957), ch. 1.

[3] T. L. Saaty and J. Bram, *Nonlinear Mathematics* (McGraw-Hill, New York, 1964), pp. 338, 339.

[4] Y. W. Lee, *Statistical Theory of Communication* (Wiley, New York, 1960), ch. 18 and 19.

[5] N. Wiener, *Nonlinear Problems in Random Theory* (M.I.T. Press, Cambridge, Mass., 1958), (paperback 1966), pp. 88–100.

[6] Y. W. Lee, "Contributions of Norbert Wiener to Linear Theory and Nonlinear Theory in Engineering," in *Selected Papers of Norbert Wiener* (M.I.T. Press, Cambridge, Mass., 1964), pp. 23–33.

[7] P. Masani, "Wiener's Contributions to Generalized Harmonic Analysis, Prediction Theory, and Filter Theory," *Bull. Amer. Math. Soc.*, vol. 72, pt. II, pp. 112–121 (1966).

[8] N. G. De Bruijn, "Wigner-Distributie als Muzikaal Notenschrift bij de Fourier-Analyse van Signalen," *Kon. Ned. Akad. Wet.*, vol. 76, pp. 19–23 (1967).

[9] E. Wigner, "On the quantum-correction for thermodynamic equilibrium," *Phys. Rev.*, vol. 40, pp. 749–759 (1932).

[10] J. E. Moyal, "Quantum Mechanics as a Statistical Theory," *Proc. Camb. Phil. Soc.*, vol. 45, pp. 99–132 (1949).

[11] E. C. G. Sudarshan, "Structure of Dynamical Theories," in *Lectures in Theoretical Physics, Brandeis Summer Institute 1961*, (Benjamin, New York, 1962), pp. 143–199.

[12] J. C. T. Pool, "Mathematical Aspects of the Weyl Correspondence," *J. Math. Phys.*, vol. 7, pp. 66–75 (1966).

[13] J. McKenna and H. L. Frisch, "Quantum Mechanical Microscopic Brownian Motion," *Phys. Rev.*, vol. 145, pp. 93–110 (1966), esp. sec. III and app. I.

[14] K. Husimi, *Proc. Phys. Math. Soc. Japan*, vol. (3)22, pp. 264–314 (1940).

Reply by Mr. Heyser

I must admit you are absolutely correct on your initial point. The generalized all-pass function is not orthonormal in the manner I indicated and thus is not a member of an inner product space. How did this terrible error occur? In going back over my notes in the derivation of Principle 2, I find an error in evaluation of that portion of the contour integral on the semicircular arc. It does not grow at a rate less than T.

Those familiar with my work realize that my publications are usually based on experimental measurement, and the delay plane is no exception. The expansions of Fig. 6 can be made as outputs of a TDS. They are observables, not abstractions. I have already succeeded in isolating certain members of that set and determining the coefficients by precisely the procedure of the now deflowered Eq. (3). Why, then, do I get the correct answer. In puzzling over this I find that I have actually used Principle 3, relating to a time shift. In order to make it easier on myself, I mapped the curves corresponding to, for example, C_2 in Fig. 6 to the straight line corresponding to C_0. Because they are never tangent, even with this mapping, curves belonging to different arrival times, such as (c) simply dropped out of the measurement.

I chose the unusual artifice of the averaging operator in looking for orthogonality not only in an attempt to keep the representation bounded in the infinite integral without a change in variable, but because it worked—at least in the practical examples I tried. The failure of orthogonality of the specific all-pass functions I introduced does not change any of the other assertions of the paper, since there are other all-pass related functions known to be orthogonal. Nor does lack of orthogonality under the averaging operator detract from the usefulness of the generalized all-pass functions to circuit analysis and subjective interpretability.

We as engineers need to get down off our high horse and recognize we had better look at other coordinates of measurement if we want to come to terms with the subjectively-oriented user of our services. *That* is why I introduced the concepts of linear spaces, not because of a specific all-pass expansion. I would shout the message if I felt that to be required in order for us to break out of our rut.

You raise some other points which I would like to address. The delay plane is, as its name implies, a two-dimensional surface that an engineer may represent by a piece of paper. Expansions in that plane may in turn be represented as lines drawn on the piece of paper. I did not anticipate that this would cause any confusion.

I did not state that arbitrary functions are expandable as a sum of generalized all-pass functions. I did state that any reasonably well-behaved frequency transfer function can be so expanded, and supported it with Principle 1. Obviously essential singularities and branch points require different treatment as I indicated in my 1969 paper [1]. Your statement that no such expansion is possible for functions with multiple poles, verified by the observation of infinite coefficients is, I must point out, incorrect. First, I specifically stated simple poles in reference to Eq. (3). Second, one accustomed to the use of conventional Laplace or Fourier tables would not be alarmed that infinities would occur for improper cascading of different simple pole expressions to obtain multiple poles. Thus Gardner and Barnes [2] entry 1.105 and Campbell and Foster [3] entry 448 yield infinities under similar circumstance. It is incumbent upon the user of any such table to exercise a semblance of caution and simplify his expressions to a form suitable for table use. Third, I derived in Part II the form to be used for multiple poles. There is no difficulty—it works.

The all-pass representations of Wiener and Lee, which are not Laguerre functions, as you state, but their transform and which are not strictly all-pass, since the numerator is of lower power than the denominator, date to a much earlier period than the 50's. Lee's 1930 dissertation, presented in open literature a few years later [4] is, in my opinion, a landmark paper since it not only introduces such functions to electrical engineering, but is the first concise application of Hilbert transform relations governing network amplitude and phase relations which appeared in engineering literature. Kronig and Cramers dispersion relations had been known prior to that but were apparently not applied to engineering. I used my reference [15] on transversal filters *specifically* because it explains the particular all-pass related expansions of Lee. Please observe, as I stated in the paper, that I am introducing parallel networks, not series networks as in the work of Lee. The difference is not trivial when it comes to implementation or relating sound to subjective concepts.

I am grateful for your pointing out the Wigner distribution, since I was not aware of it, nor can I find it in any of the references I used in compiling my paper. I was aware of a very similar expression formulated by Ville [5] then introduced by Woodward [6] to radar theory. I gather from the lack of references to Wigner's work that it was also not known to Gabor at the time he introduced the concept of logons, nor did Page seem to be aware of it when he tackled the same problem. The problem of negative energy of the modified Wigner distribution which is remedied by a smoothing function appears to support the statement that simultaneous parameterization in terms of Fourier transform coordinates is never exact. And isn't that what I pointed out?

The formalism of quantum mechanics is indeed rich in concepts of value to audio, but unfortunately most audio engineers may not be aware of the notation and how it can be applied to their problem. Hence my attempt at bridging the semantic gap by using audio related terminology. I personally prefer the view that one should reach for other tools when he does not get results he seeks with his present tools, no matter how comfortable they feel in use. I thought I had made that point clear.

REFERENCES

[1] R. C. Heyser, "Loudspeaker Phase Characteristics and Time Delay Distortion," *J. Audio Eng. Soc.*, vol. 17, pp. 30-41, 130-137 (Jan. 1969, Mar. 1969).

[2] M. Gardner and J. Barnes, *Transients in Linear Systems Studied by the Laplace Transformation,* (John Wiley, N.Y., 1942).

[3] G. Campbell and R. Foster, *Fourier Integrals for Practical Applications* (D. Van Nostrand, N.Y., 1948).

[4] Y. W. Lee, "Synthesis of Electrical Networks by Means of the Fourier Transforms of Laguerre's Functions", *Jour. Math. & Physics,* M. I. T., 11, pp. 83-113 (1932).

[5] J. Ville, "Theory and Application of the Notion of the Complex Signal", *Cables et Transmission,* vol. 2, pp. 67-74 (1948).

[6] P. M. Woodward, *Probability and Information Theory with Applications to Radar,* (Permagon Press, N.Y., 1953).

Reprinted from *JAES*, vol. 22, no. 9, pp. 674–682, 1974 November.

Geometrical Considerations of Subjective Audio

RICHARD C. HEYSER

Jet Propulsion Laboratory, Pasadena, Calif.

A general approach to audio processing is presented in terms which consider each step in the processing of a signal as a mapping between functional representations. A general result of this analysis is verification that what we hear is not necessarily what we now measure in audio systems. Translating the analytical terminology into practical terms reveals among other things that there is a subjective difference between harmonic distortion in an amplifier and harmonic distortion in a loudspeaker. Several other apparent paradoxes of subjective audio are shown not only to be correct as we hear them but predictable by this more general objective approach.

INTRODUCTION: One of the worst kept secrets in audio engineering is that what we hear does not always correlate with what we measure. This is particularly true for distortion. The lament, "the meter says it's good, but my ear says it's bad," is recognized as valid by those close to the problem [10], [11].

Until quite recently, however, there has simply been no plausible explanation for why we hear things that either do not appear to be part of our mathematical analysis or are not displayed on measuring devices which are a product of that analysis.

Those whose principal professional involvement is based on the listening experience tend to develop a subjective viewpoint with value judgements seldom related to instrumental measurement. Their own good "ears" are the tools on which they depend, although many would use instruments if such devices could measure what they hear.

Outwardly a more formal and orderly world is available to those who depend upon objective measurement. Such people tend to believe that if there is something which is heard but not describable within the framework of objective formalism, then it must be due to some property of sound perception that sets a limit beyond which objective techniques may not penetrate.

The importance of attempting to bring objective and subjective audio together lies in the fact that the end product which both strive for is an acoustic illusion which is economically acceptable. The audio product is thus generally subjective in nature but demands a measure of predictability.

In a previous paper [1] a method of linear analysis was introduced with the specific purpose of bringing together one aspect of subjective and objective audio. However, much of the discrepancy between how we measure and how we hear lies in nonlinear effects. The purpose of this paper is to investigate the effect of distortion and to try to understand not only why the "ear" and "meter" do not agree, but those circumstances in which the ear might rank systems in a different order than the meter.

PRACTICAL OBSERVATIONS

Discussions on distortion soon strike a responsive nerve when practically minded professional recording people become involved. There can be no sanctuary in equations when a reasonable number of professionals can hear an effect not predicted from those equations and describe it to others. Rather than start directly with the abstract analysis which might discourage a practically minded person from reading further, let us instead start with some of the results translated into subjective terminology. Here are some observations an engineer can make which are consistent with and derived from the analytical approach to be presented. The measure of success of the analysis of this paper is marked by the degree to which these observations are consonant with practical experience.

1) It is reasonable to expect that subtle distortion differences between two or more amplifiers may be discerned when listening through a loudspeaker system which has a measurable distortion much higher than that of the amplifiers. A loudspeaker system driven at a 3-percent harmonic distortion level does not automatically hide the sonic differences between amplifiers with distortion figures several places to the right of the decimal point.

2) Subjectively a given percentage of distortion of any particular type may be more or less objectionable in a microphone than in the amplifier following that microphone. A laboratory standard microphone with excellent linearity may be a poor choice for recording studio use, even though the criterion of performance is realism, because the resultant reproduction through conventional loudspeakers may not be as realistic on a subjective basis.

3) The addition of distortion of one type to a reproducing system already corrupted by distortion of a second type may yield a more subjectively acceptable performance, even if the net distortion is higher. Under certain conditions there may exist preferred loudspeaker–amplifier combinations for more accurate realism of reproduction. Those for whom there is a difference between the "sound" of different loudspeaker types and construction material may observe a preference in the type of loudspeaker for realism of reproduction of some program material.

4) "Improving" a system of reproduction by partial cancellation within an amplifier of some distortion terms caused by the loudspeaker may result in a less acceptable reproduction. Steady-state frequency equalization of a distortion in response caused by time-delayed acoustic interference patterns may adversely affect instrumental timbre.

5) In loudspeaker reproduction the sound image of a closely miked vocal will not appear to come from a position in space a few inches from the listener's ear.

6) A reproducing system which measures better than another based on amplifier criteria may not necessarily sound more realistic.

These observations are the result of considering a topology of signal processing and did not derive from a shopping list of subjective observations which had to be explained. The basis for this topology was presented in the delay-plane paper [1], [2] which contains the background for this more specialized work. While these results are expressible in subjective language, the steps leading to them, which will be described shortly, are abstract in the sense that, until it is all brought together in the listening experience, the only guidance is in manipulation of symbols. While the symbolism may not have subjective identification, it does have the potential for measurement.

Because this abstract manipulation of symbols runs so contrary to the very subjectiveness we are trying to understand, it is now necessary to consider how we are led to such an approach and outline the steps to be taken in this paper.

GEOMETRY OF PERCEPTION

One result of the delay-plane paper was the demonstration that it is possible to consider a geometry of representation for subjective perception of sound. That view will be elaborated upon for this paper. The assumption will be made that the final result of audio processing is the listening experience. A representational geometry will be assigned to each step in the processing of what will ultimately be an acoustic listening experience. Each stage in the processing of a signal will be considered to be a transformation from one form to another. It all comes together in the listening experience, which is the only place where subjective human perception and a mathematical model must coincide.

From the time an acoustic signal is picked up by a microphone until the time a loudspeaker recreates that sound, it will be assumed that whatever the signal is, it is not subjectively recognizable. During this time the signal will be considered from the standpoint of a topology of processing. Three postulates will be assumed for the purpose of aligning the mathematical model with audio engineering. The way we will attempt to determine the effect of processing defects on subjective perception will be to inquire how that defect may propagate in form through various stages of processing until its final presentation in a geomerical representation coinciding with subjective perception. The only "proof" of such an approach will be translating the final mathematical terms into subjective language and seeing if it makes sense.

Underlying all of this is a fundamental assertion. That assertion is that subjective awareness of order and reasonableness is describable by mathematics. This does not mean that we solve equations in order to perceive, but that basic relationships in our natural environment, which we sensorially absorb into a "rightness" of perception, have analogs in mathematical symbolism. If we properly manipulate and interpret the mathematical symbols, we can have analogs of human perception.

That viewpoint was employed in the delay-plane paper by insisting that subjective perception was correct and that if standard mathematical models did not work, that was a failing of the model not of the perception. The approach taken in that paper was to lay the present mathematical model aside temporarily and start from the assumption that the way people hear and describe sounds is the correct form. Another quite different mathematical model was generated from this language of subjective perception. This different model required an introduction of what was at first glance a different branch of mathematics than that of the original model. However, a closer view of the new model showed that it actually embraced the original model as a special case. Thus the apparent paradox of mathematics that is correct but does not work was resolved by observing that the original equations (the ones most of us now use) are correct but simply are not in the proper form for subjective identification. A substantial portion of that paper was devoted to describing the newer mathematical tools and providing one way of expressing our original equations in a form more in alignment with subjective perception.

Now we want to inquire into nonlinear as well as linear effects in a way which absorbs both into a common description. We are now looking at general relationships as they can modify the "form" of final presentation. We are concerned with the forest, not the trees, and will therefore back away for a better perspective. No distinction will be made concerning what it is that is being analyzed. It could be pressing a disc, listening to a loudspeaker, or turning a knob. Processing will be considered from the standpoint of taking something, doing something to it, and thus producing another thing.

POSTULATES

All audio processes considered in whole or part have something that can be identified as an input and another thing that can be identified as an output. These can manifest themselves as sound itself, an electrical current, deformation of a groove wall, aggregation of magnetic particles, or any of a variety of things. The program content of input and output may be anything whatsoever—an impulse, sine wave, or a portion of Beethoven's Ninth Symphony.

In the broadest sense the processing of a signal may be thought of as the observation that something is done to the input in order to produce an output. In this paper it will be assumed that the appropriate mathematical representation of the input is mapped into the appropriate representation of the output. Thus the processing is represented as an operator. This leads to the following postulate:

Every dynamical process may be described by a corresponding operator. (A)

As examples of what this means, an amplifier corresponds to an operator which changes the total energy level without altering the coordinates of representation. On the other hand, a microphone is a complex operator which may not only change the energy but alters the form from acoustic pressure and velocity to electrical voltage and current. From the standpoint of the audio engineer, not all operators need be identified with specific devices. Thus the expectation operator is used to take the appropriate moment and the convolution operator specifies one form of scanning.

The second postulate relates to a description of input and output:

Each signal may be described as a figure[1] in an appropriate function space. (B)

As an example of what this means, the dynamic sound pressure at the location of a microphone diaphragm will be considered to be a figure with space coordinates representing both the direction of maximum energy flow and the partitioning of energy in terms of pitch, time delay, and intensity. Note that each possible program is considered a separate figure. As in the delay-plane paper, the dimensionality of a representation will refer to the number of independent coordinates of representation and not to the number of possible functional forms expressible in those coordinates. The reason for this is to avoid, at this time, considering infinite dimensional spaces. Quantum and information theoretic analysis traditionally consider a signal space inhabited by points representing all possible signals and with a dimensionality equal to the number of possible signals. In this work we depart from that convention in order to align the mathematical description of a signal with the subjective description normally used by a practical audio engineer. For this reason we keep to Euclidean spaces for much of this work but reserve the privilege of using other spaces when this is analytically simple and can be tied to subjective concepts.

We are now presented with the problem of describing what is meant by linear and nonlinear operators and are quickly faced with the dilemma that while linear analysis is well understood, nonlinear analysis has no such general agreement. We will therefore invoke traditional engineering concepts that operators such as those generating harmonic and intermodulation distortion will be considered nonlinear. Devices which alter the gain in a manner dependent upon signal level will also be considered nonlinear because of their dependence upon program dynamics.

In engineering terminology, the output of a nonlinear process consists of a nondistorted version, which would have resulted had the process been linear, plus distortion terms. A nonlinear process will produce separable independent replicas, one or more of which might change form dependent upon the energy level of the input. The output of a linear process thus consists of a single representation which cannot be split apart, in anything but a trivial manner, into separate superimposed versions of the input.

In this general view of the audio process we are considering much more than a time series representation. Our consideration of distortion must therefore allow for a description of more than nonpredictive operators invariant with respect to the translation group in time [3], [4]. The concern with linearity rests in the interpretation of the resultant process by a listener. The question of concern is, "does it sound realistic?" A major premise of this paper is that a more linear process leads to a more realistic performance. The word linear is therefore used here in its subjective context relating to realism of performance rather than in the strict mathematical sense.

The conventional mathematical definition [5], [6] that 1) a linear operator is additive and homogeneous (or that it fulfills additivity and continuity) and that 2) a nonlinear operator is an operator that is not linear, is not considered by this author to fit well with subjective concepts in audio due to the considerations of mapping between spaces of different dimensionality. An echo chamber is an operator that is additive and homogeneous, yet the substantial change in ambiance and subjective spatial modification it may introduce in a "dry" program is such that we can consider a distortion to have been introduced.

A postulate for defining nonlinearity for the purpose of this paper is the following:

A linear operator is one-to-one and a nonlinear operator is one-to-many. (C)

What this refers to can be understood from the simple example of a sine wave input. A linear operator will produce a single sine wave output at the same frequency. A

[1] The word *figure* is used here to designate any unified structure formed out of elements of the corresponding function space. The words *signal* and *figure* are used interchangeably because they mean the same thing in this representational geometry. The subjective sound image, or illusion of sonic presence, is the final form of this figure when we want to study subjective properties.

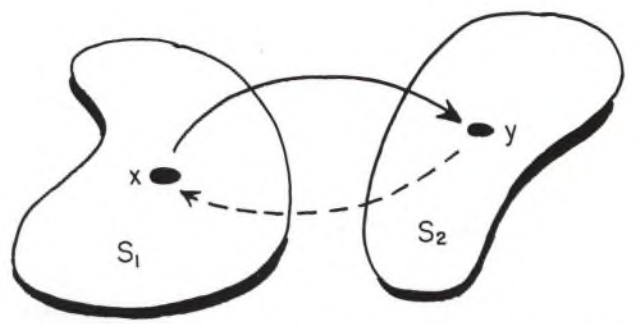


Fig. 1. Illustration of a linear operator as an invertible one-to-one map. A signal in one form is changed into only one possible signal in another form by the operation. The operator is shown as the arrow. The signal is a single entity shown as a point and the space of operation is shown as the amorphous regions.

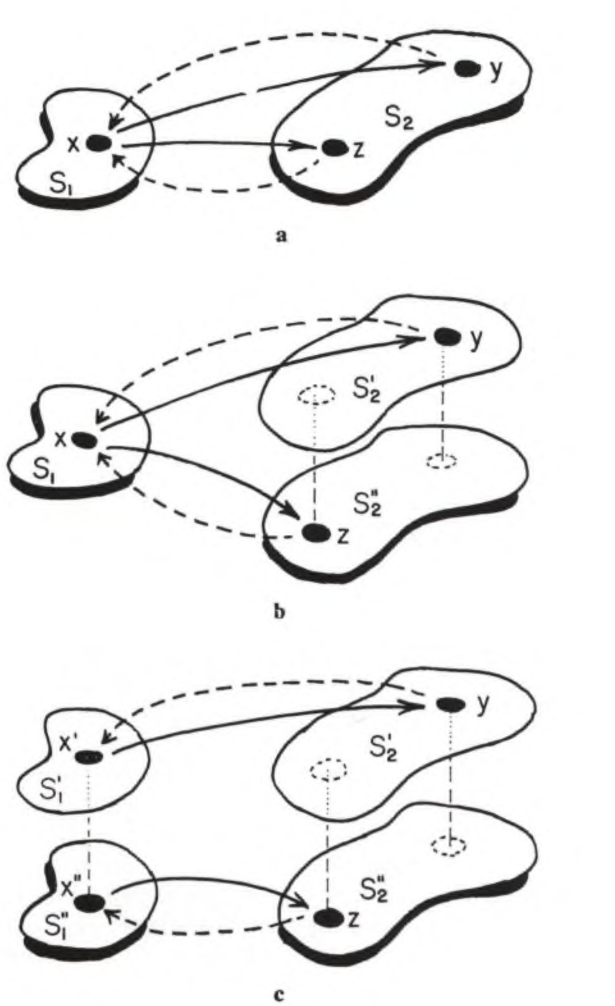


Fig. 2. Illustration of a type I nonlinearity as an invertible on-to-many map. **a.** A signal in one form is changed into two or more signals in another form by such a nonlinearity and one can always find another operation capable of undoing the process to get back to the original form of signal. **b.** This multiplicity can be visualized as separate versions, each one occupying its own position in its own space. **c.** In certain cases it is possible to consider this creation of a multiplicity of forms as due to linear maps from appropriate partitions of the original signal.

nonlinear operator will produce many sine wave outputs—the fundamental, which would have been the resultant of a linear operator, and harmonically related distortion products. Other examples spring readily to mind. Each of the "many" replicas of the input may be considered a valid "one" replica in its own space. A linear operator is singular (in the sense of oneness) whereas a nonlinear operator is multiple. We are allowed to this only because we treat the program as a single entity that is not composed of constituent parts.

There is another consideration that must be made before launching into a general analysis. If there is a second operator that can "undo" what a first operator does, then the second operator will be called the inverse of the first. Many types of distortion due to a nonlinear transfer characteristic can be eliminated by passing the signal through a second device with an inverse characteristic. This is a very old idea in audio engineering.

MAPPING BETWEEN SPACES

The following definitions can be formulated.

1) A bounded transformation between spaces of any dimensionality and which possesses a unique inverse will be defined as a linear operator if it is one to one in the sense that for every figure x in one space there is one and only one possible figure y in the other space, and if for any nonzero scalar α,

$$\alpha x \text{ transforms to } \alpha y$$

2) A bounded one-to-many invertible transformation will be defined as a nonlinear operator of type I.

3) A bounded one-to-many noninvertible transformation will be defined as a nonlinear operator of type II.

The clues here are the phrases one-to-one an invertible. If, as in Fig. 1, a figure x in space S_1 is uniquely mappable to and from any one figure y in space S_2, then the process is linear.

In engineering terms the figure x might be a particular signal input to an amplifier. What is symbolized as a point may be a signal corresponding to an entire symphony. The amplification process maps the input signal to an output signal. If there is only one possible output y for every input x, regardless of the total energy level of x, the amplifier is linear. If you could always design another device that produced the proper x from the proper y, the process is invertible. This is signified by the mapping arrows going both ways.

A type I, or retrievable, nonlinearity is represented in Fig. 2**a**. A signal x in space S_1 maps to two or more separately distinguishable signals in space S_2, at least one of which would have been the result of a linear operation, and these are uniquely mappable back to S_1 to become x. Under these circumstances we can consider the process as a superposition of two or more linear mappings to spaces such as S_2' and S_2'' as shown in Fig. 2**b**. Only one possible input signal is allowed with this definition of nonlinearity.

A type II nonlinearity is characterized in Fig. 3. In this case we have lost uniqueness if we attempt to recover the input signal.

Microphones in a studio pick up the sound of a performance. The outputs of the microphones are amplified, monitored through loudspeakers and metering devices, processed, equalized, and then recorded. The recording goes through a number of processes and changes of form before a successful release as a tape or disc. At some later time the recording is transferred back to electrical signals

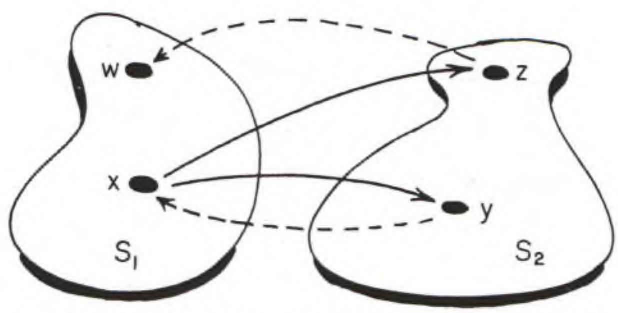


Fig. 3. Illustration of a type II nonlinearity as a noninvertible one-to-many map. The mapping operation is such that there cannot be found an inverse operation capable of uniquely recovering the original signal.

and then to acoustic signals to be heard through two or more loudspeakers.

What is suggested in this paper is that you consider each step of processing in this familiar chain of events as represented by a functional operator. A signal in one functional form is mapped to another functional form at each step. This mapping takes the signal from a given space of representation and maps it either onto another space with a different set of coordinates or maps it back onto other values in the same space. Aspects of this way of looking at a dynamic process are common in many branches of applied mathematics. All we are doing here is setting up a model for the processes used in audio engineering. The phrase audio engineering is meaningful because every action capable of altering what is ultimately to be a representation of a dynamic process may be considered an operator. Thus even a stamping process is an operator and vinyl imperfections or temperature cycle difficulties can create a type II nonlinearity. The function spaces of representation will be assumed to be of class L^2 because of practical interest in finite energy signals. Reference should be made to the previous paper [2] for an understanding of the significance of a Hilbert space representation in practical audio analysis.

The dimensionality assigned to any function space will depend upon the form in which the signal is monitored. This is not as mysterious as it may seem because all we are doing is modeling a physical process. If you monitor through a device such as an oscilloscope, the space is C^1. If you listen to it as though it were the original sound field, it is C^5. If you parameterize on a delay plane it is C^2, etc. Furthermore the process of monitoring or analyzing, a function space is itself a linear operator.

All devices traditionally considered linear by audio engineers, such as passive equalizers and active devices operated in a linear mode, are represented by linear operators when defined in the way we normally think of them. Devices we normally think of as having soft nonlinearities, such as compressors and moderately overdriven amplifiers, fall into the type I category, while clipping amplifiers and such irretrievable nonlinearities fall into the type II category.

One surprising result of these definitions is that a multipath medium, such as sound in a room, is a type I nonlinearity for aperiodic signals. A repeating time pattern of sound is the analog of the repeating frequency pattern of simple harmonic distortion. Just as we can think of the direct and reverberant sound in a room as different entities related by room geometry, so we can consider a nondistorted and distorted replica of a signal as due to a "geometry" of the associated dynamical processing. This result and similar equivalence partitions in higher dimensional representations may also be inferred from the isomorphism of L^2 spaces [2].

COMBINATION RULES

At this point some basic postulates of the gross aspects of audio engineering have been defined. Some combination rules that can be inferred from them are the following.

1) Two operations will be equivalent if they produce isomorphic figures in the same function space.

2) A nonlinear operation will spread the signal energy among the coordinates of final presentation in a manner different than that of a linear operator on the same signal.

3) Two operations which yield a similar coordinate distribution of energy may be considered equivalent if the agent of perception is incapable of distinguishing the detail differences of energy partitioning.

4) In comparing two operations in a resultant space of higher dimensionality, the operations may be ranked as equivalent on a basis of reduced dimensionality if the partitioning of energy on the reduced coordinate basis is congruent.

5) A comparison of two nonlinear operators to determine which is more nearly like a linear operator may not be valid unless the resultant spaces have the same dimensionality and are then ranked on the redistribution of energy among the coordinates.

6) A nonlinear process in a space of dimensionality M will not be equal in measure to a nonlinear process in a space of dimensionality N if the basis of comparison is made in N by assuming a dimensionality M for that space.

These are, of course, highly abstract. Practical examples illustrating each of these rules are as follows.

1) The inverse square attenuation of an acoustic signal passing through free air will sound the same as a gain reduction of an electrical signal version of that sound when you take into account the possible changes in acoustic signal shape due to passage through the air.

2) This is the central result of this particular way of describing nonlinearity. A distortionless representation of a dynamic signal will not modify the way in which the signal is represented among the appropriate coordinates. For example, the direction from which a sound appears to come should not be dependent upon the pitch components of that sound if the original sound had no such dependence. If one compares the result of a linear and a nonlinear operation on a basis of equal total energy, the nonlinear process must alter the coordinate representation of the signal. In our traditional measurements in a C^1 space, such as steady-state frequency response, this alteration amounts to a simple waveform change which may be difficult to judge for its subjective effect as seen on a spectrum analyzer. However, when we map onto a space more consonant with human subjective perception, such as a C^5 or even a C^2 space, then the redistribution of signal energy among the coordinates may be apparent. We may then find an alteration of time delay with pitch or angle change with intensity which is not present in the linear undistorted version.

3) If, under a given set of circumstances, you cannot hear the difference between two reproductions, then as far

as you are concerned they are the same.

4) If you measure a loudspeaker as though it were an amplifier, you have absorbed several coordinates of representation into a one-dimensional measurement. An amplifier which measures exactly the same as the loudspeaker on every distribution of energy to within a proportionality constant, must sound the same as the loudspeaker under listening conditions characterized by the measurement. This is because the loudspeaker and the amplifier have the same form in the reduced dimensionality of measurement, hence must have the same form in the higher dimensionality of listening. The fly in this ointment is the necessity to guarantee that the appropriate mapping operation is performed in transferring from one dimensionality to another. It is not sufficient, for example, to place a single microphone at a space position presumed to be concident with a listener's ear and measure the loudspeaker as though it were an amplifier. This is not a proper map because the azimuth, elevation, and range spatial coordinates of loudspeaker reproduction have not been properly specified. This is where microphone and loudspeaker nonlinear measurements most frequently go awry.

The detailed kinetic and potential energy density distributions, both spatial and spectral, of the sound field at the point of measurement must be specified and mapped to a complex C^1 form before a valid comparison can be made with an amplifier which is also measured C^1. Only then can they be ranked. You must properly measure both of them the same way before you can compare them.

Even if the amplifier has a simple transfer function nonlinearity, the sonic effect, when heard through headphones or loudspeaker, will be a redistribution of spatial as well as spectral and temporal energy because this is a mapping from a C^1 measurement to a C^5 experience. This is not the same effect in C^5 as a pure spectral distortion, which is generally what the simple loudspeaker distortion test measures. As a consequence the sonic effect of apparently simple amplifier distortion is much more complicated and probably more apparent than the same measure of loudspeaker distortion when measured in the conventional manner.

The requirement for topological congruence or similarity in energy partitioning in a reduced space can also offer a possible explanation for sonic effects relating to the apparent radial distance of a virtual sound source relative to a listener. The spatial energy density partitioning of a sound field due to a source is not the same in near-field conditions as it is in far-field conditions. The analytical relationship between total energy density and its kinetic and potential energy density partitioning was derived by this author in an earlier paper [17]. This relationship holds true for spatial coordinates as well as spectral or temporal coordinates. (It can be used, for example, to get the total energy density of sound in a room from spatial measurements of only the potential energy density measured by a pressure microphone.) When a vocal is close miked by a pressure-responsive microphone a map from C^5 to C^1 has taken place because the information is now a time-dependent voltage. There is no unique C^1 information to distinguish this from a distant miked situation, except cues of spectral distribution and reverberance, because only potential energy density at a point in space has been recorded. When mapped back to C^5 by loudspeaker reproduction, the vocal will therefore never appear to come from a point in space a few inches from the mixer's ear. It will instead come from a place much closer to the monitor loudspeaker. If the proposed combination rules are reasonably accurate, the close miked sound can be subjectively moved behind a monitor loudspeaker by adding reverberation, but it cannot be substantially moved forward. It would appear from consideration of the topology of reproduction that three-dimensional velocity as well as pressure fields must be recorded in order for a completely accurate reproduction to be possible.

Because an acoustic difference exists between the spherical wavefront of a local source and the more nearly plane wavefront of a distant source, combination rules 1) and 4) infer that there may be a subjective difference between discrete loudspeaker sources and large array distributed sources. Concavity or convexity of distributed sources may also subjectively be discerned by their modification of acoustic cues.

5) This is why two harmonic distortion measurements dependent only on the magnitude of the distortion component may not sound at all alike. The meter says they are the same, but the differences which may seem subtle in a one-dimensional measurement can have a devastating effect on energy partitioning in a higher dimensional representation, such as the one we use for listening. A little thought will show that cascading nonlinear operations, such as a distorting amplifier followed by a distorting loudspeaker, may cause a significant redistribution of coordinate energy over that of either nonlinear operator considered separately. It is even possible that the resultant C^5 representation may be more nearly accurate with such cascading than had either operator been more linear. Some amplifiers may sound better with some loudspeakers. Of course, this is true for any combination of operators, not just loudspeakers and amplifiers.

6) Harmonic distortion in a loudspeaker measured as though it were an amplifier is not the same thing as harmonic distortion under listening conditions if you ignore the higher dimensionality of listening when you make the measurement.

LOCAL NONLINEARITIES

The approach to audio analysis presented here is a continuation of an exploratory program initiated in a previous paper and directed toward a better understanding of subjective concepts. The intent of this paper is to explore some of the broad brush aspects of audio processing in such a way as to have the results expressible in human identifiable terms. This requires considering the signal as a single entity not composed of identified constituent parts and considering the operator definable as either linear or nonlinear. An actual program is a dynamic entity which certainly possesses a detailed description, and many processes become nonlinear only after some defined threshold in energy level has been achieved. A complete analysis of this breakdown is beyond the intent of this paper. However, apparent failure to consider them does not compromise the general conclusions which have been reached, as will now be considered.

The "unfolding" of a dynamic process with progression of time may mathematically be considered from the standpoint of a contact transformation. A simple intuitively oriented example of this was given in a previous paper in defining the coordinate called pitch [2]. Thus from a subjective standpoint what is happening "now," if we are listening C^5, is in fact a single entity of experience which is constantly unfolding before us, and our memory of preceding values gives cohesion of presentation. In a purely intuitive fashion we can consider a signal as having a domain of representation only large enough to cover the equivalent epoch of concern. The nature of an operator need only then be defined on the basis of this epoch of concern. If, for example, a playback cartridge "shatters" for groove accelerations above a certain level, then only those signal components processed under such conditions need be considered to be under nonlinear operation in order to investigate the subjective effect of shatter. This common-sense approach is consistent with the concept of mapping.

MEASUREMENT CONSIDERATIONS

The practical implications of these analytical results is that it should be possible to measure audio systems for subjective ranking. Because we usually measure in a different coordinate system than the one in which we listen, we should be more careful in the interpretation of a measurement than is our present practice. A defect which produces a dramatic aberration of reproduction on a measuring instrument may have very little subjective effect on the sound we hear through that system. This will be particularly true if the defect has a counterpart in natural sound that could have arisen under normal listening circumstance. Examples of these include narrow dips in frequency response similar to natural interferences of multiple sound paths, simple all-pass transmissions [1] that are partial representations of natural sound paths, and certain overload characteristics having a nonlinear characteristic similar to that of human sound perception.

On the other hand, nonlinearities which can give rise to an unnatural redistribution of coordinate energy can be subjectively obvious, even though the amount of remnant distortion is very small. An example of this is print-through or groove wall echo which redistributes energy in an unnatural way in time of perception and can subjectively detract from realism at levels that defy conventional measurement.

When measuring audio systems in a one-dimensional fashion, there are certain candidate nonlinearities capable of causing an unnatural redistribution of position, timbre, and delay when we listen to a reproduction in those higher dimensions. In addition to the usual nonlinearities engineers test for, we should carefully note

1) Slew rate limits,
2) Precise way in which overload occurs
3) Any hysteretic nonlinearity.
4) Correlations for time, frequency, or dynamic range.

Detail analysis of such nonlinearities is beyond the intent of this paper. However, in a number of cases the difference between a "clean" amplifier which has a moderate measured distortion and a "dirty" sounding amplifier with almost unmeasurable distortion by standard method, has been correlated with one or more such coordinate spreading mechanisms.

APPARENT PARADOXES

At the present time there are no hard and fast rules to guide in making objective measurements with this technique. There is, broadly speaking, a mental discipline that one can employ to gain an appreciation for how distortion in a component can make itself obvious in listening to a reproduction when that component is in the system. One way to illustrate how to employ that discipline is to try to resolve apparent paradoxes between what we now measure and what we hear.

One such paradox concerns harmonic and intermodulation distortion. Does the geometry of processing indicate whether there is a difference between distortion in an amplifier and distortion in a loudspeaker? The answer is yes, there is a difference. We measure an amplifier in the same coordinate basis as it functions—the single dimension of frequency. When we listen to a reproduction through that amplifier the nonlinearity which gave rise to the original frequency-only distortion is now allowed to cause a crossmodulation of space, pitch, and time cues. This is an entirely different deformation of the subjective sound image than the distortion of spectral balance we measure by simply placing a microphone near the loudspeaker.

Another paradox relates to the question, is it possible for some loudspeakers to sound better with certain amplifiers than others? Again, the geometry of processing indicates that this can happen. A perfect amplifier will allow the full audibility of loudspeaker distortion consisting of crossmodulation of spatial, spectral, and temporal coordinates. A distorting amplifier, on the other hand, will itself crossmodulate these same coordinates prior to the loudspeaker. It may happen that the dual action of amplifier and loudspeaker can result in a net distribution which is closer to realism from the standpoint of human value judgement than that due to the loudspeaker alone. One way this can happen is if the resultant redistribution of coordinate energy happens to be close to distortion effects found in natural sound. Even if the net distortion is higher than that of either device alone, it simply disappears subjectively because we are accustomed to ignoring this sound in natural acoustics.

SUMMARY

There is a very good reason why the "ear" and the "meter" do not always agree concerning distortion. In the majority of cases the meter is not being used to measure what the ear hears.

This of course is what many people have insisted all along. But to this author's knowledge no one has previously tried to raise this conclusion above the level of conjecture, let alone seek an analytical basis.

Starting from the premise that any effect which a reasonable number of audio professionals can independently perceive and describe must be real, research was undertaken to objectively describe this, even if it meant abandoning mathematics of unquestionable rigor. This is obviously not a popular way to proceed since it runs contrary to one's instincts in analysis. What was found [1] was a method of characterizing audio systems which simultaneously verified many observations of audio professionals while recognizing the accuracy of the mathematics.

In that endeavor it was possible to use analytical techniques which have been well developed. In this paper our concern lies with what engineers know as distortion. Unlike the previous paper this work must be speculative. This is because there are no previously established guidelines when describing subjective audio, particularly when considering distortion. Any serious attempt at bringing mathematical principles into subjective concepts is faced with the difficulty that there is no apparent place to start. There is not even the guarantee that it is possible for any such analysis to be performed.

This paper represents an experiment in analysis. Starting from the earlier derived result that there is a geometry of sound perception, all steps in the processing of audio are considered from the standpoint of a transformation from one geometrical form to another. The only geometrical representation which is required to be identifiable with subjective concepts is that in which a replica sound field is involved. In order to guide the analysis through each step in the processing of an audio signal, three postulates are assumed which tie the work to engineering. From that point onward the rules of analysis deriving from such representational geometry indicate what is happening. No matter how abstract the audio signal becomes in this process, it is only the final geometry of sound illusion that is of interest. The only test that can be made on such an experiment in analysis is whether the final form can be tied to subjective audio and if the predicted results actually occur. It is not fruitful in such an analysis to inquire, for example. what is meant by harmonic distortion in a cutter. The only thing that is meaningful is to ask what is the effect of cutter harmonic distortion on the final geometrical form of the sound image.

We want to find the relations which may govern the "form" of dynamic representations and what kinds of deformations may exist. This is a highly abstract process as can be appreciated from the fact that problems as unrelated to each other as an improperly shaped cutting stylus and a reverberant listening room are treated by the same method.

In order to bring the analysis back to reality some classic paradoxes are submitted to the rules to see what results. The results are encouraging because the predictions appear to justify the way people hear things, even if that is contrary to what is now measured by our present instruments. However, simply justifying some familar observations is not sufficient to assert that such a highly abstract process is worthwhile. One wants to know what is predicted that was not known beforehand.

One significant result that appears to be new is that congruence of representation (a true acoustic holograph) requires more information than simplistically obtained from single microphone intercepts, as is now our practice. In order that no hasty misinterpretations occur, this does not mean that our present methods of lateralization of a sound image are necessarily incomplete. They, like a wide-screen motion picture presentation, can preserve angularity very well. But, like a motion picture, the acoustic image may be just as devoid of total three-dimensional presence.

This conclusion may precipitate an instant controversy which unfortunately cannot yet be resolved. Be that as it may, the prediction is that we must record both pressure (a scalar) and velocity (a vector) before we can truly begin approaching accurate reproduction from the standpoint of recreating a replica sound field at a later time and a different place. If properly processed, some of this information may be calculated from a multiplicity of pressure-responsive intercepts, as was shown in an earlier paper [7].

Another result is the prediction that there is a class of distortion which has apparently not been previously identified. This distortion will tentatively be called "representation distortion" since its effect is that of modification of the coordinate representation of a subjective sound image. In nontechnical parlance it could be called "smear." Representation distortion has no meaning in one-dimensional representations but comes about in higher dimensionality of listening. Both time delay distortion [8] and scale distortion [1] are types of representation distortion. Extra wide (acoustically) divas and wandering acoustic images are examples of this distortion.

Does this analysis mean we have a basis for psychoacoustics? Not at all. All that is being done is to set up some postulates relating to the most general ideas of processing and observe what rules of analysis result from them. These rules are then translated into subjective terminology to see if parallels exist in human subjective perception. In this way we are attempting to determine if there may be some objective way of considering the completely personal world of subjective sound images. By concerning ourselves not with the substance of the sound experience, but with its form or "acoustic gestalt," we are not automatically restricted to perfect reproduction as the only standard of comparison. The most acceptable sound illusion may, for example, be the object of analysis. The approach taken is to define and then adhere to guidelines that can be objectively manipulated and measured. This means we can build an instrument (or algorithm) obeying these guidelines and apply test signals.

The test of any model we might come up with is how well the results coincide with what actually happens. When dealing with subjective sound impressions we are in a different world than, say, describing the performance of a mixing board. We cannot hook meters onto people and apply test signals. In fact, we must be aware of the statistical differences of subjectivity and inquire whether the majority of listeners perceive certain effects, or what role experience and training has in subjective impressions.

The idea that perhaps human perception of order and reasonableness is describable by mathematics dates back to Boole [9]. To this author at least, the same should be true for subjective awareness because through our senses we are interpreting a physical world which is itself describable. We are accordingly probing subjective audio with a different mathematical tool, while at the same time fully appreciating that the very concepts advanced, namely, that alternate methods must exist, may obsolete the present approach. Put simply, if the methods of this paper do not work in every case, this is a sign of incompleteness, and other techniques should be looked for.

This, after all, is what was done in the earlier paper to methods that have been around for over a century, and it is only fair that we participate in the same rules. The winner in all cases is the audio engineer who is presented with better and better tools.

REFERENCES

[1] R. C. Heyser, "The Delay Plane, Objective Analysis of Subjective Properties: Part 1," *J. Audio Eng. Soc.*, vol. 21, p. 690 (1973).

[2] R. C. Heyser, "The Delay Plane, Objective Analysis of Subjective Properties: Part 2," *J. Audio Eng. Soc.*, vol. 21, p. 786 (1973).

[3] N. Wiener, *Extrapolation, Interpolation and Smooth*

ing of Stationary Time Series, (M.I.T. Press, Cambridge, Mass. 1966), p. 11.

[4] M. A. Gerzon, "Predistortion Techniques for Complex but Predictable Transmission Systems," *J. Audio Eng. Soc.*, vol. 20, p. 475 (1972).

[5] E. U. Condon and H. Odishaw, *Handbook of Physics*, 2nd ed. (McGraw-Hill, New York, 1967), p. 1–84.

[6] T. L. Saaty, *Modern Nonlinear Equations,* (McGraw-Hill, New York, 1967), p. 3.

[7] R. C. Heyser, "Determination of Loudspeaker Signal Arrival Times: Part III," *J. Audio Eng. Soc.*, vol. 19. p. 902 (1971).

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

[9] G. Boole, *An Investigation of the Laws of Thought on Which Are Founded the Mathematical Theories of Logic and Probabilities,* (1854) (Dover Publ., New York, first American printing of the 1854 edition).

[10] R. O. Hamm, "Tubes versus Transistors—Is There an Audible Difference?" *J. Audio Eng. Soc.*, vol. 21, p. 267 (1973).

[11] R. H. Trumbull and R. S. Mintz, "More About Tubes versus Transistors," *J. Audio Eng. Soc.*, (*Forum*), vol. 22, p. 24, (1974).

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.

PREPRINT NO. 1008 (K-4)

SOME NEW AUDIO MEASUREMENTS

BY

RICHARD C. HEYSER
CALIFORNIA INSTITUTE OF TECHNOLOGY
JET PROPULSION LABORATORY
PASADENA, CALIFORNIA

PRESENTED AT THE

51st CONVENTION

MAY 13-16, 1975

AN AUDIO ENGINEERING SOCIETY PREPRINT

SOME NEW AUDIO MEASUREMENTS

Richard C. Heyser
California Institute of Technology Jet Propulsion Laboratory

ABSTRACT

When we test an audio system or device we tend to ask the listener, "Do you hear the distortion we are testing for?", rather than, "Are we testing for the distortion you hear?". In some cases the distinction between these points of view can be tremendous. Several tests will be described which were derived from the consideration of a geometry of perception and some of the preliminary results will be discussed.

INTRODUCTION

The type of audio measurements generally made today are the same as those made almost half a century ago. These "Golden Oldies" that have survived the test of time are well known to all audio engineers and need not be described here. They are good tests. They have a firm mathematical rooting. They are comfortable to use.

Then why look for additional measurements to apply to audio systems? The reason is deceptively simple. Whether we like to think of it this way or not, an audio engineer shares the professional goal of a magician. When all is said and done, the goal of audio is the creation of a mental illusion in the mind of a human listener. When we listen to a good two-channel stereo reproduction, we have come to expect the acoustic illusion of phantom images and acoustic reality. That phantom image of a center stage vocalist standing between the two reproducing loudspeakers is an illusion. If we set up measuring microphones in the listening room and measure the sound field it is immediately obvious that a simultaneous excitation of two loudspeakers cannot give rise to a spherical wavefront emanating from a position in space between the loudspeakers. Yet we all "hear" such a thing. What is more important we have the capability of verbally describing what we hear. What we need then is some way of measuring the pieces of the audio system so that the numbers that come out of those measurements are related to the final illusion of sound through the one link we have at present with that illusion - the words we all use.

In a previous Journal paper (1), this author set about to do just that. By taking the words we all use and approaching them from the scientific view that they represent a hierarchy of structure, a commonality was indeed found which allowed the verbal descriptions to be cast into a genuine mathematical structure. What resulted did not at first glance appear to be the mathematics of our present measurements. However by taking a harder look it was shown that the math of our present "Golden Oldies" is a special case of the mathematics of this new structure.

If a mathematical structure for describing the listening experience exists, can instrumental measurement be far behind? The answer is that there are already a number of new measurements based on this equally new geometry of perception. The purpose of this paper is to officially debut some of these measurements and to reveal the geometrical basis behind several others which this author has been routinely applying in published reviews in a leading consumer audio magazine.

TESTS

In the geometry of the illusion of sound, one defect of reproduction that is commonly encountered is a spatial shift of the sound image. This may take the form of a simple lateral shift, or a lateral spread, an apparent compression in depth, or in extreme cases a loss of spatial cohesion. There are several measurements that relate to this illusion through the technical investigation of representation distortion.

Representation distortion (2) refers to the crosscoupling of acoustic energy among the coordinates of representation. An undistorted reproduction will not allow the position of a signal to be modified by the tonal values, for example. In other words, Where, When, What tone structure, and How loud are independent entities in a distortionless system.

One measurement that has been found useful in evaluating spatial shift is the measurement of,

1. Incremental transfer function characteristic
   a) Single Tone

A specialized piece of apparatus facilitates this measurement. This is a dual ganged precision potentiometer. One section is a transmit potentiometer and causes an increase in transmission with clockwise shaft rotation, as is conventional. The other section is a receive potentiometer that causes a decrease in its transmission that is an exact compensation for the characteristic of the transmit potentiometer. Standard dB stepped potentiometers are probably best suited for such construction. The purpose of this assembly is to increment the test signal through the transmit pot to the system under test, then decrement the system output through the receive pot so that the test instrumentation always sees a constant level for a perfect system. The effect of distortion will be a departure from perfection. An additional modification which is handy is a third potentiometer or switch wafer attached to the common shaft which will allow a voltage to be made available that is a direct measure of the test signal level.

In this first test a single sine wave is fed through the transmit/receive (T/R) assembly and the output that is to be tested for representation distortion is fed to a coherent processor that multiplies the output signal against the complex conjugate of the test signal. This completely eliminates the angular rate of the test sine wave. Reference should be made to (1) for a definition of some of the terminology which will be used here. A simple approximation to this process may be made by using a heterodyne tracking filter. The intent of this processing is to map the cisoid test signal into a single point on a complex plane. The instrumentation is symbolized in figure 1.

In the simple instrumentation of figure 1, a sine wave generator feeds the system under test. The normalized output of that system is then fed to two four-quadrant multipliers. The original sine wave, and a cosine version made by means of a 90° phase shift network form the other driving voltages for the multipliers. The outputs may be low pass filtered to appear as an X and Y dc voltage.

If the voltages X and Y are used as the horizontal and vertical components of an oscilloscope display, the signal should appear as shown in figure 2. What was originally a sine wave is now mapped to the point P. The instantaneous energy in the test output is given by the distance from the origin, ρ, and the instantaneous phase of the signal is given by the angle Θ.

As the ganged T/R potentiometer is switched through the various drive levels, the oscilloscope display of figure 2 should remain completely steady if the system is distortionless. If it does not, then representation distortion exists. The exact interpretation of this distortion, in subjective terminology, then depends upon the way in which the final reproduction takes place. Several examples will now be given and it will be assumed that the final listener will be using a conventional two-channel stereo reproduction.

If the point moves from some position P_1 to P_2 with increase in drive level, then the incremental system gain has changed. The exact way in which the gain changes is very important in determining the effect on the final illusion of sound. The dashes on figure 3a symbolize the position the spot may occupy (greatly exaggerated for illustration) for 1 dB increments in drive level. In this case there are equal drops in gain with increase in level, something like the old triode amplifiers before the days of feedback. The effect on the sound image (neglecting harmonic distortion for the moment) will be a reduction in apparent dynamic range without much lateral spread. The sound image may in fact appear to move back. The reason is that if this is a measurement of one channel of a stereo system, the relative amount of left and right channel signal is about equally modified by increase in intensity, but the absolute amount is diminished.

If the trajectory is as shown in figure 3b then the sound image will be laterally smeared as well as pushed back. The reason is that new the relative amount of left and right component are unequally processed. The effect is that of a pan pot manipulated by a Maxwell Demon who adjusts the pot in accordance with the instantaneous program dynamics.

If the spot rotates about the origin, as it does in figure 3c, then the sound image will experience a lateral shift without significant change in depth. Again the details of energy change with level are important to determine precisely what happens. In addition we now can suspect a representation distortion between intensity and tone because the rate of change of angle (frequency) is now a function of the rate of change of intensity.

A not uncommon trajectory observed when measuring loudspeakers is shown in figure 3d. The subjective impressions of stereo spatial impreciseness of reproduction closely mirrors such measurement.

b) Harmonic Distortion

The measurement of harmonic distortion seems so simple that it comes as somewhat of a shock to realize that our conventional one-number characterizations are almost meaningless when we look at the geometry of reproduction. If in figure 1 we insert a phase lock generator (PLG) between the sine wave generator and the four quadrant multipliers, we can make a more meaningful measurement of harmonic distortion. The phase lock generator is adjusted to the desired harmonic of the test signal and made phase coherent with it so that the positive zero crossings of the fundamental and harmonic coincide. In this way the map is now made so that the appropriate cisoid harmonic is mapped to a meaningful point on the complex plane.

The details of analysis are now a bit more difficult. The amplitude and phase of each harmonic species should be monitored as a function of drive level. If the harmonics suffer different phase displacements with drive level (that is, the equivalent 3c differs for each harmonic when the order of harmonic is considered) then we know that a representation distortion has occurred between intensity and tonal values. Not only has an additional tonal structure been imposed, but the timbre of that structure will change with intensity not only because of amplitude but because of phase as well.

Space does not permit any proper discussion of what the results may mean in the subjective image, but one point deserves special consideration. The intensity and phase of the harmonic may be a function of how long the signal is applied. Therefore a momentary as well as continuous measurement should be observed. If the point migrates with time to assume a steady state position different than the momentary value, then expect a "fatiguing" sort of sound which tends to become "sour" for high level continuous signals even if it is "clean" on transient bursts. This is a relatively common defect found in loudspeakers.

c) Crossmodulation

The same basic instrumentation of figure 1 is used, but a second signal is added to the original sine wave generator. The composite is processed by the system under test but the analysis is performed on only one of them. The intent is to determine how the presence of one signal may influence the other.

What is commonly found in loudspeakers is that if the sine wave under analysis is a higher frequency (such as 440 Hz) and the added signal is a low frequency (such as 41 Hz) then the stationary spot will now become a closed locus such as shown in figure 3e. This locus will then change with drive level. The amount of amplitude modulation is measured by ΔA, the amount of phase modulation is measured by $\Delta \Theta$, and the overall behavior shown by the complex locus.

Representation distortion is now clearly shown in that spatial, tonal, and intensity values are interrelated. In addition it is frequently found that temporal values are also involved. An average angle shift in a higher frequency caused by the presence of a lower frequency will definitely cause a retardation or advancement in the arrival time of the higher frequency. This is never measured by the "Golden Oldies" but can be readily heard as a virtual quavering of the sound image by percussive bass when this defect is gross.

d) Crescendo Test

If the second signal added to determine crossmodulation has stochastic properties (such as random noise) then the instrumentation of figure 1 must be augmented. If the outputs X and Y are integrated then these may be vectorially combined to give an inner product. The orthogonality of cisoids means that the mean average energy of these inner products will linearly increase for the sine wave component and quadratically increase for the stochastic signal. In order to make a practical measurement it is only necessary to integrate for a finite time, such as one second, and plot the complex value thus obtained as a point on the complex plane. Thus the circuit of figure 4 is added to that of figure 1.

The switches S_1 and S_2 are momentarily shorted then opened. The integrators begin to process the X and Y voltages. At the end of one second the zero order hold samplers (ZOH) are updated by a strobe pulse and then, as soon as update is completed, the switches are again

momentarily closed to resume the process. In this way voltages X' and Y' are generated which are dc voltages updated to the new value every second. This additional processor has now mapped the complex plane of X and Y into another space of discrete representation. The trajectory of the signal in this discrete representation is then analyzed in the same manner as that of the original plane.

This test now lets us evaluate the transient capabilities of the system under test. The stochastic signal is not coherent with the original sine wave and may have peak energies a thousand times that of the sine wave. One could envision this test as a measurement of the manner in which an inner musical voice in a reproduction is modified by the presence of momentary stochastic signals such as applause. It is frequently found that loudspeakers can cause a "ducking" of inner voices under such excitation without substantial spatial shift. In other cases the inner voice is shown to smear laterally with the presence of incoherent energy.

All of these measurements have been directed at determining the representation distortion that may be caused by a change in intensity. These next tests are directed toward the influence of tonal values.

2. Time Gated and Apodized Frequency Response

This test was adequately described some time ago in the Journal (3). However certain aspects of this test will be considered as they relate to subjective values. It has been experimentally determined that the pitch of a momentarily applied tone may be determined for a minimum average period of application of about 13 milliseconds, substantially independent of frequency (4). The tonal attributes (timbre) of a reproducing system may therefore be expected to be established within the first few milliseconds of early sound arrival. In an attempt to determine the timbre related "sound" of a loudspeaker in a room, this author has been measuring the frequency spectrum of the first 13 milliseconds of sound arrival in what has been euphemistically called the "room test". The results are encouraging inasmuch as the anechoic frequency response, measured on the same speakers, has in many cases differed significantly from the early sound, yet the subjective impression of timbre has rather well correlated with the early sound measurement.

In order to perform such an evaluation a test signal is used which has the sharpest autocorrelation function for time delay. This author uses a frequency sweep, but a true impulse will serve as well. The signal is gated in the time domain so as to retain only those components selected for the arrival time. This time function is them multiplied by the particular weighting function that the user wants to impose, such as Hamming (5). Then the complex Fourier transformation is obtained for this gated and apodized signal. The result is an amplitude and phase measurement as a function of frequency.

3. Phase Response

It is now becoming universally recognized that the phase response of the frequency spectrum is associated with subjective values (as is the often measured amplitude response). In particular, the existence of non-minimum phase properties seems to be more detrimental to the sound illusion than minimum phase properties. The "room test" referred to above appears to have provided some possible reasons for that observation. The scattering of sound from an acoustic boundary is governed by the laws of physics. With a few exceptions it has been observed in frequency measurements made on scattered sound that the nature of scattering, considered as a transfer function, is generally minimum phase. Even in a multipath situation the composite sound is composed of minimum phase components even if the net combination is non-minimum phase. Natural sound, then, is generally composed of superimposed minimum phase components. Any substantial non-minimum phase component, therefore, may possibly be subjectively "wrong" because it is not something that can arise under natural acoustic situations and stands apart from them. For the present this must remain a hypothesis, but there is adequate experimental evidence that phase measurements are important to the subjective image.

4. Impulse Response

The time gated frequency response establishes the general timbre of the early sound, but the details of which sound gets to a listener first must arise from the properly interpreted impulse measurement. The energy density measurement (6) gives a good idea of the subjective cohesiveness of impulsive sounds, such as percussion and hand claps. The partitioning between the impulse and doublet (6) response is a measure of the representation distortion of tonal values caused by time delay values. The technology required to measure this effect is now under development.

CONCLUSION

The word "conclusion" is used here only to signify that this brief summary has come to an end rather than the possibility of a conclusive overview. This paper is a preliminary statement of research still in progress to determine some objective measures of subjective sound. The instrumentation that has been described has now survived over five years of use and appears suited to the task set out for it. Some of these may eventually take their place beside the other tests we now all use.

REFERENCES

1. R. C. Heyser, "The Delay Plane, Objective Analysis of Subjective Properties" (in two parts), J. Audio Eng. Soc., Vol 21, p. 690 (1973), p. 786 (1973)

2. R. C. Heyser, " Geometrical Considerations of Subjective Audio", J. Audio Eng. Soc., Vol 22, p. 674 (1974)

3. R. C. Heyser, "Acoustical Measurements by Time Delay Spectrometry", J. Audio Eng. Soc., Vol 15, p. 370 (1967)

4. H. F. Olson, Musical Engineering (McGraw-Hill, New York, 1952) p. 250

5. R. B. Blackman and J. W. Tukey, Measurement of Power Spectra (Dover, New York, 1959)

6. R. C. Heyser, "Determination of Loudspeaker Signal Arrival Times", J. Audio Eng. Soc., Vol 19, p. 734 (1971)

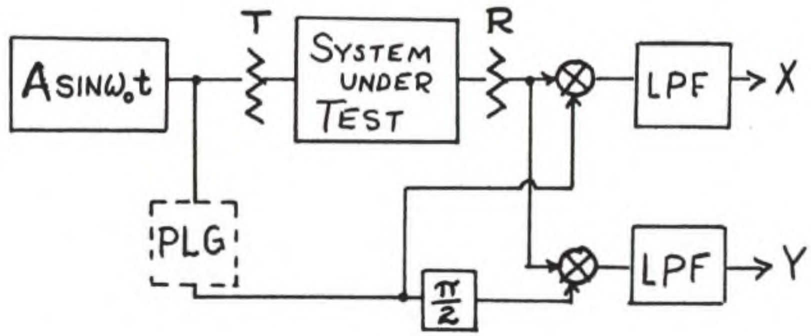


Figure 1. Basic instrumentation for coherent test

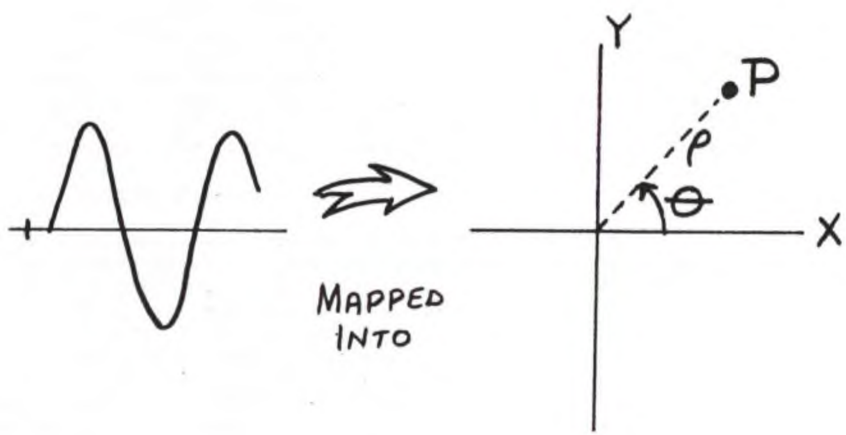


Figure 2. Basic operation of mapping a cisoid into a point, P, that may be viewed on an oscilloscope screen and measured with conventional instrumentation

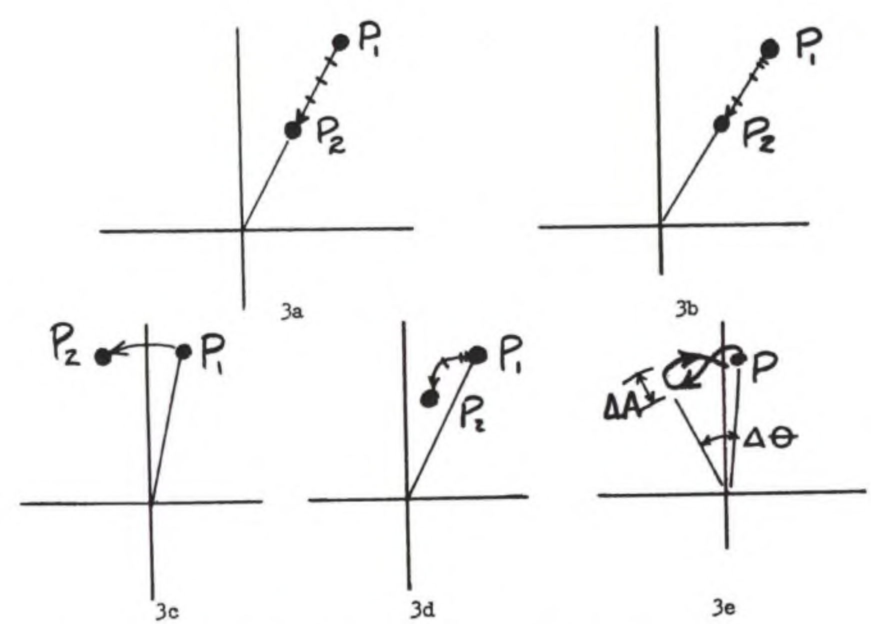


Figure 3. Oscilloscope displays showing various types of representation distortion

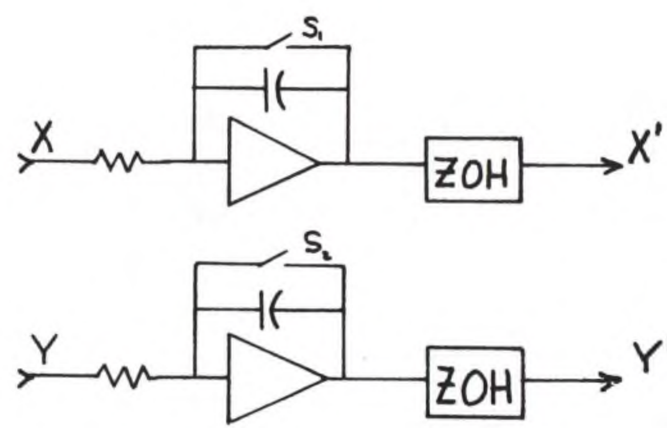


Figure 4. Circuit added to figure 1 for crescendo measurement

GEOMETRY OF SOUND PERCEPTION

PREPRINT NO. 1009 (H-1)

BY

RICHARD C. HEYSER
CALIFORNIA INSTITUTE OF TECHNOLOGY
JET PROPULSION LABORATORY
PASADENA, CALIFORNIA

PRESENTED AT THE

51st CONVENTION

MAY 13-16, 1975

AN AUDIO ENGINEERING SOCIETY PREPRINT

GEOMETRY OF SOUND PERCEPTION

Richard C. Heyser
California Institute of Technology Jet Propulsion Laboratory

ABSTRACT

Much of audio engineering is directed toward a goal of high quality sound reproduction capable of generating the subjective illusion of acoustic reality. Coping with subjective sound impressions has proved baffling because our normal methods do not appear capable of handling subjective descriptions. It will be shown that by defining a geometry of the subjective illusion it is possible to translate some of the apparently esoteric descriptions of subjective audio into objectively measurable terms.

THE SOUND ILLUSION

The end product of the audio industry is the listening experience. The end product is not meter readings, oscilloscope displays, or equations. Our meters and equations are only tools and models which we use to try to measure the end product.

Some people forget this fact and invert the whole process. To them, the equations are the end product and what we perceive is not only incidental to that consideration, but the listener may only hear what the equations indicate he can hear, and furthermore it must be in the form the equations predict.

There is a problem with such a view - it doesn't always work. What we predict, using the equations of analysis that have been around since before Rayleigh, does not always correlate well with what many people claim they can hear.

That has been a tough nut to crack because we know the mathematics is correct. This causes arguments and a "head-in-the-sand" syndrome among those engineers who wish the problem would go away.

Let us still attempt to preserve the mathematics but proceed by taking a different approach. Let us assume that if a reasonable number of critical listeners can independently hear a defect in audio reproduction and describe what they hear, and if that defect does not seem to be evident under contemporary analysis, then we must not assume,

(a) the listeners are "golden ears" freaks who can be ignored
(b) the existing mathematics is either imperfect or incapable of modeling the resultant process.

What we should assume is that if our present math tools seem to fail when applied to such subjective concepts it is because we are not using them properly.

GEOMETRY

To deal with subjective concepts we must also consider form and texture as math attributes. This is apparently a new idea in audio. Furthermore, we must recognize that it takes a number of independent attributes to describe a sound. If you don't believe it, try writing a musical score as a frequency spectrum. The difficulty in doing this is self evident, but this never fazes the analyst who gets so thoroughly wound up in the idea of a time domain and a frequency domain and their sufficiency of description of the audio process that he never looks for higher dimensionality.

There is a branch of mathematics which handles form and texture and multiple-dimensionality. Furthermore, if you really look hard you find that our present time domain and frequency domain mathematics can be considered to be a special twig on that branch.

What is that branch? It is geometry. By geometry I don't mean the tedium of axioms and postulates we encounter in secondary school. I mean the following,

Geometry
- The science that treats of the shape and size of things
- The study of invariant properties under specific groups of transformations.

In addition, we will sometimes need to use a special form of geometry called,

Topology
- Rubber sheet geometry
- The study of things that don't change under plastic deformation of the coordinate basis.

A previous paper (1) introduced concepts of topology into linear audio analysis and demonstrated that,

(a) It is possible to describe subjective impressions of form with math models
(b) There is a geometry of representation
(c) Present models (time domain and frequency domain) are correct and sufficient, but they are a special case of the more general geometry.

This demonstrated that much of the apparent disagreement between objective and subjective views springs from the same thing that caused construction difficulties on the Tower of Babel - namely, they don't speak the same language. Geometry gives a common language.

However, the subject that really causes the more violent disagreements between the subjective and objective person is that of distortion - how to measure it and how to interpret what the measurement means. Distortion is a manifestation of a nonlinear process. A second paper (2) carried the idea of a geometry of representation forward into nonlinear processes in order to look at the problem of distortion.

A number of things were found in that paper, but perhaps the most important finding was that we could identify a new type of distortion from the technical point of view. This is called representation distortion. If we have a system of description for an apparent source of sound (its coordinates of representation) then an undistorted sound will have a well defined distribution among those coordinates. For example, the sound of a violin in an orchestra will appear to come from a definite position in space, have a definite intensity, be playing with a definable timbre, and occur in the proper sequence of relative time. If the system of reproduction distorts that sound, then the representation will be altered and redistributed. The violin may now be smeared in space as it gets louder or plays certain tones. Or possibly the violin sound may be redistributed in tone and intensity such that the timbre changes with the instantaneous program dynamics. This is representation distortion. The form of the sound is warped. The sound is smeared.

What is exciting about this technical result is that this type of deformation fits the verbal descriptions that many listeners often use when they attempt to describe the subjective effect of various distortions. Let us then consider how we might conceptualize a measurement of representation distortion.

A GRAPHICAL REPRESENTATION

We need to measure the space, time, tonal, and intensity properties which can be imparted to an audio representation. The delay plane paper (1) showed how to treat this as a five-dimensional thing. To most people the idea of five dimensions is a bit mind boggling. So let's try to reduce it to a three-dimensional thing. At least we can then sketch it on a piece of paper.

Suppose we imagine that we "freeze" the audio image (the thing we hear and think of as a phantom sound source) at a moment in time. We are left with the three dimensions of space, tone, and intensity as shown in figure 1. By "space" we mean where the sound is. In its own right this is two-dimensional since "where" means the azimuth and elevation (left-right and up-down) of the sound relative to the listener. (Range, or how far away a sound is, becomes tied up with the coordinate of time of arrival). In order to proceed with this discussion assume that the axis labelled "space" is the altazimuthal coordinate and that if we need to determine position we can always get more detailed and go into that axis as a two-dimensional thing.

- 3 -

The coordinate labelled "tone" is really the measure of the spectral distribution of the sound in terms of its pitch values. The spread of sound along this axis is what we call timbre. This axis is not frequency. Frequency is a very special coordinate in another type of representation and is, mathematically speaking, not what we mean by pitch. The way we measure both of them is such that the numbers can be made the same if we have a pure pitch tonal that lasts forever. Consequently if you want to think of a pure pitch tonal of middle C as corresponding to 262 Hz, go ahead. The numbers will be the same, but that is all. Perhaps someday we shall have a proper name for the measure of pitch.

The third coordinate axis is intensity. This is a measure of how strong a signal is.

In our attempt to make things simple and three-dimensional, we have assumed that the clock is frozen at "now", whenever that may be. There is another three-dimensional sketch, using the same coordinates, that corresponds to each moment in the past, and there is presumably a set of such sketches for what will happen in the future. But we cannot see the future. We don't know what will happen. Perhaps a heuristic way of thinking about this display is that each plot represents a separate frame of a motion picture which we are viewing, and we are forbidden to see what is going to happen but we can look back over the past frames to see what has already happened. When we look at the way things change with time, we may then think of this as a measurement involving a fourth coordinate of relative time.

Having defined a method of representation, let's take a look at how we might use it. Figure 1 represents a series of these three-dimensional "snapshot frames" corresponding to successive moments in time. Point A is a single element that has the same position frame-after-frame. It will be a single pitch tonal that comes from the same position in space and with the same intensity during the entire time span represented by figure 1. These plots correspond to what the listener hears. If the listener is using a stereo reproducer, there are two signals which need to be independently measured and combined so as to give the subjective impression of figure 1. In this case there will be a left channel sine wave and a right channel sine wave. Their frequency determines the pitch. Their joint amplitude determines intensity. And their relative amplitude and phase determine the apparent position in space from which the sound originates.

- 4 -

Now let's consider figure 2. This is a tone that not only changes pitch but also intensity and space location. The listener is hearing a tone moving about in space. Again, because of the simplicity of the example we need only measure the joint amplitude, phase, and frequency of the signals going to left and right channel.

Things get more interesting in figure 3. Here we have a single tone, A, originating from one point in space with a constant intensity during frames 3a and 3b. Suddenly, in 3c, another tone, B, is introduced. The effect on A is that it is broadened in pitch, position, and intensity. A is crossmodulated by B. In figure 3d the increased intensity of B now rather seriously warps the space, tone, and intensity properties of A. We have a representation distortion. The listener complains that A is blurred when B is present.

How could this come about? Suppose A is a stereo signal that is twice as intense in the left channel as it is in the right channel. A appears to come from a left-of-center stereo stage position, as the listener perceives it. Suppose B is a stereo signal that is twice as intense in the right channel as in the left channel. B is a stage-right image.

The right channel carries the smaller part of A and the greater part of B. If the gain of the stereo channels (slope of the output/input characteristic) changes with signal strength then the amount of crossmodulation in the right channel will be different than in the left channel. The right component of A will be altered by B in a different manner than the left channel component.

It is apparent that A will experience a lateral shift, from the listener's standpoint, that is equivalent to a pan pot manipulation. A may also have its spectral values altered as well as intensity.

PRACTICAL OBSERVATIONS

All of this is so obvious that any step by step description seems trivial. Yet stop to think what this means in terms of testing an audio system. The listener complains of a warping of the sound image. He is correct. This comes about (in this case) because the sound illusion is created from the joint contribution of two components. He hears a sound apparently coming from the position in space between the speakers even if he knows that it originates from the position of the speakers themselves. Yet, in today's technology, we measure each channel alone as though the processing of that channel is the only thing needed to measure distortion. This is not correct, as indicated from the geometry of the situation.

- 5 -

The detail behavior of each channel and the proper interpretation of how they combine to give the illusion of sound is necessary. The fact that a channel may have 0.1 percent crossmodulation between two sine waves at a certain intensity has almost nothing to do with the amount of distortion perceived by the listener and in no way tells us how the subjective image is warped. It is the detail nature of this measured distortion that we must use. Once one becomes accustomed to thinking about the geometry of the problem it becomes easy to imagine types of measurable distortion that are benign as far as their effect on the sound illusion but have high numerical values. A constant curvature transfer characteristic, for example, can lead to such a situation.

Conversely, there are distortions which can contribute to a highly smeared sound image yet have very low numerical value when measured by ordinary methods. A transfer characteristic possessing a nonlinear phase or time delay characteristic, for example, falls into this catagory.

An important consideration can arise for certain types of spectrum spread with intensity, as symbolized in figure 4. Human perception is itself nonlinear. It has to be for survival in a world with possible intensity changes greater than a million-million. Accordingly it may be a perfectly natural phenòmenon to have an image point A create corresponding points B and C in our perception of sound. Incrementing A to A_1 may have as a natural byproduct the change of B to B_1 and C to C_1. In other words, we are accustomed to a certain "warping" of the sound image. We do not hear it as a warping, of course, because to us this is natural sound. All coordinates actually share in this warping and are untangled in our "software" of sound perception.

We must be aware of this possible warping when we make a harmonic distortion measurement. It is not sufficient to measure a one-number harmonic distortion for comparative measurements of audio systems. We must measure the amplitude and phase of each harmonic component as a function of signal level.

Also don't get angry if a listener claims that one audio system measuring greater than two percent harmonic distortion sounds "cleaner" than another system measuring 0.01 percent. He may be right. The higher distortion may mimic the properties of natural hearing and only sound louder.

- 6 -

There are many, many more examples of the use of this geometrical concept in describing the illusion of sound created in the listener's mind and the measurable components of signals which go to make up that illusion. More complete description of them would be too lengthy to go into here. However, some of the more apparent problems that may be handled include,

(a) The fly-on-the-wall impression of stereo reproduction which can occur when the lateralization coordinates are accurate but the cues leading to depth in the sound image are distorted. The result can be a compression of the sound image into a two-dimensional painting stretched between the two loudspeakers.

(b) The spatial shredding of distortion fragments which will spread over a greater distance than the fundamentals giving rise to them. This generally results from rate limiting in the reproduction chain or transfer functions with high order curvature (crossover notch, clipping, etc.).

(c) Time base distortions leading to unnatural illusions. These include time base fluctuations (wow and flutter), pre echos, post echos, and multiple arrivals (yes, the listening environment is important to the sound illusion).

CONCLUSION

Plato said, "All is geometry". It is fair to say he did not know a great deal about audio engineering when he made that statement. However there is still a great deal of truth in that quotation when it is applied to audio. It offers an exciting solution to the long standing problem of being able to measure what we hear.

REFERENCES

1. R. C. Heyser, "The Delay Plane, Objective Analysis of Subjective Properties" (in two parts), J. Audio Eng. Soc., Vol 21, p. 690 (1973), p 786 (1973)

2. R. C. Heyser, "Geometrical Considerations of Subjective Audio", J. Audio Eng. Soc., Vol 22, p.674 (1974)

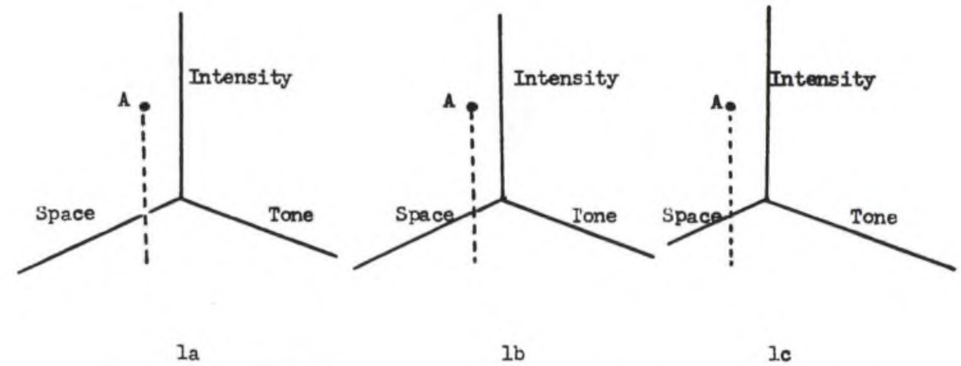


Figure 1 Representation of a tone that remains fixed in position, intensity, and pitch.

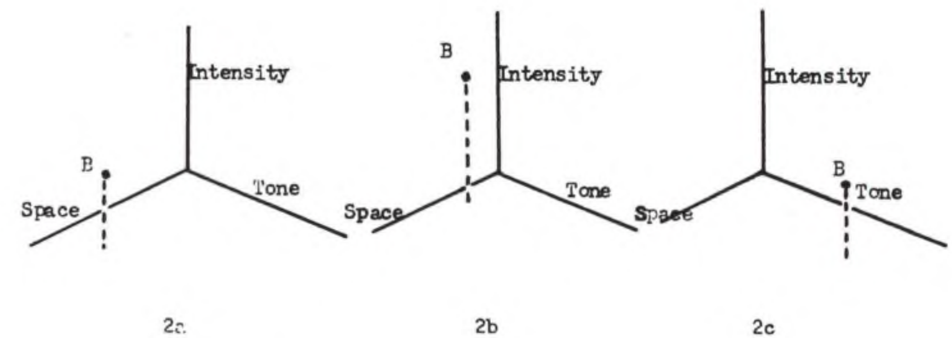
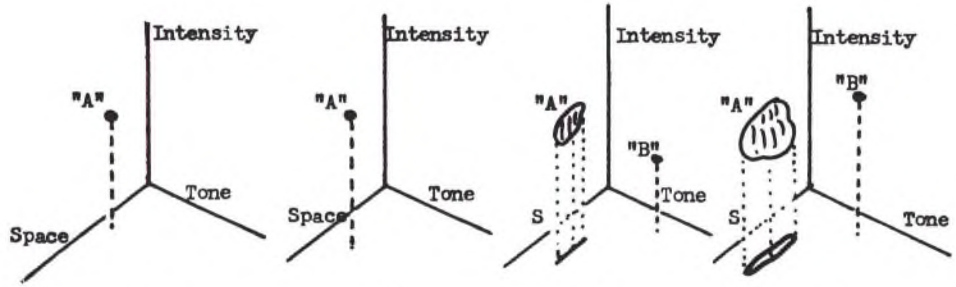
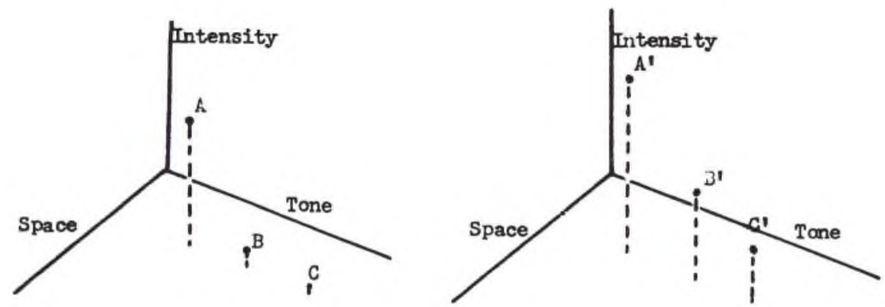


Figure 2 Representation of a tone that changes in position, intensity, and pitch for successive moments in time.



3a 3b 3c 3d

Figure 3 One type of representation distortion. A tone "A" when presented alone has a well defined pitch, position, and intensity. Upon the introduction of a second tone "B", the energy of "A" becomes redistributed in space and intensity. As "B" increases in level the smear on "A" becomes more pronounced and expands into tonal spread as well.



4a 4b

Figure 4 Human perception may be nonlinear to the extent that an otherwise pure tone, A, may create harmonic terms B and C. As A becomes more intense the harmonic terms become proportionally larger.

Reprinted from *JAES*, vol. 23, no. 7, pp. 562–565, 1975 September.

SOME USEFUL GRAPHICAL RELATIONSHIPS

RICHARD C. HEYSER

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

A recently published mathematical law of energy density partitioning is used to derive some simple graphical relationships that must always be true for a proper physical measurement. In most of the measurements made by engineers, whether of acoustic, electrical, magnetic, or physical properties, it is possible to use these graphical relationships and tell, simply by looking at the data plot; 1) if the data were correctly taken; 2) if the system under analysis is minimum phase; 3) what the other component of a single component plot should be.

INTRODUCTION: Graphs and plots constitute a powerful language which can convey a great deal of information to the knowledgeable engineer. I would like to pass along some graphical relationships which are simple to learn and easy to apply to audio measurements. Armed with such information it is possible to tell a great deal about some system just by glancing at the data plots.

Systems and devices of greatest concern to engineers, whether they are electrical, acoustical, or mechanical, are tested by measuring the cause and effect relationship between an applied stimulus and the observed reaction to that stimulus. If we want to measure everything about the reaction, we must make a detailed observation of its energy density relations. In a previous paper [1] this author presented a principle of energy partitioning which governs such relations in linear passive systems. Stated simply, the positive square roots of the components of energy density partitioning are related by Hilbert transformation. The fact that this principle was presented in a technical paper about loudspeaker reproduction may have obscured for some readers the fact that it is truly a general rule. Some examples of where it may be applied in audio engineering include

1) real and imaginary components of acoustic velocity in damped pipes,

2) resistive and reactive components of acoustic driving-point impedance,

3) In-phase and quadrature components of the frequency response of rooms,

4) amplitude and phase of minimum-phase transfer functions,

5) In-phase and quadrature components of structure-bone seismic waves,

6) transmission-line admittance versus distance from termination.

If you have a measurement that is supposed to represent the complete information about a particular exchange of energy, look at the curves that may be plotted from these data. Note where they have the sharpest bends, the flattest portions, the places where the slopes are steepest, the places where there are peaks and valleys, and whether the curves curl to the right or left. Then compare what you find against the rules stated below. A minimum-phase network response when plotted as a separate gain and phase must conform to these rules; if it does not, then it is nonminimum phase. Every other measurement with a real and imaginary part or an in-phase and quadrature part must conform. In some cases the mathematical relationships are not precise, but the rules are presented in a form that indicates how the curves will appear to your eyes.

RULES

In order to simplify the description of these visual aids, we will use the symbol (a) to stand for the curve representing amplitude, in-phase component, or real component and the symbol (b) to stand for the complementary curve of phase, quadrature component, or imaginary component, respectively. This covers all the graphs you might see under one set of terminology.

Also, we will use the following convention. The data are assumed plotted so that an increase in a value, such as higher gain, larger positive value, or increase in phase lead, will be represented by an upward deflection on the plot. This is, or should be, a conventional audio standard.

An inflection in a curve is the place where the direction of curvature changes from one polarity to another. If a curve has been curling in a downward direction as the frequency, time, or distance values increase, and transitions to an upward curling as it goes through the inflection point, then for the sake of simplicity we shall call that an "upward inflection." A transition from an upward curling to downward curling will be called a "downward inflection."

An extremum is a place where the curve is locally a peak or dip. A place of maximum curvature will refer to a place where the curve has the sharpest local bend. If the curve bends to the right with increase in frequency, time, or distance, then we shall call it a downward curvature. A peak has a downward curvature. A dip has an upward curvature.

Now to the rules.

1) If the data are plotted using a *logarithmic* scale for the independent variable, then

i) except for places of sharp transition, the numerical value of (a) will be approximately proportional to the negative of the slope of (b), while the value of (b) will be approximately proportional to the positive of the slope of (a), with the limiting case that a constant slope of (a) corresponds to a fixed value of (b) (Bode [2]);

ii) a local extremum on either curve will correspond to a point of inflection on the other curve;

iii) a place of maximum curvature on either curve will generally correspond to a point of inflection on the other curve;

iv) the relation between direction of maximum curvature and direction of inflection is such that a downward curvature in (a) corresponds to an upward inflection in (b), while an upward curvature in (a) corresponds to a downward inflection in (b). Also a downward curvature in (b) corresponds to a downward inflection in (a), while an upward curvature in (b) corresponds to an upward curvature in (a).

2) If the data are plotted using a *linear* scale for the independent variable, then

i) except for places of sharp transition, the numerical value of (a) will be approximately proportional to the negative of the slope of (b), while the value of (b) will be approximately proportional to the positive of the slope of (a);

ii) as a very special case of this relationship, except at places of sharp transition in (a), the group delay of a minimum phase transfer function will be similar in shape to the amplitude response;

iii) a local extremum on either curve will correspond to a point of inflection on the other curve (Heyser [3]);

iv) a place of maximum curvature on either curve will generally correspond to a point of inflection on the other curve (Heyser [3]);

v) the relation between direction of maximum curvature and direction of inflection is identical to rule 1 iv) above.

EXAMPLES

These rules apply to overall trends as well as localized structural features. An example of the way these rules can be utilized is shown in Fig. 1. Curve *a* is the amplitude of the frequency transfer function of some network and *b* is the phase. Extremum and places of maximum curvature are shown as circles which one can imagine are pipes around which the curves are bent. The places of inflection are shown as horizontal dashes. It is obvious that rules 1 i–iv) are obeyed at local features. Note also that as an overall trend there is a general curvature in the amplitude plot at around f_4, where the low-frequency and high-frequency asymptotes join, and this is signified by a corresponding inflection in the phase background. The fact that all the rules are obeyed means that if you are handed this response and are told that it belongs to a network transfer function, you know that the network must be minimum phase. If there is any place where the rules fall down, then you know that either the data are incorrectly taken or that this network is nonminimum phase. If the places of maximum curvature do not line up with points of inflection, do not waste any time on it, the data are incorrect. If the data purport to represent any two-terminal thing, such as impedance, and if the rules are not obeyed, the data are in error. If you are only given one piece of the data, such as the amplitude response, you can quite accurately sketch what the phase should be if it is either a minimum-phase network or any two-terminal measurement. If the person plotting the data were thoughtless and did not identify which curve was a and which b, or labeled them improperly, you can instantly identify the proper curve, since there is only one combination that works.

As another example, Fig. 2 shows the frequency transfer function of a nonminimum-phase device, in this case a parallel T circuit. The places of maximum curvature and inflection still line up, but the important difference is that the slope of the phase goes the wrong way to be a minimum-phase network. If this particular amplitude–phase behavior were imbedded within the more complicated variations that may be found in a loudspeaker system, you can instantly spot it as representing a nonminimum-phase point. If this plot were supposed to represent a complex impedance, instruct the person making the measurement to do it over because he made a mistake.

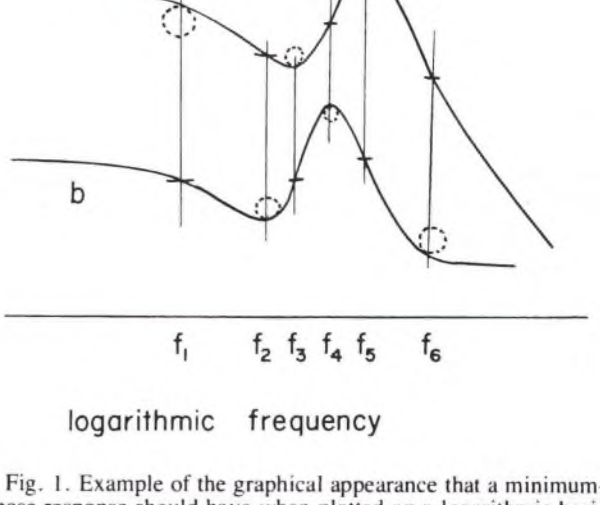


Fig. 1. Example of the graphical appearance that a minimum-phase response should have when plotted on a logarithmic basis and having an amplitude *a* part and a phase *b* part. Places of highest curvature (shown by circles) and inflection points (shown by dashes) should very nearly line up and be related as shown in this figure if an upward value corresponds to increased gain and a phase lead. Asymptotic slopes in amplitude should correspond to constant phase values.

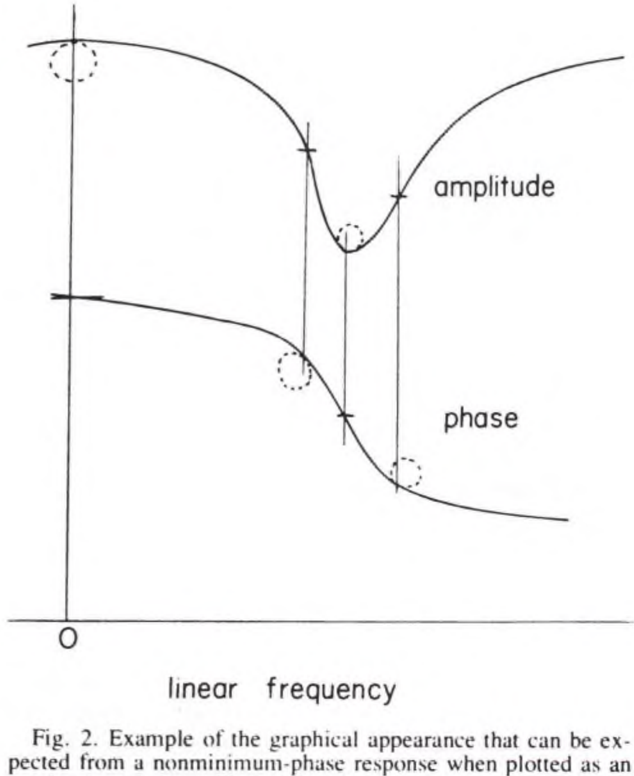


Fig. 2. Example of the graphical appearance that can be expected from a nonminimum-phase response when plotted as an amplitude and phase. Places of maximum curvature and inflection points still line up, as they do in Fig. 1, but the direction of the inflection will be opposite to that of Fig. 1. When a linear scale is used for the independent variable, in this case it is frequency, the graphical relations continue to be valid right through the zero reference.

Rule 2 ii) is a very interesting consequence of the Hilbert transform graphical relationships. The negative of the derivative of the phase response as a function of frequency is approximately proportional to the amplitude response. This means that group delay [3] measurements and amplitude measurements will be look-alikes if you are measuring a minimum-phase system. If you have plots of group delay and amplitude response, you can instantly tell whether the device being measured is minimum phase. Reference should be made to Muraoka *et al.* [4] to see examples where this rule may be applied to phonograph cartridges.

Once you become accustomed to these rules it becomes almost second nature to look over data and identify trends, and occasionally even spot mistakes that could otherwise cost you a lot of lost time.

DERIVATION

A complete analysis deriving these rules (as well as other more complicated relationships for both rectangular and polar data plots) would be highly technical and detailed enough for a long paper in its own right. That is not the purpose in this note. The interested engineer can find the extremum–inflection and curvature–inflection relationship in [3, Appendix A]. The energy basis is in [1, Appendix]. The derivative relationship can be found by proper manipulation of Eq. 9 of [1] and its association as a generalized function relationship [5]. Bode [2] is the principal source for rule 1 i) and its application. In fairness to Bode, he derived his enormously valuable rules without recourse to Hilbert transformation, although a footnote acknowledgement of the work of Y. W. Lee [2, p. 303] indicated he was not unaware of its existence. The skew reciprocity property of the Hilbert transform which gives rise to the polarity of the curvature–inflection characteristics, and incidentally allows instant identification of (a) and (b) even if the curves are unmarked, may be seen in [1, Fig. A-4].

OBSERVATION

There is one technical observation which this author finds fascinating. It would appear feasible to augment the all too sparse tables of Hilbert transforms by including those appropriate solutions which have already been obtained for electroacoustic problems. A quick comparison of the driving point impedance relationships in Olson [6], for example, reveals functional relationships not found in some of the better math tables [7].

REFERENCES

[1] R. C. Heyser, "Determination of Loudspeaker Signal Arrival Times" (three parts), *J. Audio Eng. Soc.*, vol 19, pp. 734-743, 829-834, 902-905 (Oct., Nov., and Dec. 1971).

[2] H. W. Bode, *Network Analysis and Feedback Amplifier Design* (Van Nostrand, New York, 1955).

[3] R. C. Heyser, "Loudspeaker Phase Characteristics and Time Delay Distortion," (two parts), *J. Audio Eng. Soc.*, vol. 17, pp. 30-41, 130-137 (1969).

[4] T. Muraoka, H. Onoye, and A. Takayanagi, "Measurement of Phonograph Cartridges by the Pulse-Train Method," *J. Audio Eng. Soc.*, vol. 22, pp. 504-510. (Sept. 1974).

[5] R. C. Heyser, "The Delay Plane, Objective Analysis of Subjective Properties" (two parts), *J. Audio Eng. Soc.*, vol 21, pp. 690-701, 786-791 (Nov. and Dec. 1973).

[6] H. Olson, *Elements of Acoustical Engineering* (Van Nostrand, New York, 1955).

[7] H. Bateman, *Tables of Integral Transforms*, vol II (McGraw-Hill, New York, 1954).

Note: Mr. Heyser's biography appeared in the November 1974 issue.

Reprinted from *JAES*, vol. 24, no. 9, p. 752, 1976 November.

ON DIFFERENTIAL TIME DELAY

JAMES MOIR

James Moir & Associates, Chipperfield, Herts., England

Recent issues of the *Journal* have contained two excellent contributions by Heyser[1] and Preis[2] dealing mainly with the effects of phase shift, group time delay, or, as I prefer to call it, "differential time delay," on the waveform of a complex signal.

The relevance to sound quality is not stated explicitly, both contributions dealing with the effects of time-delay differences on the preservation of wave shape, rather implying that wave shape preservation is essential if high-quality reproduction is to be achieved. Now I do not believe that there is any engineering evidence that wave shape preservation is essential to the maintenance of good sound quality, provided that the time delays that produce the wave shape deterioration are kept within the CCIR requirements; roughly, the differential time delays should not exceed about 8 milliseconds at 6–8 kHz.

An experiment that is widely quoted in this connection is that of Hilliard who noted that effects were noticeable on the transient sounds of tap dancing when the differential time delay exceeded about 2–3 milliseconds. As part of a larger investigation of these effects we have recently carried out similar experiments. These involved the use of a very simple high-quality reproducing system consisting of a Bruel & Kjaer half-inch microphone, amplifier, and several types of high-quality loudspeakers, including Quad electrostatics, a type of loudspeaker that our tests have shown to have lower values of differential time delay than any others we have measured.

The program material was chosen, not for its musical value, but because the sounds were those that would generally be called sharp transients, for example, the clicks from the overcenter flicking of a thin metal tongue in a child's toy and the sound of striking two pieces of metal together. The equipment allowed us to delay the components above about 4 kHz with respect to the components in the band below this. If the observer was allowed to continue to listen to such a noise while the time delay was varied, he slowly became aware of a "just detectable" difference in sound quality when the delay difference was between 2 and 3 milliseconds, with a more obvious effect when the time delay reached around 5–6 milliseconds. On music there were no effects observed until the delay was generally much greater than 6 milliseconds.

Differential time delays of this order produce such gross changes in the waveform of a complex sound that the time-delayed signal can hardly be recognized as having any relation to the undelayed signal. Waveform preservation is absolutely essential in many applications, but before we become involved in the complexities and costs of modifying our sound systems to minimize differential time delays (phase shift), could we have some evidence that they are of significance in a monaural channel system. Our own evidence is that provided they are held somewhere around the CCIR specified values, they are of no significance whatever,—but perhaps we have missed something.

[1] R. C. Heyser, "Some Useful Graphical Relationships," *J. Audio Eng. Soc.*, vol. 23, pp. 562-565 (Sept. 1975).

[2] D. Preis, "Linear Distortion," *J. Audio Eng. Soc.*, vol. 24, pp. 346-367 (June 1976).

Replies:

RICHARD C. HEYSER

California Institute of Technology, Jet Propulsion Laboratory, Pasadena, CA 91103

In the abstract of my January, 1969, *Journal* paper I stated, ". . .A concept of time delay is introduced which provides a physical description of the effect of phase and amplitude variations as a frequency-dependent spatial smearing of the effective acoustic position of a loudspeaker." And in the abstract of the April, 1971, *Journal* paper I repeat, ". . .the effect of imperfect loudspeaker frequency response is equivalent to an ensemble of otherwise perfect loudspeakers spread out behind the real position of the loudspeaker creating a spatial smearing of the original sound source." I think that still sums up the effect on sound quality. We have no disagreement in principle. Imperfect wave shape is important only as it relates to distortion of the perceived acoustic image. Thank you for allowing me to reaffirm this position.

D. PREIS

Gordon McKay Laboratory, Division of Engineering and Applied Physics, Harvard University, Cambridge, MA 02138

These interesting preliminary findings reported by Mr. Moir will perhaps encourage psychoacoustical investigators to undertake more comprehensive perceptual research on linear distortion. Indeed his report would have been more useful if curves were included indicating the precise amounts that frequency (magnitude) response and group-delay response were altered as a function of frequency. To the extent that comparison is valid, tolerance on group-delay distortion (that is, the maximum deviation from flat group delay) for international telephone transmission of speech is 10 milliseconds in the 300–800-Hz band and 5 milliseconds in the 800–2400-Hz band. The group-delay distortion tolerance for wideband program material is likely to be different.

Reprinted, with permission, from the *Proceedings of the IREE*, vol. 37, no. 3, pp. 67–76, 1976 March.

Concepts in the frequency and time domain response of loudspeakers

*R.C. Heyser**

SUMMARY: Up until slightly less than a decade ago, the measurement of the frequency response of loudspeakers was generally restricted to the amplitude of the sound pressure. The phase of the sound pressure was all but ignored as was the time-domain acoustic response. This was dramatically altered when published measurements, made using a newly developed coherent processing technique, showed that many loudspeakers had significant non minimum-phase portions of their frequency spectrum and that all loudspeakers had a complicated behavior of their frequency response for the period of time following an impulsive excitation. The effect of this was to prove that the long established measurement of anechoic steady-state amplitude response was not sufficient to properly characterise the acoustic response of loudspeakers. An industry had to change its standard of measurement.

Once phase measurements could be made accurately, a nagging inconsistency in measured time delay was observed which led to the formulation of a deeper physical concept. This, in turn, led to the prediction, with subsequent verification, of time-delay distortion; to a derivation of the analytical basis for kinetic and potential energy density partitioning in dynamic processes; to the recognition that the time-domain response must also be considered an entity with a magnitude and a phase and to the introduction of a particular frequency response measurement based upon a very simple phycho-acoustic observation. Present developments in loudspeaker analysis are concerned with understanding the role that linear and non-linear electro-acoustic properties play in the listening experience.

* Jet Propulsion Laboratory, Pasadena, California.

Ms received by IREE 27.11.75

Introduction

Of all audio components, the loudspeaker stands alone in the difficulty of measuring performance and interpreting the results of such measurements. Until quite recently this situation has discouraged any serious attempts at loudspeaker performance evaluation. Now the field is beginning to stir with life as more and more experimenters are discovering that meaningful loudspeaker measurement is not all that mysterious.

This paper presents an overview of some of the technology of loudspeaker testing with which this author has personally been involved in the past decade. I will discuss my own work, not to denigrate or ignore the enormous contributions of other experimenters in this field but to try to bring a personal perspective to this fascinating subject and to outline some of the thought processes which I have found necessary in coming to grips with this difficult field.

Progress in analysis is seldom achieved in quantum jumps. More often than not it is slow and deliberate. In many cases when the path seems blocked by an impenetrable wall, a passage can be found by going back and taking a more careful look at those nagging inconsistencies which did not seem important when initially encountered but which in fact point the way to a deeper underlying principle.

The technology which will be described has been applied in a variety of fields related to the subject of acoustics, including seismic measurements, sonar transducer calibration, acoustic surface characterisation and medical ultrasound. [1-5] However for simplicity the description will be restricted to the aspects of loudspeaker measurement.

Frequency response

What exactly is meant by the frequency response of a loudspeaker? It is clear from conventional circuit theory that the frequency-domain representation of the sound pressure at a listening point caused by a unit voltage applied to the loudspeaker terminals must have both an amplitude and a phase. Today this fact is commonly accepted. Less than a decade ago it precipitated a controversy. The general reaction at that time was, "phase — what's that?". Even though Wiener, [6] Ewaskio and Mawardi [7] and Stroh [8] had actually measured phase, this fact had not percolated to the working level.

In order to discuss frequency response for loudspeakers it is necessary to define it as the complex Fourier transform of the pressure response to an impulse of electrical energy. [9, 10] The loudspeaker frequency response, H(ω), thus consists of an amplitude response A(ω) and a phase response φ(ω) under the relationship,

$$H(\omega) = A(\omega)e^{i\,\varphi(\omega)} \tag{1}$$

It frequently happens that the amplitude response is better characterised as a logarithmic function $\propto(\omega)$, which leads to the simplified form,

$$H(\omega) = e^{\propto(\omega)}\,e^{i\,\varphi(\omega)} \tag{2}$$

The frequency response of a loudspeaker then consists of two plots. The plot magnitude of sound pressure, expressed in decibels as a function of frequency, which represents $\propto(\omega)$, is simply called *amplitude.* The plot of phase angle of sound pressure as a function of frequency, $\varphi(\omega)$, is simply called *phase.* Taken together these two plots completely characterise the linear loudspeaker frequency response under the relation,

$$\ln H(\omega) = \propto(\omega) + i\,\varphi(\omega) \tag{3}$$

This appears quite straightforward. However, the acoustics industry had been measuring $\propto(\omega)$ and ignoring $\varphi(\omega)$ for forty years. Now they were told it was necessary to measure $\varphi(\omega)$ as well. The general reaction was, "why?". The answer to that question not only provided a solution to the long standing mystery in loudspeaker measurement but opened a Pandora's Box of future analysis.

Minimum-phase

For a long time loudspeaker designers had been aware of two apparent mysteries in their art. First, the loudspeaker that measured better did not necessarily sound better. Thus while amplifier manufacturers were proudly announcing the uniformity of "frequency response" to a fraction of a decibel, the poor loudspeaker manufacturer did not dare publish what he actually measured for his product. Some of the better "sounding" loudspeakers had a frequency response that was more akin to an elevation profile of the Swiss Alps.

This led to the second mystery, because it was common knowledge that if one took a "good sounding" production speaker and "flattened" the frequency response by adding a bit of mass here and scraping some structure there, the resultant product invariably sounded worse.

With the advent of phase measurements one of the answers to this mystery quickly became apparent. Audio designers had become accustomed to the observation that a flat amplitude response automatically meant perfection. That, of course, is simply not so in general. The phase response can in its own right render a very unacceptable reproduction even if the amplitude response is quite flat. By not measuring the phase response, the early designers did not really have a valid measure of more accurate reproduction.

The problem was caused by the fact that almost every other audio device which was easy to measure, such as amplifiers, equalisation networks and the like, was of minimum-phase type. It was common experience that the best square wave and tone burst response for such networks always resulted when the amplitude of the frequency response was the most uniform. To make matters worse, many textbooks were so exclusively devoted to minimum-phase systems, without even a warning of this fact to the unwary reader, that many engineers did not know of the existence of non-minimum-phase properties.

A number of experimenters had observed that the square wave response of a loudspeaker did not always improve with frequency equalisation but little note was taken of that fact. The same was also found true of magnetic playback and disc equalisation but these were apparently considered anomalies apart from loudspeaker reproduction. While an early paper pointed out the possibility that loudspeakers could be non-minimum-phase, [11] this fact was lost in subsequent debates on the significance of phase. [12 - 14]

The amplitude and phase response of a network are uniquely related only if that network is of the type known as minimum-phase. [15] Loudspeakers are generally non-minimum-phase over at least part of the frequency spectrum. [10] Thus the phase measurement assumed an immediate and important role for the estimation of one measure of quality of reproduction.

Time delay distortion

While this provided a possible answer to one simple question, the measurement of phase quickly led to a new set of problems. It was observed that actual loudspeaker systems in some cases had a rather complicated frequency response when both amplitude and phase were considered. What did this mean? Also, in particular, what was the importance of the variations in the frequency response on the quality of transient reproduction? It was becoming evident that the response in the time-domain was very important as well.

An earlier proposed acoustic model of delayed spectrum signatures seemed to point the way toward a partial understanding of what was involved. [9] This model is a three-dimensional plot of pressure versus frequency versus time after the application of an impulse. It is a plot of the moment-by-moment frequency response as distinct from the steady-state frequency response. When the three-dimensional properties are projected on a two-dimensional surface in pseudo-perspective but in a manner preserving the capability of direct scaling of data without the concern for perspective foreshortening, the format was referred to in that paper as an isometric display.

I had the privilege of making a great many measurements of the response of rooms and of the early direct sound of loudspeaker systems with the results displayed in isometric format. When measuring loudspeaker systems alone it became evident that there is a microsecond-by-microsecond change of the frequency spectrum, both in amplitude and phase, after the application of an impulse. Clearly the loudspeaker has a spectral behavior that is a function of time following the application of a transient. The questions immediately posed by this observation are: what is this effect, how does one measure it and what does it mean to the reproduction of sound?

In the classic definition of frequency response, time is not allowed into the measurement. When all one is concerned about is the amplitude of sound pressure, the time it takes the sound to get from the loudspeaker to the microphone never enters into the picture. Suddenly, with phase a part of the measurement, the experimenter can no longer ignore the time between the application of a voltage and the measurement of the sound pressure a few meters away. Move the microphone backward or forward a few centimeters and the phase response appears different.

Of course the rather constant velocity of sound in air is clearly a part of the measurement. In order to remove the air path from the measurement, it seems self evident that all one needs to do is to subtract a constant time delay. In order to do this one must measure the exact distance from microphone diaphragm to loudspeaker cone, then knowing the speed of sound in air calculate the air path delay and subtract it from the measurement.

When we do this we find a rather puzzling thing. No matter how we try to refine our measurement, the phase response shows that the effective acoustic distance between microphone and loudspeaker is always longer than our physical measurement. The amount of this discrepancy is usually so slight that a casual experimenter might attribute it to experimental error. However, as in many things, this nagging inconsistency is pointing out a much deeper physical property that needs to be uncovered.

When we subtract the actual air path delay from a loudspeaker measurement of frequency response, we are left with what the loudspeaker itself does to the signal. Clearly the loudspeaker is contributing an excess delay.

How can we deal with this delay? Our first reaction is to turn to the use of group delay [16, 17] but when we do that we immediately encounter logical inconsistencies. A headlong application of group delay to loudspeaker measurement discloses that for many frequencies the pressure wave is calculated to occur before the electrical signal giving rise to it! This violation of causality is absolutely unacceptable. Group delay clearly cannot be used. The reason for this is that group delay, defined as the negative of the rate of change of phase with frequency, cannot be used as a measure of time delay when the amplitude response shows a non-zero rate of change with frequency. In fact, it has recently been demonstrated that group delay can never be used as a measure of time delay when applied to any non-trivial minimum-phase network. [10] Exit group delay. Then what is left?

The solution to this problem must be offered by a completely new look at network time delay. Without going into details since it is covered elsewhere, [10, 18] there is one special type of network for which the time delay and group delay coincide. In this network the time delay of each frequency component is always positive and coincides with the definition of group delay corresponding to the negative of the slope of the phase versus frequency. The network is the all-pass transmission function which is strictly non-minimum-phase.

What has this got to do with loudspeakers and other networks which have a frequency-dependent amplitude response? The surprising answer is that properly mixed parallel combinations of all-pass networks can be used to imitate any reasonable minimum-phase response function, regardless of the variations of its amplitude with frequency.

This means that since we know the true time delay properties of the all-pass networks and since signal components can be linearly added if we have a non-interacting parallel combination of such all-pass networks, there is a proper analytical tool that can be used to calculate the amount of time delay that can be expected from a loudspeaker or any network element.

If a signal is fed to a loudspeaker, the resultant response can be equated to that of an ensemble of parallel all-pass elements. Each all-pass element has a perfectly uniform amplitude response and a frequency-dependent phase response with a slope that never becomes poisitive with increasing frequency.

A dip in the loudspeaker frequency response corresponds to a cancellation by destructive interference in the equivalent ensemble response of all-pass elements. A peak is the converse. Since this is a model, we know that the actual loudspeaker, as a network, is not composed of these conceptual elements but the fact remains that if we compute and measure as though it were made of these elements, we will achieve results indistinguishable from the physical case.

To help visualise this concept, each all-pass element in the overall loudspeaker response can be considered as being produced by a hypothetical perfect loudspeaker with a flat amplitude and a flat phase response. As seen from the microphone (or listener) this perfect loudspeaker is then considered to occupy a position in space that varies with frequency in such a manner as to give the equivalent all-pass response. The acoustic position of this hypothetical perfect loudspeaker will always be at or behind the physical position in space that we may assume for the average position of the real loudspeaker.

The complete loudspeaker response, with its complicated amplitude and phase frequency spectrum, can thus be considered to be produced by a swarm or ensemble of these otherwise perfect loudspeakers. When we apply this rather strange model to physical measurement, we find that all the numbers fit properly.

As in any endeavor, it becomes necessary to give a shortened name to a hypothesis. While it may not have a noble ring, I chose to call the ensemble time delay spread of frequency components *time delay distortion*. [10] In later publications by others, this term has been picked up and used to refer to the gross measure of average acoustic delay. That certainly is one aspect of the theoretical concept but clearly there are more detailed aspects. In subsequent work it was demonstrated that time delay distortion is itself a special type of a more general signal deformation which has been called *representation distortion*. [19]

Graphical relationships

The ability to measure phase response and the recognition of time delay distortion puts a greater emphasis on the significance of minimum-phase response and its importance to sound quality. If a loudspeaker (or any device) is minimum-phase, then a simple network correction of irregularities in the amplitude response will automatically correct the phase response. In that case both the frequency domain and the time-domain characteristics will be simultaneously improved. A major presumption at the present time is that this will lead to a more accurate reproduction (assuming no other reproduction property is upset when this is done.) On the other hand, a simple amplitude equalisation of a major aberration that is non-minimum-phase can lead to a definite distortion in transient response. This is because the resultant reproduction will correspond to a time-delay warping of signal components even though the amplitude of those components is correct, a situation which seldom has a physical counterpart in natural sound and hence may stand apart from natural sound.

It becomes important, then, to be able to determine when a response is minimum-phase. The technical relationship between the amplitude and phase responses for a minimum-phase system is that of Hilbert transformation [10, 20, 21]

$$\alpha(\omega) = \frac{1}{\pi} P\int_{-\infty}^{\infty} \frac{\varphi(x)}{\omega - x}\, dx$$

$$-\varphi(\omega) = \frac{1}{\pi} P\int_{-\infty}^{\infty} \frac{\alpha(x)}{\omega - x}\, dx \qquad (4)$$

where P indicates that the principal part of the integral is to be taken at the pole ω and x is a dummy variable.

If one has a computer at his disposal he need only program it to compare the measured amplitude and phase responses as Hilbert transform pairs and have it print out the discrepancy. Most of us, however, do not have access to a computer, so a set of graphical relationships which will let us look at separate plots of responses and tell whether the system is minimum-phase will be most helpful. That observation precipitated the following derivation. [10]

The integrals are the counterpart of Cauchy's differential relations on the frequency axis. [10] Because for a minimum-phase transfer function there are no zeros or poles in the right half s-plane, the Cauchy-Riemann equations are valid within and on the boundary of the right half s-plane if we use the logarithmic representation of the frequency response, [10]

$$\ln H(s) = \alpha(s) + i\varphi(s) \qquad (5)$$

where

$$s = \delta + i\omega$$

The result of this is that at a local minimum or maximum in the transfer function, a frequency of maximum curvature of amplitude corresponds to a point of inflection of phase and conversely. There is a skew symmetry, dictated by causality, that gives a preferred slope/extremum relation as follows:

when $\partial^2 \alpha/\partial\omega^2$ is a maximum positive, then $\partial\varphi/\partial\omega$ is a maximum positive,

while when $\partial^2 \varphi/\partial\omega^2$ is a maximum positive, then $\partial\alpha/\partial\omega$ is a maximum negative.

It also follows that points of inflection correspond rather closely to frequencies of general maximum curvature in the complementary plots. The original paper should be referred to for a derivation and discussion of the limitations in these relationships.

It should be noted from these relationships that at a frequency of stationary phase (i.e., $\partial\varphi/\partial\omega = 0$), the amplitude of a minimum-phase function has its maximum rate of change with frequency. This clearly invalidates the reckless use of the principle of stationary phase which one frequently sees in "proofs" that the time delay is single valued and equal to group delay. Causal time delay is not equal to the group delay in a minimum-phase system.

There is another variation on the Hilbert transform which can be used to infer graphical relationships, [21]

$$\alpha(\omega) = \frac{-1}{\pi}\frac{d}{dw}\int_{-\infty}^{\infty} \varphi(x)\ ln\ |1 - \frac{\omega}{x}|\ dx\ ,$$

$$\varphi(\omega) = \frac{1}{\pi}\frac{d}{d\omega}\int_{-\infty}^{\infty} \alpha(x)\ ln\ |1 - \frac{\omega}{x}|\ dx \qquad (6)$$

In this analysis we are discussing relationships and how they will appear to the eye of the engineer who has just plotted the separate amplitude and phase response of a loudspeaker and has both plots in front of him. His concern rests with determining, by looking at the plots, whether the device is minimum-phase and if not, then at what frequencies the non-minimum-phase properties occur.

With that more restricted view we can observe that the logarithmic quantity under the integral in equation 6 behaves analogously to a generalised function which gives enormous weight to the function inside the integral at the singularity $\omega = x$ and which drops off rapidly for values of ω farther away from x. Its action, much like an impulse, tends to pull the function outside the integral so long as the function is reasonably well-behaved in its derivatives.

In other words, for a minimum-phase function, we have the approximate relations,

$$\alpha(\omega) \simeq -\frac{1}{\pi}\frac{d}{d\omega}\varphi(\omega)$$

$$\varphi(\omega) \simeq \frac{1}{\pi}\frac{d}{d\omega}\alpha(\omega) \qquad (7)$$

We can expect these relationships to be more accurate for smooth trends in amplitude or phase.

From the above we can produce the sketch of graphical relationships which we should see for a minimum-phase loudspeaker. This is shown in fig. 1 for both a logarithmic and a linear frequency scale.

Bode's [15] famous rule that (only for a minimum-phase network) the phase approaches a constant value for an asymptotic gain slope on a logarithmic frequency basis and that the gain and phase are approximately related to each other as derivatives on a logarithmic frequency basis is included in this sketch. This latter observation of Bode came out of an integral he derived that is the logarithmic based equivalent of equation 6.

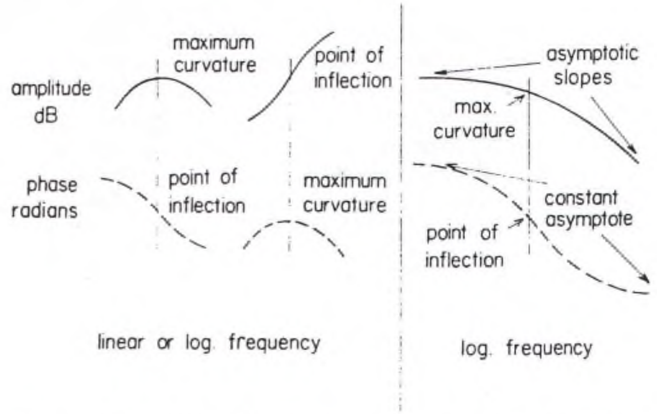


Fig 1. Characteristic amplitude (solid) and phase (dotted) plots which may be used to identify a minimum phase response.

There is one interesting observation that can be made from these equations. From equation 7 we see that the group delay of a minimum-phase network will be similar to the amplitude response of that network, that is, the shapes will look alike. This leads to a shortcut in analysis since group delay can usually be measured more readily than phase response, particularly for systems with inherent complicated delay behavior such as phonograph cartridges. If a measurement of amplitude response and group delay have similar properties as a function of frequency, then the device is principally minimum-phase.

Time response

In a certain sense, a loudspeaker is intended to produce at the location of a listener, a sound pressure wave that is, in some measure, a replica of the electrical signal applied to the loudspeaker. One of the most important properties of loudspeakers relates to the processing of aperiodic signals.

The time delay distortion concept suggests that we can imagine sets of delayed reproductions of a program which each have the peculiar property that the time delay is frequency dependent. This will lead to a time smear of any reproduction.

It seems evident that a test stimulus for time-domain measurement should be an impulse. A true impulse is an abstraction which no one can generate in a laboratory. It was in fact considered a pathological entity by mathematicians until they succeeded in formalising the impulse as a generalised function. [22] What the engineer normally means by an impulse is a sufficiently narrow spike of signal which has a more or less uniform frequency spectrum over the range of interest for the test.

These are serious problems raised whenever an experimenter uses a narrow spike of high voltage to obtain the impulse response of a loudspeaker and from this compute the frequency response by means of the Fourier transformation. Fourier transform relations are only true in a linear system and it is incumbent on the experimenter to show that all loudspeaker drivers are truly operating in a linear mode when presented with a large spike of voltage. This author has had considerable experience measuring frequency-domain properties by a direct method and has observed many cases where thermal and mechanical limitations enter in a different fashion in modifying the response for differing ratios of peak to average applied power. Be that as it may, let us assume that for testing purposes it is possible to measure the impulse response accurately.

What does the measured impulse response look like? The answer quite frequently is that it is a confusing pattern of decaying oscillatory components. We would like to extract meaningful information from this and one such attempt will now be described.

In the pressure waveform of an impulse response there are many instants when the pressure passes through "zero", that is, the local equilibrium pressure in the measuring environment. Does this mean there is no "sound" at these moments? The answer is, no, it only means there is no pressure excess at these moments. The fact that there are pressure rates of change at these times of zero pressure excess indicates that there is energy density available in perhaps another form and indeed there is. Again, we need to probe deeper if we wish to understand what is going on.

The "philosophy" can be found elsewhere [23] but the fact is we need to wrench ourselves away from another habit pattern in analysis. The frequency response of a loudspeaker is now universally recognised as a quantity with an amplitude and a phase. I would like to stress the fact that it is also of analytical benefit to think of the time response as possessing an amplitude and phase, or, what is the same thing, inphase and quadrature components which can be combined in Pythagorean fashion to make an amplitude and phase.

Signal analysts will instantly recognise this as a call to concern ourselves with the analytic signal representation of the time response. [24] From energy balance considerations it now appears evident why Gabor was forced to develop the analytic signal. I heartily recommend that pioneering paper as background reading for anyone involved in signal theory.

When we apply this concept to loudspeakers we are led back to our old friend the Hilbert transform. We can look upon the "impulse response" as the "x-axis" projection of the complex analytic signal. There will be a conjugate "y-axis" projection which it turns out is the Hilbert transform of the impulse response. The magnitude of the complete analytic signal corresponding to the impulse response is obtained as the square root of the sum of the squares of the two projected components. These quadrature components define a right triangle with the hypotenuse equal to the magnitude. The magnitude is the length of the phasor whirling about a point which moves along the time axis much as the tip of a propellor describes a pattern about the axis of the propellor shaft as an aircraft flies through the air. Only, in our case, the length of the propellor changes along the flight path as does the angular rate.

In order to understand how the analytic signal representation may be used in analysis, it is necessary to introduce a concept developed for that purpose and presented elsewhere. [23]

If we define the total energy density for the state variable x as E(x) and presume that it is equated to the sum of a potential energy density, V(x) and a kinetic energy density, T(x), we have,

$$E(x) = V(x) + T(x) \tag{8}$$

We then define a complex function in the following manner,

$$\sqrt{E(x)} = \sqrt{V(x)} + i\sqrt{T(x)} \tag{9}$$

Now, what do we know about the total energy placed in the loudspeaker by our test signal? We know it is finite,

$$\int_{-\infty}^{\infty} E(x)\ dx = \int_{-\infty}^{\infty} |\sqrt{E(x)}\,|^2\ dx < \infty \tag{10}$$

The complex function $\sqrt{E(x)}$ is of class $L^2(-\infty, \infty)$, that is, the Lebesgue integral [21] of the square of its modulus over the entire space of the state variable is finite. It is known that analytic functions of the form $\sqrt{E(x)}$ which are of the class L^2 have complex components which are conjugate. [21] Thus we can infer that $\sqrt{V(x)}$ and $\sqrt{T(x)}$ are Hilbert transforms of each other.

Returning to acoustic measurement, the pressure is proportional to the square root of the potential energy density. [23, 25] If the pressure impulse response is one term of the analytic signal, then the complete analytic signal represents the magnitude and partitioning of the energy density components.

The significance of the analytic signal to loudspeaker measurements now becomes evident. When an electrical impulse is fed to a loudspeaker the sound does not instantly emerge. Furthermore, the total acoustic signal will arrive in discrete bundles for those components which are the result of diffraction and reflection in the loudspeaker and its enclosure. The magnitude of the analytic signal will show a peak for each discrete component of energy arrival. The rate of change of the phase of the analytic signal bears a relation to the spectral distribution of the separate energy arrivals. When the analytic signal is used, the mysterious bumps and wiggles of the scalar impulse response are quickly revealed as discrete arrivals in signal energy and the subsequent decrement in energy for each arrival.

As a side benefit of this energy relationship, we know
h
that, [21]

$$\int_{-\infty}^{\infty} |\sqrt{V(x)}\,|^2\ dx = \int_{-\infty}^{\infty} |\sqrt{T(x)}\,|^2\ dx = \tfrac{1}{2}\int_{-\infty}^{\infty} |\sqrt{E(x)}\,|^2\ dx \tag{11}$$

Thus an energy calculation based solely on either a measured pressure response or a measured particle velocity response will be correct to within a factor of two even if the analytic signal is not used. Also the parameter x is not restricted in this relationship, which is quite consistent with Parseval's equation. [26] This becomes important in mapping a signal representation into other levels of dimensionality as will be touched upon later.

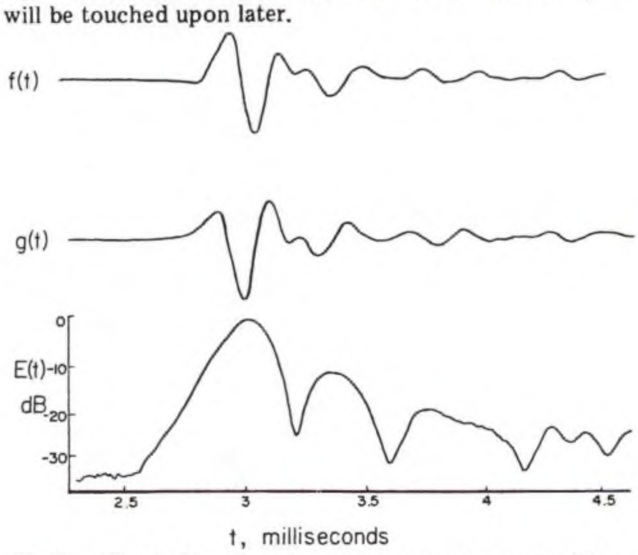


Fig 2. Typical loudspeaker measurement showing the impulse response, f(t) its Hilbert transform, g(t), and the total energy density, E(t).

Figure 2 is a plot of a typical measurement on a loudspeaker showing the impulse response, its Hilbert transform (which this author has called the doublet response [23]) and the total energy density computed from these components. The measurement of the total energy density as a function of time is now a standard measurement in the evaluation of loudspeakers for a consumer publication. [27] In all cases the peaks in the energy response have been identified with discrete arrivals due to physical structure and inherent loudspeaker problems, leading in a number of cases to an instant correction of the problem. The impulse response, while it contains all the necessary information, is in a form that makes it very difficult to perform a similar diagnosis.

Room measurement

Up to this point the assumption has been made that the loudspeaker is measured apart from the environment of normal listening. It is necessary to bring the room into the measurement.

It is here that the characterisation of a loudspeaker becomes burdensome. Furthermore, no two listening situations will be identical, which means that an engineering choice must be made in determining the type of measurement to be made.

There is some experimental evidence which suggests that the first few milliseconds of "early sound" of loudspeaker reproduction should be used to characterise performance. It is common experience that certain loudspeaker types have a characteristic coloration of spectral balance that gives rise to a change in timbre of the reproduced sound. What is it that allows a person to identify the coloration so well that he can often identify the loudspeaker by manufacturer simply by listening to a reproduction of sound by that loudspeaker, even in an unfamiliar acoustic environment? If we search for timbre-related characteristics, we find that on the average the pitch of a pulsed tone appears established within the first 13 milliseconds for most observers. [28] This minimum time period is surprisingly independent of the pitch that is perceived.

A measurement has now been in commercial use for several years which uses this observation as a basis for determining the "room response" of loudspeakers. [27] The loudspeaker is placed in a preferred listening position in a room having a 2.5 meter floor to ceiling height, which corresponds reasonably well to the dimension found in many homes in the world. The measuring microphone is placed at a position where experimental evidence has shown the average listener may be seated, three meters from the front of the speaker enclosure and one meter above the floor. The signal applied to the loudspeaker has a spectral density uniform throughout the frequency range of interest and an autocorrelation function whose time width is the inverse of the width of the frequency spectrum. This condition gives the greatest time acuity for a given spectral measurement.

The first sound arrives at the microphone approximately nine milliseconds after the signal is applied, for a three meter air path. The method of measurement used for this particular test is Time Delay Spectrometry, which will shortly be described. However, the measurement will first be described in terms relating to the equivalent use of an impulse test signal. The output of the measuring microphone contains all signals, including first arrivals and multiple reflections. A time gate is opened in this microphone signal for the first 13 milliseconds immediately after the first sound arrival so as to allow only this part of the sound to be passed to the processor. This gate thus extends from 9 to 22 milliseconds after application of the loudspeaker voltage.

The gated microphone voltage is multiplied by a time weighting function which has a gain progressing from a low value at 9 milliseconds to a maximum value halfway through the gate then progressing down to a low value at the 22 millisecond cut-off. The purpose of this weighting, also called apodisation, is to reduce the equivalent sidelobe clutter in the subsequent Fourier transform, performed to yield the frequency spectrum of this sound.

This frequency spectrum is the response of the loudspeaker at the listening position which occurs during the timbre-establishing period of 13 milliseconds. The floor, walls, and ceiling enter into this measurement in the same manner as in normal listening. Later scattered arrivals are effectively eliminated and are presumed to contribute the sense of room ambience, while loudspeaker coloration of the "early sound" is contained in the measurement.

The width of the time gate can, of course, be opened up or restricted to any value one chooses. This particular gate width was chosen for measurement of published results for loudspeaker evaluation because it is long enough to satisfy a psycho-acoustic parameter, yet short enough to eliminate the details of room furnishing and room length and width from influencing the measurement, a consideration important for widespread duplication of results.

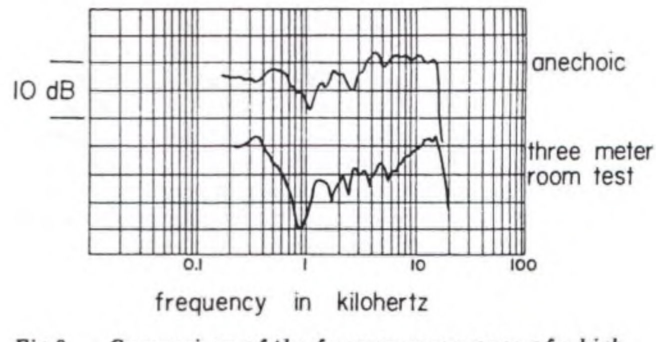


Fig 3. Comparison of the frequency response of a high quality loudspeaker measured under free field anechoic conditions and for the first 13 milliseconds of sound arrival in a typical listening situation in a room.

Figure 3 is an example of the anechoic response of a loudspeaker and the room response of the same loudspeaker measured by this method. There are many cases, as in this example, in which the anechoic frequency response bears little relation to the measured room response. It is gratifying to the assumptions that went into the design of the room measurement to note that the subjective evaluation of the loudspeaker "timbre" is usually in close agreement with this measurement.

Time delay spectrometry

Chronologically, this next subject did not follow from the items just discussed. Rather it preceded them. The greatest obstacle to the test of a loudspeaker by itself is the elimination of the acoustic interaction of the room and environment in which the test is performed. In large laboratories this is done by means of an anechoic room, a very expensive facility in which the walls, floor, and ceiling are designed to absorb sound rather than allow reflections to occur back into the test area.

Under certain conditions it is possible to perform the testing out of doors and away from large physical objects that might interfere with a measurement. Not only is this method subject to the vagaries of local weather but with the rising decibel level of many urban areas the ambient noise level can militate against an accurate test at moderate sound levels.

Additionally, there are circumstances when a measurement is needed of the room and its acoustics without the masking effect of the source speaker. Such a case can arise if one is interested in the reflection coefficient of a particular segment of an acoustic surface. Or in some cases a loudspeaker may depend for its operation on a portion of the listening room and this negates the use of an anechoic environment.

A method of generalised analysis was invented to meet the needs of anechoic measurements in a reverberant environment. This method is called time delay spectrometry (TDS). [9, 29] While originally developed for loudspeaker testing *in situ*, the measured properties of actual loudspeakers quickly revealed a need for a more extensive investigation of testing needs.

There are three practical considerations that arise out of the general propagation of signals. First, the velocity of propagation is finite, which means there will be a time delay between the application of a signal and a system response to that signal. Second, the velocity may not always be independent of either frequency or geometric configuration, such as dispersive seismic waves, doppler offsets due to relative velocity between transmitter and receiver, or because of medium motion. Third, many cases arise in which there is multipath propagation between a source and a receiver.

Time delay spectrometry recognises these considerations and in fact incorporates them in its formulation. The basic concept of TDS can be illustrated as shown in fig. 4. A constant-amplitude source signal is used which has a defined phase angle versus time, such as curve A. The general form of this signal is exp ($i\theta$ (t), where θ (t) can be anything whatsoever, linear, quadratic or even pseudo-random.

The time rate of change of the phase angle of a constant-amplitude phasor is called the instantaneous frequency. [30] For the sake of illustration, this parameter is called "frequency" in fig. 4.

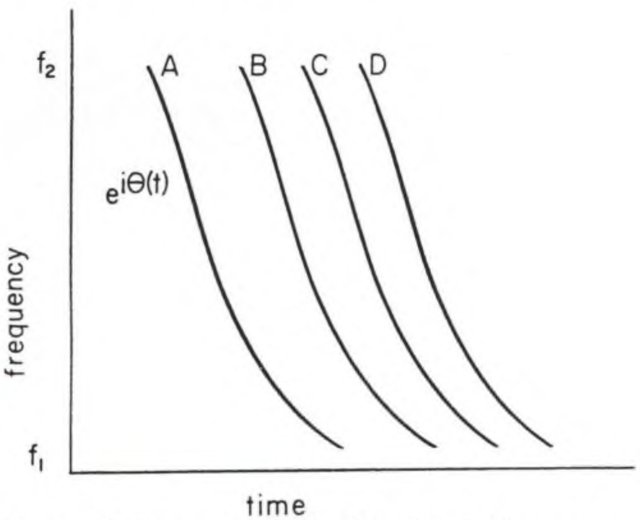


Fig 4. Illustration of the principle of time-delay spectrometry. A constant amplitude coherent signal source, A, sweeps a known trajectory of frequency versus time. This signal is used to drive a loudspeaker for test in a room. The sound picked up by a microphone will consist of a direct sound, B, and later multiple arrivals, C and D. The frequency versus time of all arrivals will have the same trajectory as source A but be delayed in time of arrival depending on path length.

If the signal A is a source signal at a transmitter, then the signal intercepted at a receiver will generally consist of a first arrival, B, and subsequent multiple arrivals, C, D, etc. At this point it is instructive to change the display to that of fig. 5(a). The source signal A can be considered to generate a relative time reference of zero delay for all versions of that signal.

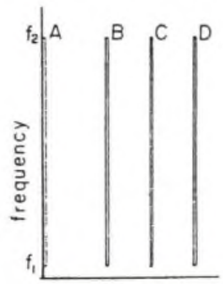


Fig 5a. All multipath arrivals, for a nonmoving, non-dispersive medium, have a constant delay relative to the source A, when the excitation shown in figure 4 is used.

The curves A, B, C, etc. now appear as shown if the medium of propagation is non-dispersive and there is no doppler offset between transmitter and receiver. The signal A is defined as perfect. The responses B, C, etc., will not be perfect but will spread in relative time and distort in frequency response.

If we wish to isolate the first signal B (i.e., measure the anechoic frequency response) we can symbolically cut along the dotted line shown in fig. 5(b) and reject all signals outside this boundary. This is our gate.

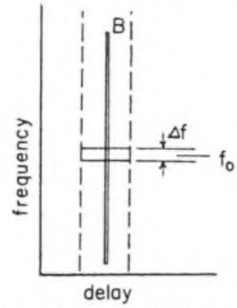


Fig 5b. Isolation of any multipath component, such as B, may be obtained by time gating the delayed spectrum around the arrival delay of that component (shown as dotted line) and rejecting delayed arrivals outside that gate. Coherently processing the energy within the strip Δf will give the value of the complex frequency spectrum of B at the central frequency f_o. Scanning f_o through all frequencies, performed by the time trajectory of A in fig 4, gives the complete frequency response for component B.

Because we have generated a signal of known characteristic over the frequency range from f_1 to f_2, the complex value of the signal energy in each incremental strip as Δf will approximate the value of the frequency response of B at the respective values of frequency. We have caused the desired signal B and the undesired components C, D, etc. to become separated by virtue of the time delay between them and have provided a measurement of the frequency spectrum of the selected signal. This is the historic reason for the name time delay spectrometry.

If the medium is dispersive, then the dotted line to be cut is as shown in fig. 6 for making accurate frequency measurements. The width of the relative time acceptance need only be large enough to encompass the time smear (time delay distortion) of the system under test. In the case of the room test described above, the width is set at 13 milliseconds.

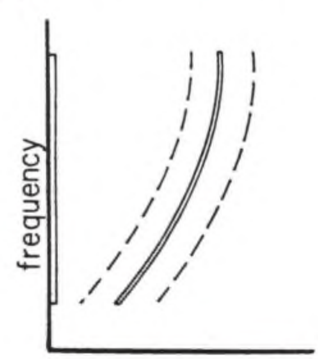


Fig 6. A dispersive medium will alter the signal delay in a frequency dependent manner. Proper isolation of a multipath component, such as B, is made in time-delay spectrometry by time gating in a manner consistent with that dispersion (shown as dotted line).

Time delay spectrometry utilises a phase and amplitude coherent signal source and those familiar with coherent signal processing theory will recognise that the optimum gate (dotted line) and its associated apodisation for making a frequency spectrum analysis is a complex conjugate inverse filter for the signal B. This is the frequency-domain equivalent for a time-domain concept well known in communication theory. When this is done there is a maximum separation of B from the interfering multipath components and room noise interference on a mean square error basis. This allows us to make a useful measurement even when the loudspeaker signal is of lower sound pressure level then the mean average level in the ambient sound.

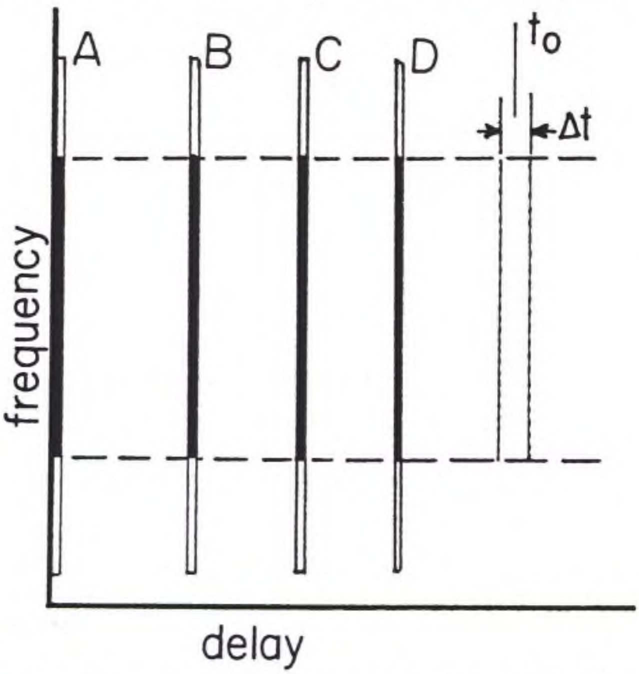


Fig 7. Illustration of the manner in which time delayed components may be processed to yield the analytic signal corresponding to the loudspeaker impulse response. A block of frequencies is selected (shown by dotted lines) and all components outside that frequency range are rejected. Coherent processing of the components within the narrow strip Δt will yield the analytic signal value corresponding to the epoch t_o after the excitation signal A. Scanning Δt through the range of delays then gives the continuous analytic signal.

The spectrum referred to in time delay spectrometry is also the time spectrum as can be seen from fig. 7. If we cut along the dotted line shown here we have isolated the frequency range of interest. To determine the time-domain spectrum of this segment, we coherently sum along the segment Δt and slide it along the delay axis.

The process of cutting and sliding is, of course, gating and scanning which can readily be done in either analog or digital fashion depending upon the instrumentation desired. For loudspeaker testing, a signal is applied which sweeps a known trajectory of frequency versus time (as in fig. 4). The microphone picks up the delayed acoustic signal and passes it through a narrow bandwidth tracking filter. The center frequency of this filter is tuned to follow the precise frequency versus time behavior of the loudspeaker voltage but with a time delay equal to the path length delay of the component to be measured. In this manner the filter is tuned to the exact frequency of the received signal at every moment. Other signals, taking longer or shorter delay times, will be picked up by the microphone but they will not pass through the filter because their frequency will never coincide with the filter acceptance frequency.

The output of the tracking filter contains the desired component of a multipath arrival. The frequency spectrum of the signal path has been converted into a voltage that changes with time. The frequency co-ordinate has been conveniently changed into a time co-ordinate.

Because we use a signal that is characterised by an instantaneous frequency uniquely related to each moment of time, we can use the fact that frequency and time can be exchanged in their role. Thus while we normally consider the causal world in which we generate a signal with a precisely known frequency as a function of time, we could just as easily think of the converse, non-casual, situation where the time metric is determined as a function of frequency.

The exchange of frequency and time leads to certain simplifications if we choose to use analog instrumentation. For one thing, the tracking filter bandwidth plays the role of the time gate in selecting the desired component, and windowing (apodisation) [31] is automatically accounted for by the type of filter selectivity chosen. For another, the frequency spectrum is presented without the need of Fourier transformation.

Whether we use digital or analog methods in time delay spectrometry, time-domain properties are best obtained as a Fourier transformation from the frequency-domain. This offers no difficulty because the frequency spectrum is a time-dependent voltage and need only be multiplied by a sine wave of the proper frequency and integrated for the length of time corresponding to the frequency range of interest.

A side benefit of this is that it is easy to perform non-causal experiments where the time metric may be non-uniform, stop or even run backwards. For example, if the source signal only extends from frequency f_1 to f_2, the non-causal single-sided frequency spectrum is obtained. The conversion from frequency to time-domain then gives the analytic signal directly without the need for a separate Hilbert transform. This is the method used to obtain the published data on the time-dependent magnitude of the analytic signal, which is referred to as the energy-time plot. [27]

Conclusion

I would like to close this discussion with a few observations concerning past, present and future work in loudspeaker evaluation. I have attempted to present a summary of some of the more important changes in loudspeaker signal analysis in which I have been privileged to participate.

The loudspeaker has always represented a severe challenge to the engineer. The reason for this, from the engineer's point of view, was presumed to be mere instrumentation difficulties. I do not believe that is the whole reason.

As many technically trained people are prone to do, I naively presumed when I first began analysing loudspeakers almost a decade and a half ago, that I could bring contemporary communication theory to bear and simply overwhelm the poor loudspeaker with technology. After all, a simple search through the literature revealed that the instrumentation then in use for loudspeaker testing was conceptually antiquated.

I soon found the error in my thinking. The evaluation of the acoustics of loudspeakers and the room containing them proved to represent a microcosm of all the difficult problems in wave propagation. A wavelength range of over ten thousand to one is bad enough but the physical extent of the important dimensions in a single experiment range from below one thirtieth wavelength to greater than thirty wavelengths for many practical loudspeaker systems. The loudspeaker as a communication medium is both dispersive and absorptive and even without the room is best characterised as a multipath medium when one begins measuring the effect of tweeter, midrange and woofer. Furthermore, much of the best linear mathematics melts away before the obvious non-linearities possessed by all loudspeakers.

What I actually found was that I could learn from the loudspeaker how to improve the holes in my own background of communication theory. A case in point is group delay, a concept which I frankly never understood in college studies but which I had been assured was properly defined "somewhere". Everyone knows that there is some actual time delay for a signal passing through a physical system but when group delay, which pretended to be that delay, was applied, non-causal solutions frequently resulted. Textbook derivations made a great deal of fuss about explaining that "anticipatory transients" did not really exist and could be explained as resulting from approximations in the presumed frequency response. It was vaguely hinted that group delay could not be used near frequencies of absorption. But how close? And was it valid farther away? If so, then how far away?

As long as such problems lurked only at the fringes of measurement, no-one seemed bothered. But suddenly with loudspeakers, I found myself deep in such considerations. The problems could not be ignored, particularly since they seemed to be at the heart of considerations of the importance of phase response to quality of reproduction.

I did finally find a solution to the mystery of time delay in a dispersive absorptive medium but only after I was forced to do so by practical considerations of loudspeaker reproduction. This solution has since been folded back into electromagnetic problems to obtain correct answers. Group delay, it turns out, is never equal to causal time delay in a minimum-phase system. One should not compute the time delay of a loudspeaker by a measurement of group delay.

That is now part of the past work in loudspeaker evaluation. I think it is fair to say that much of loudspeaker testing has caught up with linear communication theory. Pseudo-random sequences are now routinely used with autocorrelators to measure the impulse response. [32] The Fast Fourier transform is in full flood in frequency response measurement. [33] The Hilbert transform relations are now used for both the complex frequency spectrum and the analytic signals. [10, 23, 34] Inverse filters, deconvolution techniques [35] transversal filters [30, 36] and even a primitive Kalman filter [37] have been used with loudspeakers and the correction of room acoustics.

That brings us up to the present work in the testing and evaluation of loudspeakers. As the audio industry continues in this game of "catch up" with the fund of knowledge accumulated over the years in other fields of communication, the question arises, "where do we go from here?" Will we reach a golden plateau where we finally can measure a loudspeaker with these techniques and know exactly how it should sound. I think not.

I say this not as a pessimist but as one worker in the field of analysis who feels that we have much more to learn from psycho-acoustics and the lowly loudspeaker (I keep referring to the loudspeaker but remember it is only one link in a chain of reproduction and all the links have to be known better). We know the loudspeaker is non-linear; thus our present fund of linear analysis could soon become bankrupt. Also, we must remember that the function of a loudspeaker is to act as the final interface with the most inaccessible and complicated device of all, the human listener.

From my personal standpoint the present work in the analysis of the loudspeaker and its role in reproduction lies in coming to grips with both the linear and non-linear aspects of the listening experience. Some results have come out of this present analysis which I find startling. [19] For example, we are accustomed to considering the time-domain and frequency-domain as the only representations of a signal that may be used for analysis. This, it turns out, is not the case. The time response is a representation in one-dimension, time. The frequency response is also a representation in one-dimension, frequency. We can map either of these representations into the other by means of the Fourier transformation. Neither representation contains more information than the other, they merely have a different co-ordinate basis. Apparent end of story. But not so. We can also map these one-dimensional representations upward to two or more dimensions. The time and frequency-domain are thus only two of an indefinitely large number of possible domains we might use. We may have been casting our analytical gaze on the ground all this time, not realising we could look upward to the heavens.

When this concept of interdimensional mapping is applied to loudspeaker reproduction we find that it is necessary to treat the loudspeaker and its environment as a system, rather than as separate entities. This brings a whole new set of objective measurements to the fore.

The electrical current passing into the loudspeaker terminal is, as conventionally considered, a signal that is a function of time, the one-dimensional time-domain representation. However, it is also a signal that represents a spatial, spectral, temporal and intensity distribution of a virtual sound image. When we apply the rudimentary concepts of multi-dimensional mapping to the loudspeaker, we find that the signal parameters we must measure are somewhat different from those now considered *de rigueur*. For example, a classic distortion test for non-linearity involves the measurement of the harmonic distortion terms generated from the application of a pure sinusoid. As in early frequency response measurements, the amplitude of the harmonic set is considered sufficient. Well, it turns out it is not sufficient to measure amplitude when the newer analysis is invoked. The geometrical features of the higher dimensional representation depend heavily on the exact behavior of the harmonic species, including their phase angle relative to the fundamental. Furthermore, the interplay of the parameters of time of occurrence, spectral content and relative intensity play a very important role. In standard harmonic distortion analysis it is not necessary to know if the fundamental and its harmonics are shifted in relative time of occurrence as a function of total energy. The newer analysis says we must measure for this and when this is done for loudspeakers it is found that many loudspeakers actually do shift the relative time with signal level!

What this present analysis means is that we are not finished when we obtain a measurement of the analytic signal response to an impulse no matter how elegant the measurement may appear. We have just begun. One of the things we need to do is to detail the interplay of various aspects of that response and treat the loudspeaker and listening environment as a system, because this system as a whole may warp the message being transmitted to the listener.

Work is now in progress that is directed towards creating some new measurements that appear to be necessary in order to characterise loudspeaker reproduction objectiv ely. The task lying ahead of us is quite clear — develop a better understanding of the psycho-acoustics of perception and learn how to correlate what can be objectively measured against that perception. This is not a vain hope but shows some promise of coming to pass. The end product of the loudspeaker is not simply a sound field but includes those properties of reproduction necessary tc create the proper acoustic illusion in the mind of the listener. We may yet be able to measure what we hear.

References

1. Le Croissette, D.H. and Heyser, R.C., "Transmission Ultrasound Soft Tissue Visualisation", Proc. 3rd Int. Conf. on Med. Phys., 1972, p. 35.5
2. Heyser, R.C., Besseling, L.N. and Le Croissette, D.H., "Transmission Ultrasonograms using Time Delay Spectrometry", Proc. 25th Conf. on Eng. in Biol. and Med., 1972, Vol. 14, p. 138.
3. Heyser, R.C. and Le Croissette, D.H., "Transmission Ultra-

sonography", 1973, Ultrasonics Symposium Proceedings, IEEE Cat. No. 73 CHO 807-8SU, (1974), p. 709.

4. Heyser, R.C. and Le Croissette, D.H., "A New Ultrasonic Imaging System using Time Delay Spectrometry", *Ultrasound in Medicine and Biology*, Vol. 1, 1974 p. 119-31.

5. Fishman, L.S., Heyser, R.C. and Le Croissette, D.H., "Ultrasonic Transmission Measurements on Human Brain Sections", *Radiology*, Vol. 112, No. 1, 1974, p. 211-3.

6. Wiener, F.M., "Phase Distortion in Electroacoustic Systems", *J.A.S.A.*, Vol. 13, October 1941, p. 115.

7. Ewaskio, C.A. and Mawardi, O.K., "Electroacoustic Phase Shift in Loudspeakers", *J.A.S.A.*, Vol. 22, July 1950, p. 444.

8. Stroh, W.R., "Phase Shift in Loudspeakers", *Trans. IRE*, (Audio), September 1959, p. 120.

9. Heyser, R.C., "Acoustical Measurements by Time Delay Spectrometry", *J.A.E.S.*, Vol. 15, 1967, p. 370.

10. Heyser, R.C., "Loudspeaker Phase Characteristics and Time Delay Distortion" (in two parts), *J.A.E.S.*, Vol. 17, 1969, p. 30.

11. Mawardi, O.K., "A Physical Approach to the Generalised Loudspeaker Problem", *J.A.S.A.*, Vol. 26, 1954, p. 1.

12. Hilliard, J.K., "Notes on How Phase and Delay Distortions Affect the Quality of Speech, Music, and Sound Effects", *Trans. IEEE*, (Audio), March 1964, p. 23.

13. Tappan, P.W., (Editor's Corner), *Trans. IEEE*, Vol. AU-12, March 1964, p. 21.

14. Tappan, P.W., "More on Phase Distortion", (Editor's Corner), *Trans. IEEE*, Vol. AU-3, January 1965, p. 2.

15. Bode, H.W., "Network Analysis and Feedback Amplifier Design", Van Nostrand, New York (1945).

16. Brillouin, L., "Wave Propagation and Group Velocity", Academic Press, New York (1960).

17. Havelock, T.H., "The Propagation of Disturbances in Dispersive Media", Cambridge University Press (1914).

18. Heyser, R.C., "The Delay Plane, Objective Analysis of Subjective Properties", (in two parts), *J.A.E.S.*, Vol. 21, 1973, p. 690.

19. Heyser, R.C., "Geometrical Considerations of Subjective Audio", *J.A.E.S.*, Vol. 22, 1974, p. 674.

20. Lee, Y.W., "Synthesis of Electric Networks by means of the Fourier Transforms of Laguerre's Functions", *J. Math and Physics*, Vol. 11, 1931, p. 83.

21. Titchmarsh, E.C., "Introduction to the Theory of Fourier Integrals", Oxford, (Clarendon Press), 2nd Ed., (1967).

22. Lighthill, M.J., "Fourier Analysis and Generalised Functions", Cambridge University Press, (1959).

23. Heyser, R.C., "Determination of Loudspeaker Signal Arrival Times" (in three parts), *J.A.E.S.*, Vol. 19, 1971, p. 734.

24. Gabor, D., "Theory of Communication", *Proc. IEE*, Vol. 93, 1946, p. 429.

25. Menzel, D.H., "Fundamental Formulas of Physics", Prentice-Hall, New York (1955), p. 356.

26. Butzer, P.L. and Nessel, R.J., "Fourier Analysis and Approximation", Vol. 1, Academic Press, New York, (1971), p. 209.

27. Heyser, R.C., "Breakthrough in Speaker Testing", *Audio*, November 1973, p. 20.

28. Olson, H.F., "Musical Engineering", McGraw Hill, (1952), p. 250.

29. Heyser, R.C., "Time Delay Spectrometer", U.S. Patent 3,466,652, (1969).

30. Franks, L.E., "Signal Theory", Prentice Hall, (1969), p. 81.

31. Blackman, R.B. and Tukey, J.W., "The Measurement of Power Spectra", Dover Publications, New York (1958).

32. Schomer, P.D., "Measurement of Sound Transmission Loss by Combining Correlation and Fourier Techniques", *J.A.S.A.*, Vol. 51, 1972, p. 1127.

33. Bertram, S., "On the Derivation of the Fast Fourier Transform", *Trans. IEEE*, Vol. AU-18, March 1970, p. 55.

34. Berman, J.M., "Loudspeaker Evaluation using Digital Techniques", 50th. Convention Audio Eng. Soc., March 1975.

35. Schroeder, M.R., "Computers in Acoustics: Symbiosis of an Old Science and a New Tool", *J.A.S.A.*, Vol. 45, 1969, p. 1077.

36. Kallmann, H.E., "Transversal Filters", *Proc. IRE*, Vol. 28, July 1940, p. 302.

37. Kalman R.E. and Bucy, R.S., "New Results in Linear Filtering and Prediction Theory", *ASME J. Basic Eng.*, March 1961.

Reprinted from *JAES*, vol. 24, no. 8, pp. 660–667, 1976 October.

COMMUNICATIONS

PERSPECTIVES IN AUDIO ANALYSIS: CHANGING THE FRAME OF REFERENCE, Part I

RICHARD C. HEYSER

Jet Propulsion Laboratory, Pasadena, CA

The geometrical consequences of the process of "changing the frame of reference" are considered from the standpoint of solving problems of an objective nature as well as the inferences to be drawn for subjective perception. A principle of alternatives is presented and used for the development of a mathematical expression for mapping a functional representation from one frame of reference to any other, regardless of the change in dimensionality or units of measurement involved. It is shown that Fourier transformation cannot change the number of dimensions of a representation and is a special case of more general maps. A geometrical interpretation of the meaning of distortion is presented and one aspect of this tied to mapping a problem to a sequence space for measurement. The processes of scanning and imaging are associated with changes in the frame of reference, and definitions are presented which show, among other things, that a hologram is a special case of a more general representational map which is defined as a holomorph. Other reference-changing concepts are considered from their meaning to audio analysis, including self-repairing maps, metrizability, even- and odd-dimension forms, apodization, uncertainty relations, and spatial filtering. Predictions are presented for advanced methods of intercepting and recording sound which come directly out of this geometric analysis, and examples are given of the type of processing required for recording signals in a one-dimensional holomorph that could be preserved as a modulated groove on contemporary disk recording, and a new type of two- and three-dimensional holomorph.

It has been said that geometry is an instrument. The comparison may be admitted, provided it is granted at the same time that this instrument, like Proteus in the fable, ought constantly to change its form.

—— ARAGO, *Oeuvres*, vol. 2 (1854), p. 694 [1]

INTRODUCTION: This paper, as those preceding it, is a continuation of a report of one aspect of research into one of the greatest problems of audio engineering, learning how to analyze and measure what we "hear."

Finding a solution to this problem is important because the end product of audio engineering is the subjective listening experience. If we wish to advance audio engineering, we must be able to understand how our labors bear on the end product.

This problem, which is so easy to define and discuss, is considerably more difficult to solve. The extent of this difficulty is generally underestimated by those who do not attempt to bring the human listener into the analysis.

At the present state of sound reproduction technology, the audio engineer shares the professional goal of a magician. Both strive for the creation of an illusion in the mind of the observer. In audio this illusion is that of an apparent acoustic reality. The majority of listeners "hear" almost the same illusion. An industry is based on that premise.

But if we carefully measure the sound field in the listening environment, we find that the actual sound comes, as we expect, from discrete loudspeaker sources and could not have originated from the apparent space location of the illusory sources. One can, of course, set up replica sound fields that duplicate the proper wavefronts for creating a virtual sound source, but that is not what one finds in a normal listening situation.

If we wish to understand how to "measure" what we "hear," then we must deal with subjective perception and the illusion of sound. We cannot avoid it or pretend that it does not exist.

In the research, one aspect of which is reported in this paper, a serious attempt is being made to bring both the personal experience of subjective perception *and* the description of physical processes under one common set of rules. This is an enormously complicated task because it means that absolutely nothing can be left out as a "special case." Every equation we now use must be accurately preserved within whatever "supermath" we attempt to utilize, while at the same time every subjective element such as musical training, prior experience, how the observer feels, what other stimuli are involved, and even personal prejudice must be given accounting in this same super math.

Underlying all of this work is a basic premise for which this author must assume total responsibility. That premise is that cognition, perception, and value judgment are subjective operations performed on a set of sensory stimuli taken in context with previous experiences and based on

Manuscript received January 26, 1976.

rules of form and structure. Those subjective rules, derived from the substance of physical experience, will parallel the rules of the experience itself. It is this author's thesis that a useful model for the rules of form and structure may be found in abstract geometry which thus provides a common modality for the analysis of subjective as well as objective experience and establishes a bridge between them. This is not only a geometry of the physical world, it is a geometry of perception.

The intent of this paper is to present some of the geometric concepts which are involved when we "change the frame of reference." This process is one of the most important operations of human perceptive endeavors as well as in objective engineering. It is what we do when we consider things from another persons view, when we amplify a signal, when we search for information in a signal, when we cut a record or listen to a loudspeaker reproduction.

It is all these things and more because we are considering the abstract application of geometry. The abstraction is that we associate certain geometric entities with *things* rather than numbers. To paraphrase the words of the mathematician Hilbert [2], we may talk about points, lines, and surfaces, but these words could just as easily stand for tables, chairs, beer mugs, or any object. But the additional abstraction we call for is that we not only think of objects but of actions and of conceptual associations. When we talk of images, figures, signals, or spaces in this paper we may be dealing with anything that has an organized structure composed of definable elements, be it an office building or a musical score. When we talk of mapping, or transformations, or functional relationships, or taking different viewpoints, we mean doing something to produce a new result.

We are not avoiding cognition and perception in such analysis. We are meeting it head on by dealing with the relationship among the elements composing that faculty and are using generalized terminology relating to the physical processes which underly it. Nor are we diluting the mathematical basis for analyzing physical processes because we are using those more fundamental relationships on which our simplified engineering mathematics is based. We believe the results which are here presented can be of value in audio engineering.

REVIEW OF PREVIOUS RESULTS

In order to deal with the fundamentals, let us summarize some of the results which have already been presented [3], [4].

A fundamental premise was advanced that nature conforms to rules of geometry. By *nature* we mean not only the physical world which we can measure, but also our subjective perception obtained through our senses and interpreted as either an actual or an apparent reality. In this view we align perception in accordance with our experience of the way the physical world behaves. The very words we use to describe our subjective impressions reveal the geometric structure of our perception.

By *geometry* we mean that branch of analysis dealing with form, texture, and those relationships that do not change simply because we alter our frame of reference.

The geometric premise was presented in the form of postulates that asserted that 1) nature conforms to certain geometric relationships, and 2) the effect of distortion is to change the type of geometry governing a process. Examples were given of the use of these postulates, and certain predictions of subjective effects dictated by these postulates were presented.

It was also shown that while our visual world is principally three dimensional and Euclidean, this is not true of other perceptual relationships, particularly sound. We not only need to accept other than three dimensions, but we need to develop the facility to think in terms of any number of dimensions and in a wide variety of units of expression.

Perhaps the most significant change in the way we can conceptualize the structuring of natural processes was shown to arise from an open-minded investigation of an observation that is deceptively self-evident, but seldom considered, namely, *nature does not solve equations in order to function*. People solve equations. The equations we set up and solve are nothing more than conceptual entities that model physical processes. We have no right to assume that we know "the" equations of a process or that there cannot be other models in other frames of reference that are just as valid.

In order to introduce some of these other models and frames of reference, some concepts were introduced from linear geometry. Most of us actually use some of these in our present analysis, but seldom recognize them for what they are. There is, for example, a tremendously rich chapter in abstract mathematics that is known as functional analysis. Functional analysis is an abstract formulation of the concept of an operator acting on a class of functions. In that formulation functions are regarded as elements or points of a (possibly) infinite-dimensional space, and operators transform points into points. An example of this is Shannon and Weaver's [5] pioneering work on a signal space of $2TW$ dimensions (for messages of duration T and bandwidth W) which has led to the present concept of a signal space [6] in which a signal can be considered a single entity or point of an infinite-dimensional space.

Some basic relationships were presented for dealing with one special type of linear signal space in audio terms. This was done because the signal space concept can be just as valid in audio engineering as in other fields of communication analysis. A special vector space was presented that uses generalized all-pass functions [3] which can be used to write down the time response, by inspection from a lookup table, for network frequency responses which are expressible as rational polynomial fractions in the frequency domain. This includes virtually every practical audio network. One of the values of this signal space lies in the fact that it allows us to completely bypass the Fourier transform and replace a complicated integral with a linear superposition of terms, all of which have the same form. A table was provided for this purpose.

But there is much more to it than that, for it also turns out that another way of describing the same thing is in terms of functions expressed on N-dimensional manifolds. An example of this was given as a representation of signals in terms of complex expressions in N dimensions in a form known as class C^N. The frequency response is such an expression and is a one-dimensional representation, a function only of frequency, which makes it C^1. It also turns out that the time domain equivalent of this, obtained as a Fourier transformation from the frequency domain, is also C^1. From purely geometrical considerations we know that there must be other, equally valid, expressions in two, three, and any number of dimensions we choose to use. And the generalized all-pass expansion showed us how to generate such a space, a delay space of class C^2 which, because it is two dimensional, can be expressed on a surface, called the delay plane. It was shown that the reason why a linear superposition of all-pass transfer functions could replace a Fourier integral in passing from the frequency domain to the time domain was because it actually represented a detour through the delay plane and bypassed the more direct Fourier route.

So, instead of having only two representations to play with in objective audio analysis, the frequency domain expansion and the time domain expansion, we actually have an unlimited number of representations in a great many possible domains. These all represent the same thing, but use different coordinates and dimensions to do it.

What has this mathematical abstraction got to do with audio engineering? A great deal, because when we begin looking at some of these higher dimensional manifolds it turns out that the coordinates of expression can be made to correspond to the way we *describe what we hear*. Abstraction vanishes when we realize that a frequency domain expansion can be completely accurate and yet have *no meaning to a subjective listener* because it is in the *wrong system of coordinates*. If we can transform the measurement into the proper frame of reference, we can make contemporary measurements subjectively understandable. Conversely, when we find out how to express subjective perception in the proper frame of reference, we will know what to measure.

It was also pointed out that a given numerical ranking for distortion in one frame of reference could *not* automatically be used to infer a comparable ranking for quality in another frame of reference [4]. It does not follow that one tenth percent distortion in an amplifier, as we now measure it, is always less objectionable on a subjective basis than one percent distortion, because the way in which it effects the multidimensioned subjective illusion may have little correlation with the way it modifies a special waveform in a one-dimensional measurement.

There were many other results presented in the previous papers, but this short summary of the more significant geometric concepts serves to illustrate the type of thinking that we must use if we want to apply some of these newer ideas to the problem of dealing with subjective perception in objective terms. We must think in terms of form, texture, and the relationship among things. We must accept the many analogies of form (such as time–frequency duality) that abound in engineering, not simply as happy accidents, but as expressions of an underlying structure which we can use to gain a deeper understanding of not only how some things happen the way they do, but why they should happen that way.

FRAMES OF REFERENCE

As stated earlier [4], signal processing consists of taking a signal, doing something to it, and thus producing another signal. Let us expand that concept. We need three conceptual entities: an "input," a "process," and an "output." Let us call the input f, the process m, and the output g. In standard mathematical terminology this is expressed as

$$m{:}f \rightarrow g$$

Let us now take a large conceptual step. Suppose we have an f. For the sake of illustration f may be a voltage corresponding to one channel of a two-channel stereo signal. The conceptual step is the recognition that even without our awareness, there are many g's corresponding to that f. In fact there are an infinity of g's. We happen to be observing one of the representations of that signal, which we call f. There could be another person, using a totally different frame of reference, who is also observing the same signal, except that as far as he is concerned it looks like a g.

Therefore we do not have to *do* something to a signal in order to have a different representation. We need only *change our frame of reference*. In fact, we could look on signal processing, which actually does something to a signal, as nothing more than a change of frame of reference. The "output" is, in that concept, only a different view of the "input."

This lets us take a larger view of audio engineering, and was the implied philosophy of the postulative approach to [4].

DISTORTION

What is distortion? From this viewpoint, distortion is a change in the type of geometry for the frame of reference we use. If the geometry has a special kind of metric, or measuring rod, we can say that distortion has warped the representation by producing a curvature in the space of representation. There are two kinds of curvature we may consider, analytic and singular. If the curvature is analytic, we can always "undo" the process and end up with an undistorted, or unwarped, representation. Singularities are irreversible. That is why two classes of nonlinear operator were introduced in [4], and why the definition of nonlinear operator deviated from the conventional mathematical definition.

Distortion changes a Euclidean geometry C^N to a nonEuclidean geometry. If we try to maintain a description of this distortion in the same number of dimensions N, then we find that there is the equivalent of a cross modulation of parameters when we enter into it from an otherwise undistorted frame of reference. As a subjective example, the position and tone values of a previously undistorted sound image will become somewhat cross coupled by distortion so that the position in space of a sound may be a function of its instantaneous values. This warping in representation is what we called *representation distortion* [4].

In the case where there is a metric, and the distorted frame of reference is of the special type known as Riemannian, then we know that the description can again be converted to a Euclidean form of higher dimensionality [7] with a number of dimensions no smaller than $\frac{1}{2}N(N+1)$. Subjective listeners may subconsciously accommodate to small amounts of distortion by considering the coordinate interactions as corresponding to new functional entities, thus enlarging the number of perceptual dimensions.

In a symbolic sense, m (which will stand for mapping operator) could be considered a window through which f views g. When we are at f and look through the window at g we may see a highly warped scene, from our standpoint. However, an observer in g stands in an orderly world and it is f, which he views through the window, that is untidy.

Because the coordinate bases in f and g may be totally different, even of different dimensionality, we cannot automatically make judgements about the extent of distortion *until we move ourselves over to the proper coordinate framework*. One way to do this is to pass a measuring rod through the window m. We give a person in that frame of reference a rod that represents a measurement in a preferred fashion, such as an increment in one of our parameters. Then we observe how he interprets that measurement. If the window is distorted, we will watch in horror as he bends our rod out of shape. When we complain that this is no way to treat a precision instrument, he angrily retorts that he did nothing but accept a badly shaped instrument from us, and we should know better than to use warped tools. The nonlinear window has caused a representation distortion.

This anthropomorphic discussion may convey, without recourse to mathematic symbolism, the spirit of one way we can possibly measure subjective distortion (what g observes in his frame of reference) by objective measurements (made by f in f's frame of reference). The measurer f makes an incremental analysis of a signal by changing only one parameter at a time. The change which this produces *in g's frame of reference* is determined. In this way we begin to build up a picture of just what it is that g hears when we observe a nonperfect f.

PRINCIPLE OF ALTERNATIVES

Let us go back to the concept of mapping and of implied alternate representations of a signal. We make the following assertion: "There is no privileged frame of reference, either for dimensionality or units of measurement."

We may find certain frames of reference extremely convenient and may even find it impossible to imagine that any other systems exist, but they do. Let us advance that assertion to a principle, the principle of alternatives: "There is an alternate frame of reference for every defined situation."

This principle may sound pretentious, particularly to those accustomed to concepts of geometrodynamics and quantum physics, but it actually derives from the geometrical structures which we are considering. Let us now use this principle to discuss mapping between alternate representations.

MAPPING

Consider the situation symbolized in Figs. 1–4 which illustrate the conceptual steps involved in mapping from one frame of reference to another. A signal f in N-dimensional form is to be mapped by mapping operator m into a signal g in M-dimensional form.

We have the N-dimensional space of f. The principle of alternatives allows us to speculate the existence of another M-dimensional space (Fig. 2). Conceptually, bring the two spaces together to form a composite space of either $(M+N)$ dimensions (Fig. 3**a**) if none of the coordinates of g are like those of f, or of $(M+N-R)$ dimensions if R of the coordinates of g are like those of f (Fig. 3**b**).

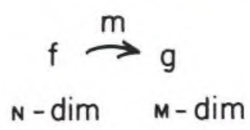


Fig. 1. Two persons viewing exactly the same thing will see it differently if they do not have the same frame of reference. If we see the situation as some f in an N-dimensional frame of reference, then how might it appear to a person who uses a different M-dimensional frame of reference? What map m allows us to take the other person's view?

Fig. 2. The principle of alternatives allows us to conceptualize that if we have a valid frame of reference, then there will also exist a valid alternate view of the same thing, and a map can be found to transfer from one view to another. Think of these alternate frames of reference as individual spaces, shown here as amorphous blobs.

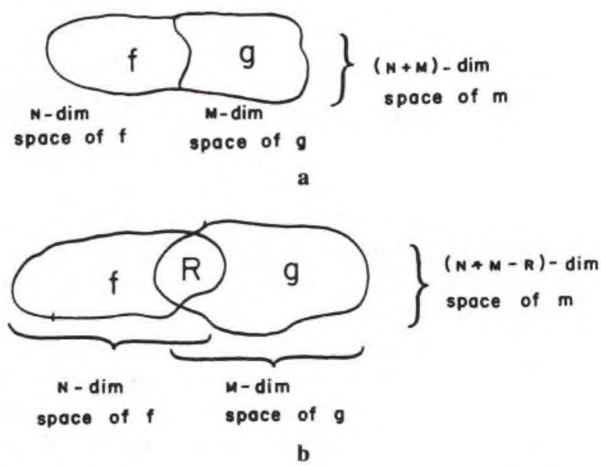


Fig. 3. Even though f and g are separate representations of the same entity, the map joining them is a geometric figure that must exist in the union of both spaces. **a.** None of the coordinates of f and g are alike. **b.** A number of coordinates, R of them, are common to f and g.

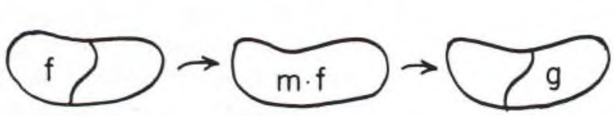


Fig. 4. The process of mapping consists of spreading f over the union of spaces forming the product of m and f to do so, then removing the original N-space part by summing the product over all N-space coordinates. This leaves g in the proper M space.

What we have done is to create a new space that contains the representations of f and g as partitions. Now define a new signal m which exists (along with its necessary derivatives) throughout the totality of the new space. The steps we will use to come up with a g in its M-dimensional space are as follows:

1) "Spread" f throughout the entirety of the $(N + M - R)$ space by the use of m.

2) "Filter out" that part of this spread signal that exists in the original N-dimensional partition of the $(N + M - R)$ space.

3) What is left will be g in its M space.

The engineering analog of this mapping process occurs when we want to heterodyne a signal from one spectrum range to another. We cross modulate the original signal with another signal that will cause the result to be spread over the desired spectrum range, then we filter out those cross-modulation products that are not wanted.

Getting back to the mapping, assume the coordinates of f are the following:

$$x = x_1, x_2, \cdots, x_N$$

and the corresponding coordinates of g are

$$\xi = \xi_1, \xi_2, \cdots, \xi_M.$$

m will be a function, defined throughout $(N + M)$, which appears in the N space as a hypersurface defined in terms of the parameters ξ on a coordinate basis of x. In the M space this will appear as a hypersurface defined on the ξ coordinates in terms of parameters x.

In other words, each frame of reference sees m as a *geometric figure* defined in terms of certain parameters which correspond to the coordinates of the *other frame of reference*. That is how the parameters of a new space will always manifest themselves, as certain properties which we view within our preset frame of reference.

To "spread" f we multiply the figure f by the m hypersurface as follows:

$$f(x) \cdot m(x,\xi) = f(x_1, x_2, \cdots, x_N) \cdot m(x_1, x_2, \cdots x_N, \xi_1, \xi_2, \cdots, \xi_M).$$

We then allow the parameters ξ to pass through *all their values* and sum the resultant for all values of the coordinates x. This will produce g,

$$g(\xi) = \int_x f(x) \cdot m(x,\xi) dx = \Big(f(x), m(x,\xi)\Big). \quad (1)^1$$

g takes the form of an inner product [3]. The inner product is one of the most significant geometric relations we encounter in engineering, and the reason lies in its mapping properties. The sum of products is the fundamental form in geometric conversion from one frame of reference to another.

The process of letting the ξ pass through their values allows the hypersurface that is m (as seen in f's frame of reference) to sweep through the space of f. It should be evident that the dimensionality (dim) of the mapping operator m is equal to the sum of the *unlike* dimensions of the signals it joins. If there are R shared coordinates between f and g, then

$$\dim m = \dim f + \dim g - R.$$

If, for example, we want to map a three-dimensional object into a two-dimensional representation in which the coordinates are not shared, such as a conventional optical hologram, then we need a five-dimensional mapping operator.

FOURIER TRANSFORMATION

This concept of mapping gives us a different picture of some of the familiar operations of engineering. For example, consider the fundamental process we know as the Fourier transformation. It is difficult to imagine modern audio engineering without using concepts drawn from this operation.

In geometric terms, Fourier transformation is a map between an N-dimensional representation in a set of coordinates x and another N-dimensional representation in a *special set of coordinates* ξ. A Fourier transform is an *alternate view* of the same signal in the *same number of dimensions* but in a *different frame of reference*. What is special is that the mapping operator is a phasoid that uses a hyperplane as its angle dependence.

The equation of a hyperplane p in the coordinates of x is given by the inner product of the coordinates x and the parameters ξ,

$$(x,\xi) \triangleq \xi_1 x_2 + \xi_2 x_2 + \cdots + \xi_N x_N = p.$$

The parameters ξ become the coordinates of the transformed expression. Because it is defined on a hyperplane which we allow to sweep through the space of x's, the units of measurement of ξ *must be the inverse* of the units of measurement of x. It can be no other way and still be a Fourier transform. What was time becomes inverse time, or frequency; what was space becomes inverse space, or space frequency, etc.

The mapping becomes, in shortened notation, *for any number of dimensions*

$$g(\xi) = \int_x f(x)\, e^{i(x,\xi)}\, dx = (f(x), e^{i(x,\xi)}). \quad (2)$$

On a technical point that will become important in subsequent analysis, any reversible mapping operator that uses a hyperplane as its basis, and which is reciprocal in the sense that the mapping operator takes the same form in the two dual spaces, is called a Fourier mapping operator [8]. The phasoid is such a Fourier operator and the mapping thus formed is now called, simply, Fourier transformation. There are, however, other Fourier-related maps that con-

[1] In this present simplified discussion, we are not considering the case where m is also dependent upon f or where the integral is taken with respect to the measure of m. Nor do we consider the problem of the stability of m which may not allow us to return to precisely the same f in our initial frame of reference once we pass through other frames of reference. Both of these situations arise in perceptual processes and are important in subjective audio, but their discussion lies outside the intent of this paper, which is to introduce the basic concepts.

nect inverse measured dual spaces, such as the Hankel transform and the Mellin transform.

The special engineering advantage of the phasoid hyperplane map (Fourier transformation) lies in its unique periodicity properties and the fact that its form does not alter under the operations of differentiation and integration—so important in the solution of problems involving energy exchange.

The process of looking at a signal from these special Fourier frames of reference does not in any way imply that either 1) these are the only two views of a signal, or 2) the map we call the Fourier transformation is the *only* way to pass between Fourier frames of reference.

If, instead of a hyperplane, we have a more general *hypersurface* $\phi(x,\xi)$, the phasoid map that is formed becomes

$$g(\xi) = \int_x f(x)\, e^{i\,\phi(x,\xi)}\, dx = (f(x),\, e^{i\,\phi(x,\xi)}). \qquad (3)$$

Now the units of measurement of the ξ and x *do not have to be reciprocal.* Nor do the number of dimensions of g have to equal the number of dimensions of f. This more general map is what this author has applied to audio measurements under the name *time-delay spectrometry* [9], [10]. In audio measurements, and in particular in loudspeaker measurements under reverberant conditions, this author has published explanations of its use to map a signal from a time domain to a delay plane, thereby spreading components in a form allowing ready isolation of the anechoic from the reverberant components, and then map to a frequency domain for conventional display. Its more extensive geometric basis can be seen from the standpoint of the hypersurface map. It should now be evident why there was a special distinction given to the phasoid as a defined entity in [3].

There is, in fact, the further distinction that a cisoid is the form the phasoid takes when the argument of the phasoid is a hyperplane. We will define an N-dimensional cisoid as the complex exponential that has a hyperplane argument,

$$e^{i(x,\xi)}.$$

What this discussion is intended to show is that a Fourier transformation is a *special case* of a more general map. Fourier transformation is a shuttle that passes between representations of the same dimensionality but of reciprocal units of measurement. If we wish to travel to different levels of dimensionality and other units of measurement (relate what we measure to what we hear, for example), we cannot use the Fourier transformation but must follow a different map, such as the hypersurface maps. It may be time to get off the shuttle if we want a change of scenery.

And, because it is so important to audio engineering, let me repeat a point made in an earlier paper on this subject [3]. If by the pitch of a musical signal we mean that attribute which can change with relative time to form a glissando, then pitch is not frequency. Frequency is the coordinate x, say, and time will then be the coordinate ξ under hyperplane (Fourier) mapping. We *cannot* form a view in a reference system involving *both* time and frequency. Our view must be taken in a reference system that has the proper attributes, and we can do so with a phasoid hypersurface map onto a reference frame of mixed measurement, such as the delay plane in which the coordinates are relative time and periodicity (or pitch) measured in the units of inverse relative time. When we do that, a glissando becomes a perfectly acceptable entity involving the change of periodicity with relative time. However, if the tone stays at one periodicity *forever,* then this is a map to one dimension where relative time no longer has significance, and then, in this one-dimensional frame of reference, we have a tone of fixed *frequency*. That is the only correct way of equating the pitch of a tone with its frequency on a one-to-one basis.

MAPPING TO A SEQUENCE SPACE

In order to know what the components of the subjective image are, in terms of the two one-dimensional functional representations which we call the left- and right-channel stereo signals, we must recognize that each dimension of the subjective illusion may in fact correspond to a complicated functional relationship in the alternate one-dimensional spaces. Furthermore, these one-dimensional spaces do not stand alone, but are linked by the acoustic properties of the listening environment.

It often turns out that the coordinate-mapping process from the one-dimensional voltages in an audio chain to the multidimensional subjective illusion is more readily handled in terms of an intermediate map onto a space of sequences. That is exactly what is represented by the Fourier series expansion of a function. When we solve for the Fourier series corresponding to a periodic function, we are making use of a one-to-one map of the function space onto a space of sequences. This is not the way most engineers look upon Fourier series expansions, but it is a geometric interpretation that comes directly out of the celebrated Riesz–Fisher theorem [11].

The purpose of this exposition is to point out that one coordinate of the subjective illusion may correspond to paired sine-wave voltages in the audio reproducing chain, while another coordinate may correspond to other paired sequences. Thus a pure tonal of a definite pitch in the subjective image may be the equivalent of a pure sine-wave signal in the audio voltage. An increment in intensity is a change in the energy of the sine waves, while an increment in spatial position may be a sequence shift of relative intensity and time of occurrence.

In order to test an audio component we generally need some sort of measuring rod. One way to generate such a measuring rod is to map a continuous functional representation into an appropriate set of sequences and analyze how these sequence values are altered by changes in the dynamics of the continuous representation. If we can identify certain combinations of sequences with subjective image dimensions, then the change of sequence sets due to the presence of program distortion may yield clues concerning the changes to be expected in the subjective illusion.

In its simplified form, this is what we do when we measure the harmonic and intermodulation distortion of an audio component. We are using a signal f with a known periodicity and mapping it to a sequence space g by using a mapping function with a periodicity related to f. Our sine-wave signals, filters, and demodulators are the engineering manifestations of the elements of Eq. (1).

DEFINITIONS

We have now arrived at the place where some specialized definitions are needed.

Alternate space. The space of representation for an alternate view.

Dual space. Those special alternate spaces which have the same dimensionality and are dually linked by the property that the vectors of each space correspond to the "coefficients of expansion" of the vectors of the other space. In functional analysis, the dual space is identified as the space of bounded linear functionals.

Duality. The dual state or quality arising out of alternate representations in dual spaces that gives rise to a property in one space having analogous form with another property in the dual space. Duality is the rule, rather than the exception, in physical problems.

Alternatives. Those equally valid viewpoints expressed in different frames of reference and corresponding to alternate spaces.

Mapping space. The composite space formed from alternate spaces and throughout which the mapping operator m must be defined in order to map a figure in one alternate space into a figure in the other alternate space. This is also called the *process space* if we perform some process to change one figure into another. The mapping space is what we must consider when we wish to examine two alternate views of a given situation. The mapping space is the *union* of the alternate spaces.

Illuminating function. The special type of space-filling mapping operator formed throughout all or part of the mapping space by means of *wave propagation* or its analogy from an equivalent source of illumination, also called *illuminating operator*. The values of sound at every place in a room that is caused by a source of sound in that room. This is obviously a special case of a more general *mapping function,* since wave propagation is a special type of relationship between parts of a space-filling function.

Hologram. Alternate representations mapped from each other by an illuminating function. This unusual definition preserves the more common expression of a hologram as that diffraction pattern of an object formed by combined interaction of scattered wavefronts from the object and the source of illumination. But in addition it opens up the hologram to a more general interpretation in any type of space, in any number of dimensions, and with any type of illumination.

Holomorph. Alternate representations connected by a mapping operator because they are both *whole forms* of the same thing. Obviously, a hologram (whole writing) is a special case of a holomorph. The adjective *holomorphic* is used in complex variable theory to denote the condition where every point in a region has a uniquely determined derivative. This adjective use, and the fact that an N-dimensional holomorph in Euclidean space is C^N is quite consistent, even though the word *holomorph* as a noun is apparently not found in the mathematical literature. All valid alternate representations, such as those joined by Fourier transformation, are holomorphs. Holomorphs formed by illuminating functions are holograms.

Global–local map. When "everywhere" in one figure is mapped to a localized region in an alternate figure.

Point-to-point. When each single point on one figure is mapped to a corresponding single point on an alternate figure. If local regions are mapped to corresponding local regions, we shall call it *local–local*. The other extreme of this is where everywhere is mapped into everywhere, a condition we shall call *global–global*. Phonograph records, magnetic tape, and television scanning are point-to-point maps.

Intersection of spaces. The shared region when spaces of representation overlap in shared coordinates. The region of dimensionality R is an intersection. Feedback and any form of shared interaction can only occur in the intersection.

Coincide. When the dimensionality of both of two spaces is equal to the dimensionality of their intersection, the spaces are said to coincide.

Include, imbed. When the dimensionality of two spaces is different, and the dimensionality of the intersection is equal to that of one of the spaces, the other space will be assumed to be of larger dimensionality and the smaller space will be said to be included, or imbedded, in the larger space. The larger space will be said to include the smaller space. The process of increasing the number of dimensions of a representation to end up with a representation that has the original dimensions plus some new ones, will be called *expanding*. The inverse of this operation will be called *contraction*.

The definitions may grate on the sensibilities of those engineers who pride themselves on their practical approach to audio engineering. But what we are doing is describing the relationship between things—and that is the true power of geometry. Form, shape, texture, and how something appears when seen from a different point of view is what we are discussing here.

When an engineer designs a gadget to do something, he has created a relationship between the input and output, whatever they may be. Engineers develop common sense by observing how things are related. We can all be at a complete loss, however, when we find difficulty seeing the relationship between things that we know must be related, but which, for one reason or another, we cannot observe, and hence consider "common sense." It is then that such definitions are useful. One of the most belittling experiences is to deride the "black art" of a craftsman who gets consistent results by a certain ritual which he cannot explain and then to discover that his actions in fact held a deeper technical significance than we understood at that time from our simplified mathematical model.

REFERENCES

[1] R. E. Moritz, *On Mathematics and Mathematicians* (Dover, New York, 1958), p. 694.

[2] M. Kline, *Mathematical Thought from Ancient to Modern Times* (Oxford University Press, New York, 1972), p. 1010.

[3] R. C. Heyser, "The Delay Plane, Objective Analysis of Subjective Properties, Parts I and II," *J. Audio Eng. Soc.*, vol 21, pp. 690-701 (Nov. 1973); pp.786-791 (Dec. 1973).

[4] R. C. Heyser, "Geometrical Considerations of Subjective Audio," *J. Audio Eng. Soc.*, vol 22, pp. 674-682 (Nov. 1974).

[5] C. Shannon and W. Weaver, *The Mathematical Theory of Communication* (University of Illinois Press, Urbana, 1949).

[6] L. E. Franks, *Signal Theory* (Prentice-Hall, Englewood Cliffs, NJ, 1969).

[7] C. Weatherburn, *Riemannian Geometry and the Tensor Calculus* (Cambridge University Press, London, 1966).

[8] R. Bracewell, *The Fourier Transform and its Applications* (McGraw-Hill, New York, 1965).

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

[10] R. C. Heyser, "Loudspeaker Phase Characteristics and Time Delay Distortion, Parts I and II," *J. Audio Eng. Soc.*, vol 17, p. 30 (1969); p. 130 (1969).

[11] E. C. Titchmarsh, *The Theory of Functions* (Oxford University Press, London, 1968).

About 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. 24, no. 9, pp. 742–751, 1976 November.

COMMUNICATIONS

PERSPECTIVES IN AUDIO ANALYSIS: CHANGING THE FRAME OF REFERENCE, Part II

RICHARD C. HEYSER

Jet Propulsion Laboratory, Pasadena, CA

SCANNING AND IMAGING—A POINT OF VIEW

A phonograph record, if we could stretch the spiral groove out to a straight line, is a point-to-point map between a time-dependent signal and a mechanical deformation of a place on the groove. Is this a good map? The answer to that depends upon what our economic requirements may be. From the standpoint of the convenience of the inverse map from mechanical to electrical reference (which we call scanning), it is a very good map. But from the standpoint of utilization of the record base, it is terrible. Actually, the entire program is on that long thread which cutting engineers find bothersome enough that it must be vacuumed up and thrown away.

Also it is a poor map from another geometrical standpoint. It is a property of point-to-point maps that each single error is forever passed on as an error with no chance for "healing" or reducing its prominence. A scratched record will always destroy program values at the damaged places. Is there a better map to eliminate these defects? The answer is a resounding affirmative.

In order to provide an illustration, let us consider an analogous situation that can be readily visualized. Assume that we want to take a two-dimensional scene, such as a monochrome photograph, and convert it to a one-dimensional signal that may be equated to a time sequence.

Editor's Note: Part I of Perspectives in Audio Analysis appeared in the October issue.

This is obviously the problem of television scanning, and the corresponding geometry is shown symbolically in Figs. 5 and 6. The scene is shown as f, the resultant one-dimensional representation is shown as g, and the mapping function by m.

If the coordinates of f and g are independent, then m must be a three-dimensional operator, and this is shown in Figs. 5 and 6 as a rectangle. We can observe f and g, but m is the process and not observable in our normal frame of reference. However, artistic license is taken in Figs. 5 and 6 to symbolize what is going on.

Mapping consists of multiplying f times m and summing the product as all values of ξ are swept over. In television we can take each "hypersurface in ξ" of m and place it over f as a transparency and pass light through the sandwich thus formed in order to get a product scene $m \cdot f$. We sum the product scene by collecting the total transilluminated light and passing it to a photocell to produce a single numerical value.

Consider that m is a deck of cards with a single card for each value of ξ. Conventional raster scan is the case where each card is totally opaque except for a pinhole, and the pinhole occupies adjacent coordinates of x for successive values of ξ. This is symbolized in Fig. 7**a**, which represents how each successive card in the deck would appear if fanned out onto a table. This is the point-to-point analog of scanning a phonograph record, and it is an instructive view of television raster scan.

The problem with this type of scan is that missing cards in the deck, particularly connected sequences of cards corresponding to a run, lose substantial information. In order to combat the "lost image points" due to bad sequences, such as an interrupted transmission, we can shuffle the deck so that no two adjacent image points are represented by adjacent cards, such as symbolized in Fig. 7**b**. This scanning method works very well and is known as pseudorandom scan. Single blemishes still disrupt the corresponding image point on g, but long interrupted sequences of ξ are drawn from such a distributed sequence of image points that the basic form of the image can still be constructed.

A more efficient map is a global-local map where the entire scene of f influences each point of g. In order to do this we can replace each card of m by another card that contains its own unique random scene, such as shown in Fig. 8. The type of scene to use depends upon the desired geometric results in g. If we have reason to believe that each value of g may be corrupted by a random disturbance during transmission and we wish to reconstruct a new scene f' from the received g, we choose the statistics of the m hypersurfaces (each card of the deck) to match those of the interference.

This process of image transmission is far more efficient from the standpoint of immunity to random influences than any point-to-point scan. Each card of the deck can be considered a vector of an infinite-dimensional vector space, if we want to think of it in those terms. This shows the conceptual tie between infinite-dimensional vector spaces and finite-dimensional manifolds.

Because it is global–local, a piece of g can be used to reconstruct *all* of f'. The penalty paid is that f' will be either noisy, or out of focus, or whatever defect the designer elects to design into the process of m as a tradeoff for lost information.

This then is the result of the geometric analysis. To map a two-dimensional image onto a one-dimensional sequence in a more efficient manner than now done, we use another series of mapping images. The mapping images should have distributions of values that are mutually orthogonal (zero inner product for unlike terms) in each of the two-

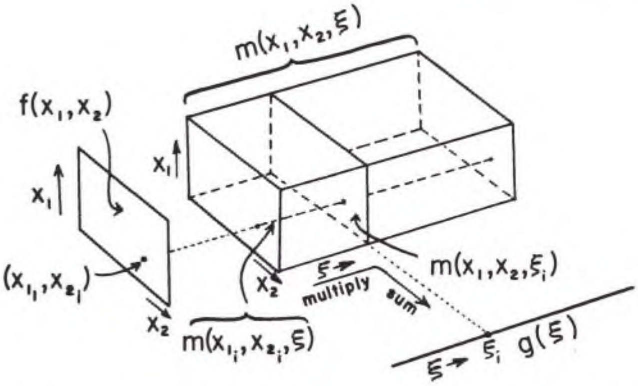


Fig. 5. Symbolic representation of a generalized mapping process between a two-dimensional signal f and a one-dimensional signal g. Each coordinate point ξ_i in the one-dimensional signal is represented in the space of f by a two-dimensional mapping image $m(x_1,x_2,\xi_i)$, shown here as a two-dimensional slice of the required three-dimensional mapping space. Conversely, each coordinate point in the two-dimensional signal (x_{1_i}, x_{2_i}) is represented in the space of g by a one-dimensional distribution $m(x_{1_i}, x_{2_i},\xi)$, shown here as a line in the mapping space. To map f into g we have two alternatives. We may either multiply all of f times each two-dimensional mapping image and sum the product scene over all space values of f to give each signal value at each point of g. Or we may multiply each scene value of f times the corresponding distribution for that scene coordinate point and sum all such products to give the one-dimensional g.

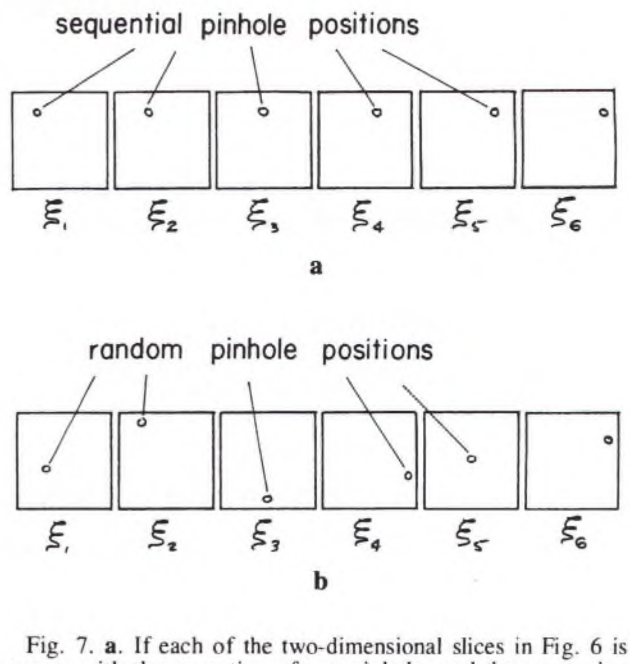


Fig. 7. **a**. If each of the two-dimensional slices in Fig. 6 is opaque with the exception of one pinhole, and the successive slices have pinholes occupying successive positions, then this is a point-to-point continuous map which corresponds to standard television scanning. If we could place slices next to each other as playing on a table, they would appear in this fashion. **b**. If the sequence of slices in Fig. 7**a** is redistributed so that adjacent slices no longer correspond to adjacent pinhole positions, then the scanning is still point to point, but now it has a pseudorandom property.

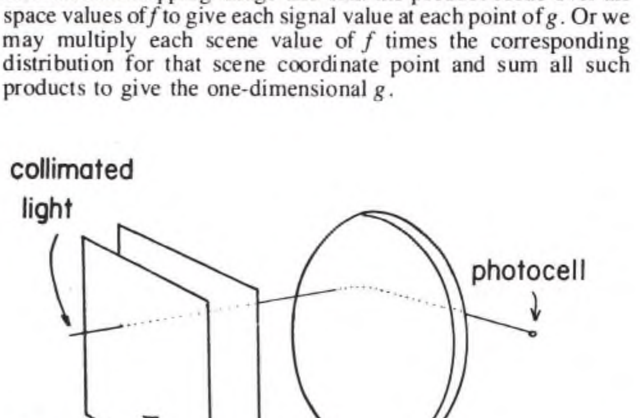


Fig. 6. Example of a television scanning process using the geometrical basis of Fig. 5. In this case the scene is multiplied by two-dimensional mapping slices and summed to give the voltage corresponding to a single value of the video waveform. For illustration it is assumed that the scene and each mapping is an optical transparency and summation is performed by optical means.

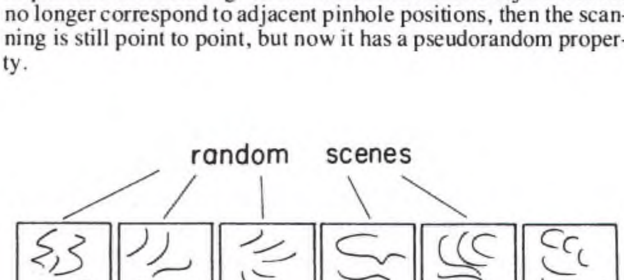


Fig. 8. A more general map uses slices which represent complete and separate images. This forms a global–local map in which each coordinate point of the video signal contains information about the entire scene being scanned. If each mapping image is orthogonal to all the rest and has the same autocorrelation, then the same scenes may be used for a reversible map from the video waveform to the original scene. If the autocorrelation varies from one ξ value to another, then this map is not reversible in one step but must pass through an intermediate space in order to reconstruct the original scene.

dimensions of the image frame of reference as well as orthogonal along the parameter that becomes the one-dimensional frame of reference.

When we do that, we have a global–local map that provides the greatest immunity against the effects of errors in handling the one-dimensional sequence, including reduced channel capacity. When we again reconstruct the two-dimensional image by the inverse global–local process, any error is spread everywhere in such a fashion that it is optimally suppressed as a differential increment on an otherwise perfect reconstruction.

This more optimum mapping is true for any process we wish to improve, and orthogonal scanning figures should be used rather than less optimum scanning maps. If we want to map an audio program into another form, we need a sequence of orthogonal audio programs, if we want the least error sensitivity.

SELF-REPAIRING MAPS AND FORM MAPS

If nothing else, this different geometric viewpoint of engineering teaches us that form and texture can be very important attributes. We can take geometric shapes within a given coordinate basis and map to a space where the shapes represent dimensions. This is analogous to an operation in linear spaces wherein a space can be mapped to a conjugate space where vectors in the original space become functionals in the conjugate space except that we can change this in turn to a space where certain sets of functionals become determining dimensions and other functionals are expressed in those dimensions.

All of this is not as weird as it may appear at first glance, because many of these things begin to sound very much like the way many persons describe their perception of sight and sound. Consider this; a musically trained person can "hear" a musical chord as a thing that is almost a discrete entity. He does not perform a spectrum analysis on each tone and its partials, but it is so strongly identified in his mind as a coordinate that he can mentally "hear" the chord and use it as a basis of comparison even if all he perceives is visual marks representing notes in a score. For others, articulated sequences of chords may be just as important. All in all, then, there are alternate spaces in which form is a determining dimensional attribute.

When one accepts this, it is not unreasonable to inquire what are the fewest number of attributes to "remember" a scene or communicate the essence of an experience, such as a caricature sketch that can identify one experience out of a great many. These considerations lie within the purview of engineering when we invoke geometry.

On a technical level it would appear from certain geometric mapping conversions that an incredibly smaller number of parameters may be needed than one might assemble from a more conventional coding scheme in order to characterize a complicated situation. Furthermore, substantial dismemberment or disfiguration of portions of an alternate representation may not have a significant effect on the reconstructed figure until a threshold in deformation is achieved, at which occurrence everything goes bad. For example, we might be able to identify a melody formed out of grossly modified sounds which, heard singly, represent something else to us.

The reason for this is that if overall form contains the information, damage can be "repaired" by choosing forms such that equilibrium is achieved by applying some type of "stress" that tends to draw out local errors to be absorbed into the proper figure. This can continue until the local errors become so substantial that the stress yields a new form that corresponds to another possible figure. An analog of this occurs in present linear mathematics in mapping and is called the "almost everywhere" (ae) limitation, an apellation used to indicate that there are sets of measure zero that "get lost" when we attempt to recover them by an inverse map.

This "repairing" can only take place if there are a predetermined number of acceptable forms and a forced choice is made as to which form the deformed signal most closely represents. In communication theory we normally use each possible signal as a coordinate of an extremely large space (for example, $2TW$ dimensions where T is the length of message of bandwidth W) and use as a measure of closeness the power criterion of mean square error. We can thus "repair" by performing a forced choice among a finite set of allowable forms and apply conditions on power spectral density as the "stress".

The use of form for conveying information in a modest number of repairable alternatives is quite interesting. While it verges on careless speculation, the communication problem of genetic transfer or of the body's immune system may yield to such views wherein the geometric form is the key. In an example closer to audio, we may begin to inquire what is really necessary to establish a melodic contour, or even the pitch of a tonal structure. And, as if it were not already obvious, the spaces of psychoacoustic "dimensions" is certainly part of this concept, as is a properly structured subjective frame of reference that uses words to describe evoked images of acoustic form. By properly structured, I mean the use of words that have a previously defined meaning.

METRIC OR NONMETRIC GEOMETRY

We have some fundamental problems when contemplating the type of geometry to use for audio engineering. One of the more serious questions concerns whether we need a metric.

Our fundamental assertion is that a signal in one form and dimensional representation can be changed into another representation in a different number of dimensions. If it is the same signal (only viewed in a different frame of reference), then what remains invariant? What is the "essence" of each signal that distinguishes it from all the rest no matter what coordinate basis we use?

If the different possible coordinates all have the same number of dimensions and are measured in the same units (such as distance in meters), then the answer is simple— the line element ds^2 generates an invariant algebraic form of the second order. We have a measurable quantity from which we can generate a metric geometry, and as a result have all of the riches of tensor calculus available for use.

But this is apparently no longer the case when we take the same signal and expand it to alternate representations which do not have coordinates measured by the same units. How can we measure the increment between a value of tone and space, or between "warmth" and "spaciousness," for that matter?

There is one second-order algebraic expression that may eventually suffice for alternate dimensional use. This is the expression for energy and its density partitioning in terms of system coordinates. Total energy density and its kinetic and potential density partitioning in all coordinates has a Pythagorean triangle relationship where the partitions in each dimension are related by Hilbert transformation [12].

Because engineering deals with systems of finite total energy, it is geometrically significant to invoke this Pythagorean relationship. This relationship was derived for signals that are of class L^2 $(-\infty, \infty)$, which means the Lebesgue integral of the square of the modulus over the entire range of independent variables is finite. In engineering terms the Pythagorean relationship holds for finite-energy systems.

It is a property of Lebesgue integration that when such an integral exists in one dimensionality, it exists and has the same value in other dimensionalities [13]. Thus we can infer that the relationship known as the Parseval formula [14], which states that the integral of the square of the modulus of a function is equal to the integral of the square of the modulus of its Fourier transform, can be extended beyond Fourier transformation to include the alternate expressions in any level of dimensionality.

This geometrical fact, although stated here for the first time, coincides with our common-sense observation that we should not be able to change the total energy of an event simply by altering our frame of reference. Hence energy is an invariant and may lead to a metric. What we do when we change frames of reference is alter the detail partitioning of the energy density. That, in fact, is why this author chose to define the terms "kinetic" and "potential" in terms of the frame of reference (see [12 p. 903]).

What is tantalizing about striving for an overall metric geometry is that when we consider the restricted versions of this analysis that are now used by engineers, we find that they all have a metric. Furthermore, if we proceed as though some metric were present, even if we cannot yet identify what its significance may be, we get what appear to be proper answers for the effect of distortion. For example, a distortionless reproduction of natural sound will preserve the space position of a musical instrument and not allow it to be a function of relative time or intensity of the tonal structure of what is being played. But a distorted reproduction will warp the representation such that position in space may be dependent upon other program dynamics.

For the present time, then, the analysis will proceed with an implied metric basis for the higher-dimensional representation of those signals which are mapped upward from a reduced-dimensional metric basis. The only condition imposed on this will be physical experience. If, at any time, our model indicates a result contrary to physical evidence, then the model will be considered improper. The model has worked quite well so far.

Why the concern for a metric? Because we want to measure what we hear. If there is some perceived property that listeners consciously or subconsciously detect, then we want to know how to measure and rank on that property. This is an enormous problem and one that warrants the best weapons in our arsenal.

EXPANSION AND CONTRACTION OF DIMENSIONS

Just as Alice in the Looking Glass, we may "go into" a single dimension and expand it as a two-, three-, or more-dimensional space. This should not be surprising, because we know that there are enough points on a line to generate a plane, or cube, or hypercube with nothing left over or lost [15].

A person who is told that one of his dimensions is actually a three-dimensional space in another frame of reference should not be disturbed, because he is actually seeing everything that transpires, except that the subtle relations between the higher dimensional attributes appear coded on his one-dimensional line. If he chooses, he can map to the proper frame of reference to observe those interactions. It is just an alternate view.

SHARP EDGES AND APODIZATION

No matter what alternate view we take of a complete geometric figure, boundaries and edges will be sharply defined because we have all the information available to us. If we take alternate views of a piece of the figure, and do not have knowledge of the whole figure to which this piece belongs, then our uncertainty in knowledge takes a variety of geometric properties as we form alternate representations.

There may be one or more special coordinate representations which present the piece as a figure with sharply defined boundaries. This includes the original space in which the piece was created as a part that was separated from the whole (the space of finite sampling of continuous signals, for example), and those other spaces which are joined to it by point-to-point continuous maps. Spaces obtained through other maps, such as global–local maps, will cause the representation of the piece to be ill defined for its boundaries—"fuzzy." This is because each point of the original space is spread into either a neighborhood of the new space or the entire space itself.

In addition to fuzzy, out of focus, edges, the new representation of the original piece of a more complete figure may have footlike appendages for those regions of the new space where the normally diffuse values combine coherently. The process of modifying the spatial distribution of values in one space for the purpose of minimizing the spread of appendages in another space is called *apodization*—literally, the elimination of feet.

The use of the word apodization for this process is found in optics, radio astronomy, and surface acoustic wave technology. The term "weighting" is found elsewhere to denote this process for those situations involving the global–local Fourier transform map. The more general process of apodization, for any geometric representation, involves modification of values in phase as well as magnitude [12]. The modification of a mapping operator by the addition of a weight kernel may be done for a variety of reasons, only one of which is to bound the extent of space values in the new representation. For this reason this author prefers the specific term apodization to denote the process done to modify spatial appendages.

In audio we find ongoing examples of apodization applied to the mapping between Fourier transform domains. We find that the Fourier transform of a rectangular pedestal is a sinc function. Thus a sharp-boundaried gap in a magnetic reproduce head, which has a pedestal wavelength response, will give sin *f*/f sidelobes in the corresponding frequency response. Because it is a geometric mapping result, the frequency appendages are nonminimum phase. If we attempt a standard frequency equalization of the drop in high-frequency response as the top end rolls off toward the first appendage notch, we can equalize the amplitude of the frequency response of our magnetic reproducer, but the phase response becomes nonuniform. We need to add a nonminimum-phase all-pass network to correct the situation. Geometrically a better solution is to apodize the reproduce gap to create a better frequency distribution. This can be done by shaping the distribution of flux lines within the immediate vicinity of the gap.

In fact it is a general result that if we want to minimize fuzzy edges and appendages in one domain due to an incomplete set of data in another domain, we should apodize the data to avoid sharp edges. Maximally flat frequency responses are to be avoided if, in achieving them, we must cause a sharp "edge" in response at either the high-frequency or low-frequency end. Most of us are aware that a sharp high-frequency cutoff causes a "Gibbs phenomenon" time smear in the reproduced signal. There is a comparable time smear that happens when there is a sharply defined high-pass frequency response [9]. The geometric reason for this is that in our attempt to equalize the energy in the signal so that all frequency components have equal energy right up to a precipitous break in response, we have caused a corresponding "smear" in the time delay of those pitch components near the frequency "edge."

UNCERTAINTY AND SPATIAL FILTERING

When two spaces are mapped from each other under global–local rules, a point in one space is spread into the entire region of the other space. As more and more points in one space are combined to define a figure, then the diffuse values in the other space gradually coalesce into the equivalent alternate figure. Finally, when all points of one space are accounted for, then the alternate figure is complete and can have sharp edges.

We cannot have a sharp-edge broken-off piece in one space that corresponds to a similar sharp-edge piece that is broken off the corresponding figure when spaces are joined by other than continuous point-to-point maps. This means, for example, that there can never be a frequency response of any audio component that has zero magnitude for any band of frequencies. If we try to filter out all values of a band of frequencies, we are attempting to break a piece off a complete figure and must pay the penalty that all corresponding global–local equivalent spaces have the figure spread throughout the space. The time domain, obtained as a Fourier transformation from the frequency domain, is such a space, and the result would be that the equivalent signal values would be spread throughout the coordinates of the time domain—all possible negative time and all possible positive time. The response would thus be able to predict the future and is disallowed.

We can have nulls at single points of frequency since this is the result of cancellation and not a broken-off piece. We also can see that the sharper and faster we try to filter out a range of frequency components, the more spread out will the equivalent time response become. But the time response will not spread into the future so long as we can only filter imperfectly. This also means that a clever audio designer will be able to obtain spectacular filtering if he can predict the future. Since nature does not solve equations, we can "predict" the future by delaying the entire signal and looking at selected components "before" they come out of the delay device. This is a very useful circuit technique that has been used in communication circuits for over half a century.

Since we cannot have any finite region of one space that contains everything corresponding to a finite region of an alternate space, there must be some engineering tradeoff between the relative size of regions in alternate spaces that will contain "most" of the corresponding figures. And there is. For spaces coupled by Fourier transformation, the tradeoff is to choose regions that have an extent equal to the second moment of the corresponding signals [16]. When this is done, there is found to be a reciprocal spread of these regions such that their product is equal to or greater than a fixed value. This is called the uncertainty relationship, which denotes nothing more than the geometrically obvious fact that the smaller the region of knowledge we have about a figure in one space, the larger will be the spread of the region we must enclose to have comparable knowledge in the alternate space. It should be clear that not only is the uncertainty relationship an interspace mapping rule, but that this will be true for any dimensionality of space and for many other maps beside the Fourier transformation [3].

It is regretable that the conventional expression in signal theory has the same form as that of certain quantum mechanical operators, because it has been mistakenly applied as a forbidden limit in signal theory. There is a true indeterminacy in quantum mechanics which reflects the limited application of the concepts of classical physics to the microcosm. But there is no such limit in signal theory.

The difference lies in prior knowledge. Think of this in geometrical terms. If we knew everything about a figure, and someone were to hand us a piece broken off that figure, then we could not only identify that piece, but we could set about reconstructing the entire figure from only that piece. If we did not know everything about the figure, but had prior knowledge that it is a statue of a horse, for example, we could probably do a reasonably good job of reconstruction of those parts of the statue associated with the piece handed us. If we had absolutely no prior knowledge about the figure, then we could do no reconstruction—the piece must stand as is.

If we have prior knowledge that the piece must belong to a figure that, in that space of representation, must obey some physical rules, then nature has given us a clue. What we then do is determine what alternate space exists in which the properties of a figure are given characteristic signatures by the way in which the physical laws are also transformed into that space. We then can map our piece to that alternate form and match up properties to winnow down the alternatives from an infinite number (no prior knowledge) to a finite set (prior knowledge). In this way we are exploiting our knowledge of the way things work and are not operating under the limitations of an interspace mapping rule. To someone with no prior knowledge of what the piece represents, we can "violate" uncertainty by giving a more exact description than he can. We are, of course, violating nothing. We are more intelligently using the alternatives available to us.

When we map a frequency transfer function from the frequency domain to the delay plane expansion, we can separate each discrete signal path that corresponds to a different arrival time. In the delay plane we can pluck out one arrival time from all the rest and map that component to the time domain. To a person who must use the more direct Fourier transform map, we appear to have violated uncertainty because the time smear of the other paths with their imperfect frequency response has hopelessly overlapped the path we present to him. We have not violated uncertainty. We chose a map that let nature give us the necessary clues. That is one reason why the delay-space concept was developed by this author and elaborated at great extent. There are other multidimensional spaces, but the delay space makes certain audio and communication problems much easier to solve.

In more complicated situations we may find it necessary to map first to one space, there extracting what we want, then mapping the result to a second space and continuing this process until we have limited the alternatives to something we can handle. We subconsciously do this in most of our daily subjective decision processes. This is really what spatial filtering is all about. In communication analysis, the process of mapping a signal to its Fourier equivalent space for processing, then mapping the result back to the original signal space, is called spatial filtering [17]. That is a fortunate terminology, because we can expand that concept to all available spaces.

There is an enormous philosophical implication here which we will not pursue further at this time, but the more general concept of spatial filtering can be used to tell us more about a signal than we may have thought possible. As a small example, we know that a musical conductor can mentally isolate one instrument out of an entire orchestra and check its performance. This, the "cocktail party effect," and many other observations such as the left and right hemisphere brain dominance for differing acoustic signals and the demonstrated difference between language and music, have been used by some to infer that there is some capability within the human that transcends engineering practice. It does, if we limit ourselves to the type of filtering we normally use, but it is not unreasonable to infer that an alternate frame of reference can yield similar results. While we may not yet know how to do this, the fact that it can be done infers that even extreme differences between subjective and objective audio may eventually be analyzed by common methods.

EVEN- AND ODD-DIMENSIONAL FORMS

There is no fundamental distinction between the concepts of mapping between representations of the same dimensionality and mapping between representations of differing dimensionality. There are, however, some considerations which must be made if the mapping process corresponds to some situation in the physical world. One consideration relates to causality if the defining geometry is that of a dynamic process. The most general mapping functions are not constrained toward preferred coordinate behavior. As stated previously, "there is no inherent indignation in these transforms for a world with backward running clocks" [12]. The engineer must use judgement and select the parameters that correspond to the reality of the situation. As an example, an inadvertent polarity reversal of phase in the frequency response will yield, through Fourier transformation, an impulse response that runs backward in time.

A second consideration relates to the distinction that exists between even-dimensioned and odd-dimensioned spaces when a scalar solution is attempted. For example, there is a difference between the solution of the equation of wave motion in two and three dimensions and the general fact that a diffusion always occurs in even-dimensioned spaces, but may or may not occur in odd-dimensioned spaces. Another example is the even/odd dimension distinction in the scalar Radon transform [18] which, along with other methods of analysis, is now coming into use in the problem of reconstructing the density distribution of N-dimensional objects from a set of $(N - 1)$-dimensional integral projections of that density distribution (radio astronomy, NMR zeugmatography, x-ray tomography, ultrasound tomography, etc.). The distinction between odd- and even-dimensional behavior is eliminated if the proper complex representation is used. For that reason it is always advisable to maintain the complete form (such as amplitude and phase, or real and imaginary, or pressure and velocity) for all audio engineering processes.

PREDICTIONS

It is possible to use some of the geometric concepts to develop improved methods of handling sound-related signals. A few of these will now be discussed.

INTERCEPTING SOUND

If our eventual intent is to develop an acoustic holograph, wherein a sound field is to be duplicated in spatial extent and dynamic properties, but at a later time and in a different place, then we need to create and intercept the proper illuminating function. That was one reason behind this author's call for a Rosette Stone signal since it, in conjunction with the regular recorded program, can approximate an illuminating function for future generations willing to expend the effort, even through our present technology does not allow us the privilege of hearing our own recordings and its included signal as a holograph.

It works this way. The original sound intercept contains the acoustics of the enclosure and the frequency and angular defects in the microphone pickup. If we record a subsidiary acoustic signal that is capable of defining the enclosure and microphone properties, then this signal can be combined with the program intercepted by those microphones to generate, in two steps, a piecewise diffraction pattern of the program, a hologram.

It will be necessary to intercept both pressure and particle velocity in order to preserve the more complete field. These microphones do not need to be spatially coincident. Both linear and simple nonlinear properties of the microphones and subsequent recording process can be removed by deconvolving the subsidiary signal from the data.

RECORDING

A more efficient method of preserving sound-related signals was discussed under Scanning and Imaging. A global–local map using orthogonal sequences of signals will yield a record with properties considered remarkable by contemporary standards, even though we can perform all the necessary steps to generate such a record today. No conceptual invention is needed, but we will require the development of faster and more powerful processors than now available.

In contemporary recording we have become accustomed to some equality between program time and position along a groove or tape. This will not be necessary with a proper global–local map. The result is that the amount of record material required for a program with a substantial frequency range will be dramatically *smaller* than dictated by present methods. There will be no places on such a record corresponding to soft passages, loud passages, bursts of high frequency, or thundering bass. Instead, each moment of the original program will be spread into *everywhere* on the record. A piece of the record will contain the entire program. Ticks, scratches, and pops will not exist, no matter how badly the recording is mutilated, but the effect will be discerned as an increase in overall background noise. The signal-to-noise ratio will be a function of the size of the record. A large number of programs, or formats of programs, can be simultaneously impressed on the same record. The type and quality of playback can be dependent upon the complexity of the playback processor available to the user, much as it is today for four-channel, two-channel, or monaural playback.

The process of generating such a record may consist of computing the orthogonal mapping for the entire program, then transferring the result to a means of recording. The record could be a spiral groove, an optical diffraction pattern, a topographical surface, or any method capable of efficient duplication as a one-, two-, or three-dimensional object.

The process of playback may consist of reading the record into a processor, which then performs the inverse map and releases the appropriate signals corresponding to the original program.

As we stated, this is not blue-sky thinking, but could in principle be done today. In fact, in the limiting case of a point-to-point map, this is a description of our present records.

SUMMARY

We have touched on some of the things which happen when we change our frame of reference. The view was advanced that there is no preferred frame of reference for any situation. This may not seem very significant when given such a simple statement, but actually this has a dramatic impact on objective analysis. It means, for example, that an engineer who has carefully measured the impulse response of a device has an accurate representation of the linear properties of that device. He knows everything there is to know about that device, *in that frame of reference*. Another engineer may independently measure the same linear properties by using a different frame of reference, the steady-state frequency response. The second engineer also knows everything there is to know, in his frame of reference. The two measurements do not even look alike, but they each represent everything there is to know about the same device. Can the languages of the measurements be translated into each other? Of course, the process is the Fourier transform and a number of excellent textbooks are devoted to that subject.

Now comes the impact. These two frames of reference, joined by Fourier transformation, are not all we may use. They are only two out of an infinite number. What are these other frames of reference? Our textbooks do not talk about them, and the Fourier transformation is clearly of no use in changing our present measurements into whatever frames of reference these other viewpoints possess. How can we go from one frame of reference to another?

The basic process of mapping a representation from one frame of reference to another was shown to consist of multiplication followed by summation—sums of products. A figure in the initial frame of reference has its values multiplied everywhere against the values of a second figure (mapping function) in the same frame of reference, and the composite figure thus formed is summed over all of the appropriate coordinates to produce a new figure in a new frame of reference. It is the mapping function which determines what the new frame of reference will be. The new frame of reference may have the same, greater, or fewer dimensions than the initial frame of reference. Every time we change our frame of reference, or point of view, the form taken by this basic process is that of a summation of elemental products, and appears in engineering as the inner product, crosscorrelation, convolution, basic transformations, and the definition of generalized functions.

We then saw that an outcome of this was that Fourier transformation is a special case of a more general interspace map. Another surprise is that a hologram can be considered a special case of a more general form, which we called a holomorph. And a number of, what might appear to be unrelated concepts, were shown to arise from considerations of interspace mapping—changing the frame of reference. Thus the conventional uncertainty relationship between Fourier transform quantities is nothing more than an interspace mapping rule and is itself a special case of a more general interspace mapping rule. Apodization is seen as a similar interspace mapping rule.

We discussed the number and type of dimensions of a representation, which led to some more general considerations of scanning and imaging. And the significance of using the more complete representation of the energy density relationships in a process was tied to the use of the complex form which thus avoids difficulties when mapping between odd- and even-dimensioned spaces.

The real intent of this paper was not the presentation of a potpourri of mapping ideas, but the presentation of some useful engineering results which come from their use. Thus the call for a Rosetta Stone signal on our present recordings is revealed as a method of presenting future generations with a means of eventually creating an acoustic holograph, once they discover how to generate velocity as well as pressure fields. Geometric analysis had indicated that this possibility exists, but the necessary background material is only now being presented.

We have seen that a generalized holomorph is a more efficient method of recording information than spiral grooves or linear-sequence magnetic patterns. The fact is that we can create and process such holomorphs with contemporary technology. That does not mean we should change everything we now do. But the fact remains that a better way is within our reach, and eventually we may head in that direction. When that occurs, our present tapes and disks will become museum pieces, which was another motivation for the call for a Rosetta Stone signal on contemporary material.

The subjective person was not left out of the material in this paper. We discussed the geometric significance of distortion and indicated some methods of probing a signal to determine how the subjective illusion may be distorted to create an image that is warped from reality. All the material presented in this paper has been submitted to several years of scrutiny and comparison with observed data prior to submittal in order to minimize the likelihood of improper interpretation. That includes the measurement of distortion through modifications in the type of geometry, which was briefly touched upon in this paper. Preliminary results are very encouraging, but are too lengthy to include at this time.

As a final observation, we would like to emphasize that the material in this paper, and those previous to it, is derived from and for audio engineering. It is in no way an attempt to force principles of engineering and observations of perception into any existing mathematical structure. Rather, it is an attempt to find whether there is a mathematical structure capable of dealing with both physical and perceptual processes. We believe there is, but only time and experience will tell if abstract geometry is that mathematics.

APPENDIX

HOLOMORPHS OF AUDIO SIGNALS

The conversion of an audio signal into a holomorph so that it can be presented as a "record," and the subsequent reproduction of that record, will need to be done by software processing. In order to illustrate how this may be done, we will consider two simple cases: converting a time-dependent signal to a one-dimensional global–local holomorph that may be preserved as a modulated groove on a contemporary disk recording, and converting the same voltage to a two- or three-dimensional holomorph that may be handled by some other method of recording.

If a one-dimensional signal $f(x)$ is to be converted to a one-dimensional holomorph $g(\xi)$, we need a two-dimensional mapping function $m(x, \xi)$. The value of m as a function of the coordinates x and ξ can be considered to represent the height of a complicated surface with respect to the plane formed by orthogonal x and ξ axes. This surface is stored within the processor either as a lookup table or as a computable entity.

The nature of the mapping function is such that if we cut single slices out of the figure along either the x direction or the ξ direction, the resultant curve will be noiselike with a zero average value and an autocorrelation that is very nearly a delta function. Also, any two slices will be orthogonal so that the integral of their products will be very nearly zero. The mapping function is a two-dimensional pseudorandom function, the purpose of which is to yield a global–local map.

To make a recording we first read the entire program $f(x)$ into the processor. Once this signal is stored, the processor multiplies this entire signal against the signal corresponding to a slice of the mapping function parallel to the x axis and corresponding to a particular epoch ξ_i, then integrates the product of these two functions of x over the entire range of x to produce a single number $g(\xi_i)$. This process continues for slices over all the available ξ, and the numbers are connected to produce the signal $g(\xi)$. This entails a great many multiplications and summations.

As a note, if the mapping surface were an infinite-height "fence" running diagonal along the $x\xi$ plane, then each slice would be a Dirac delta function corresponding to the epoch in ξ. The g value produced by the multiplication and integration process then corresponds to the generalized function that defines a point-to-point map of $f(x)$ into $g(\xi)$. The output of the processor would then be identical to the signal we now use for recording. This would be an absurd way to generate a signal to be fed to a cutter amplifier, but by drawing this comparison we can see how our point-to-point recordings are a special case of more general recording methods.

The one-dimensional signal $g(\xi)$ will be noiselike in character, but can be cut into a modulated groove as any other program. The purchaser of such a record would find that it had unusual properties, however. First, the unprocessed playback would have the sound of white noise, much like the sound of an off-station FM receiver. Second, no single place on the record corresponds to a unique place in the time scale of the reproduced program. Instead, each point along the groove corresponds to the entire program, so that if the record were scratched, the reproduced program could never have the metronome like tick we now are accustomed to. The effect of a catastrophic scratch on the mechanical record will be an almost imperceptible increase in background noise throughout the entire electronically reproduced program. In addition, if the user were careless and dropped the record, breaking it into many pieces, he could retrieve the most convenient piece and fasten it to his turntable so that the reproducing stylus could still trace the portion of the grooves remaining. The output of his processor would still contain the entire program, but would have a, possibly audible, white-noise background.

In order to reproduce such a holomorph, we need a playback processor that is preloaded with the same two-dimensional mapping function relationship $m(x, \xi)$, since we will want to convert the holomorph coordinate ξ back to the coordinate x. The entire record signal $g(\xi)$ is played into the playback processor, and the inverse computation proceeds after the signal is stored. This time the processor multiplies the $g(\xi)$ against slices taken parallel to the ξ axis, but perform the multiplication and summation in the standard method to produce the original $f(x)$.

In the recording process we know when the program starts, but in reproduction we might need help in synchronization of the processor programs. We may want the freedom to be careless in placing the playback stylus on the record. In order to eliminate the need for a precise synchronization mark to align the ξ coordinate of the record with that of the processor, we can use the orthogonal properties of the mapping function and let the processor form a "sliding multiplication and summation," or convolution, during the reproduction. The processor can then perform a simple matched filter operation to align coordinates.

If a one-dimensional signal $f(x)$ is to be converted to a three-dimensional holomorph $g(\xi, y, z)$, we need a four-dimensional mapping operator $m(x, \xi, y, z)$. This can be generated in the processor as a sequence of three-dimensional mapping surfaces $m(x_i, \xi, y, z)$, which are slices of the four-dimensional figure. Each such slice then corresponds to a particular epoch of the coordinate x. The nature of the mapping surfaces is the same as that of the previous example in that the figure is formed as three-dimensional pseudorandom sequences.

In order to consider the method of recording, let us assume that the eventual holomorph will be a three-dimensional spatial figure which we can hold in our hand. This holomorph will have a defined orientation that will establish a base reference plane, and the audio program will be contained as surface shapes on this figure with the signal values corresponding to functional values of each point of the surface with respect to the reference plane. Let us assume that the space coordinates of the reference plane are y and z and that the space coordinate of the signal value is ξ.

We must emphasize that this particular shape of holomorph is only one of many we may consider. We could even conceptualize a solid holomorph in which each possible space position within and on the object conveys the information in an appropriate state variable such as magnetic domain orientation, optical opacity, or whatever, limited only by our inventiveness. Nor do we wish to infer that a holomorph we can hold in our hand is limited to three dimensions; there are many dimensional attributes beside that of Euclidean space which may be utilized.

The entire program $f(x)$ is read into the recording processor. This program is then sampled at each epoch x_i and the sampled value multiplied times the ξ-dependent points of the three-dimensional mapping surface $m(x_i, \xi, y, z)$. That is, the y and z coordinates of this surface remain unchanged, but the height above the yz plane is multiplied everywhere times the sampled value. If the sampled value is zero, then the surface collapses to the yz plane. If the value is negative, then the surface is pulled "inside out" in the manner of an umbrella inverted by the wind, and the height above the yz plane is multiplied times the magnitude of the sampled value.

This will produce a new set of three-dimensional pseudorandom surfaces, each one of which corresponds to a particular epoch in x. The processor then adds up all the heights at the corresponding y and z positions on each of these to produce a single three-dimensional surface with height measured along the ξ direction. This is read out of the processor as a voltage (or digital number, or whatever) as a function of the coordinates y and z.

We have several alternatives available to us in the generation of a physical object representing this holomorph. The ξ values could be converted to embossed heights on a solid base with dimensions y and z with miniature mountains and valleys as a vertical relieve structure. Or, we could use the ξ values as density modulations of a two-dimensional optical transparency or as magnetic domain orientations of a two-dimensional card. The reason for this is that we deliberately chose a mapping operator that could yield a simple conversion from three dimensions to two dimensions as an integral projection operator.

For the remainder of this discussion we will assume that an optical transparency is used for the holomorph. This optical transparency can be held up to the light and observed as a mottled pattern of optical values. This is not a hologram, it is a holomorph. We did not use an illuminating function, we used a general mapping function. Because we did not use an illuminating function, an acoustic program corresponding to a single impulse, such as a pistol shot, will not appear as a zone plate on our holomorph, but will be a "salt-and-pepper" distribution with no discernible structure whatsoever. A spatial array of fringes does not exist here. Had we gone back to the mapping operator and imposed zone plate dynamics on the three-dimensional map, the analog of wave propagation, we would have created a hologram and would see fringes.

The holomorph can be reproduced by using a processor that runs the process backward. The entire holomorph is multiplied times three-dimensional slices of the original four-dimensional mapping signal. These three-dimensional slices are made parallel to the x axis and thus sampled ξ values of the holomorph are retained in the yz reference system and multiplied times the ξ values of the mapping figure corresponding to epochs in ξ and summed to produce a number f. The sequence of these numbers is $f(x)$.

There are a number of methods that could be used for reproducing $f(x)$, including purely optical techniques. Any discussion of these would go beyond the intent of this paper, which is to show how geometric principles of interspace mapping can be used to predict better ways of preserving and processing audio signals. These special cases of holomorphs were presented to show what can be done. There are a great many variations on them, and it should be obvious, for example, that a one-, two-, or three-dimensional holomorph can contain a large number of programs in a great many dimensions (channels), limited only by processing economics.

An obvious technical advantage of a global–local holomorph over present recordings is the substantial independence of the reproduced program from the physical properties of the medium on which the recording takes place. First, the amount of material is far less than that needed for conventional recording. Second, the dynamic range of a reproduced holomorph can be exceedingly high and is principally limited by the processor. There is a direct tradeoff between the physical size of a holomorph and the signal-to-noise ratio for the same recorded program. This means that a film-chip holomorph that is the size of a business card may contain several symphonies in discrete four-channel sound, let us say. The signal-to-noise ratio on any reproduced program may be a large number, say 100 dB. If we were to tear off a small piece the size of a postage stamp and were to reproduce only that piece, we could still reproduce any of the programs and suffer an increase in noise related to the area of the piece we are using. A piece one tenth the area of the original holomorph would reduce the signal-to-noise ratio by 10 dB. A third advantage lies in the reduced distortion in the reproduced program due to physical nonlinearities in the holomorph. The optical gamma, for example, of a film-chip holomorph does not directly influence the linearity of reproduction, but manifests itself as a decrease in signal-to-noise ratio.

REFERENCES

[12] R. C. Heyser, "Determination of Loudspeaker Signal Arrival Times, Parts I–III," *J. Audio Eng. Soc.*, vol. 19, pp. 734-743 (Oct. 1971); pp. 829-834 (Nov. 1971); pp. 902-905 (Dec. 1971).

[13] N. Wiener, *The Fourier Integral and Certain of Its Applications* (Dover, New York, 1958), p. 17.

[14] E. C. Titchmarsh, *Fourier Integrals* (Clarendon Press, Oxford, 1967).

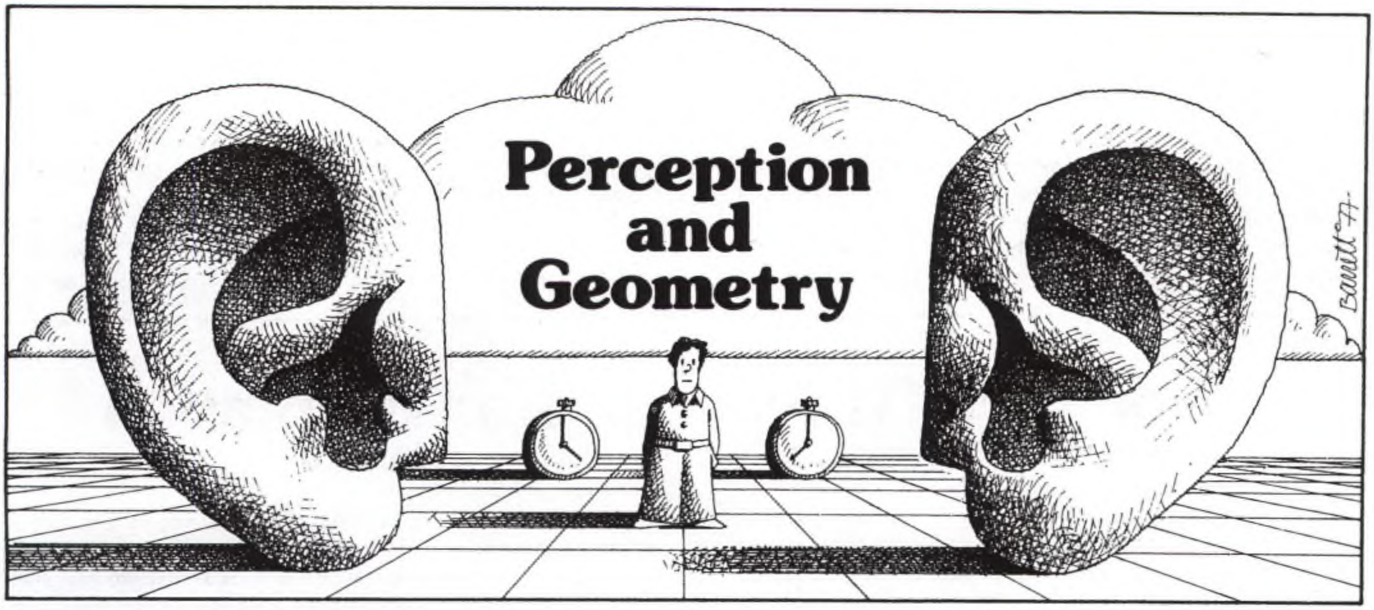
[15] G. Simmons, *Topology and Modern Analysis* (McGraw-Hill, New York, 1963), p. 36.

[16] D. Gabor, "Theory of Communication," *Proc. IEE (London)*, vol. 93, p. 429 (1946).

[17] A. Vander Lught, "Coherent Optical Processing," *Proc. IEEE*, vol. 62, p. 1300 (Oct. 1974).

[18] I. M. Gel'fand, M. I. Graev, and N. Vilenkin, *Generalized Functions*, vol. 5; *Integral Geometry and Representation Theory* (Academic Press, New York, 1966).

Mr. Heyser's biography appeared in the October issue.



Perception and Geometry

Richard C. Heyser

The perception of sound is a highly personal experience. It is neither art nor science, but our own private view through one of the windows of the senses.

We can share that view through words and actions, so we know that others experience it also. But it is left for fools like myself to dare sift and quantify the ingredients of that experience in some hope of understanding what it is and how to make it more enjoyable.

I have wondered, as we all have, how we might be able one day to put numbers on the stuff of perception. We are a long way from doing that. But in my own personal way I have been working on an allied problem. The problem of developing closer ties between what we measure in the physical world and what we seem to perceive of that same physical set of stimuli. I have come up with a few answers and I would like to share them with you. The results are applicable to audio analysis.

The technical details of what I am about to describe have been presented in a number of papers in the *Journal of the Audio Engineering Society*. In this article, I want to present the reasoning behind the technical details.

The basic idea is extremely simple. If we write down the most commonly used words which we all use to describe what we hear, we find that there is a definite structure to those words. We can arrange the descriptive terminology into categories which are reminiscent of a geometric framework. The words have a gestalt basis and are linked to relationships in the totality of our sense experience, including vision, taste, and touch. I therefore suggest that we should use geometry to probe the interplay of these word concepts.

Here, I feel, is a link between subjective perception and objective analysis. Rather than use numbers, we should invoke form, texture, and the relationships among things. Model perception with gestalt, and use abstract geometry to analyze gestalt.

The term abstract, as I use it here, refers to the analysis of "things" which are not named and quantified in the general analysis, but which can be named and numbered when we are ready to do so.

I would like to state that my approach was greeted with great excitement. I would like to state it, but I cannot. For one thing, the use of abstract analysis is in far left field, as far as most technical persons are concerned, if not outside the ball park altogether.

For another, the type of analysis that is required for even the simplest example in audio is pretty much uncharted. Among other things, we have to develop geometric tools for changing the dimensionality of an expression. And that's just for starters.

The Problem of Frequency

OK, where do we start if we want to apply the idea to audio? Well, I think the answer is easy. Start by cleaning up the mess we call frequency.

Let me state the problem. And in the statement I will give some of the answer. Then we can go on and develop the answer more fully.

The frequency description of a signal and the time description of that signal are tangled up with each other in a very fundamental way. The parameter that we call "time" and the parameter that we call "frequency" are not independent of each other. And no amount of Band-Aid engineering with running transforms or things called instantaneous frequency is going to change that fact.

Yet in subjective audio, we know darn well there is the property of pitch which is frequency-like, and that pitch can change with relative time. So if we want to apply the existing high power mathematics of time domain and frequency domain to what we hear, we seem to need a joint frequency-time description. Ultimately, when we try that trick, we run into the fundamental relationship between time and frequency, a relationship which we ourselves created from the definitions we gave these things.

But rather than blame ourselves, we choose to imagine that nature has intervened and somehow, magically, put a limit on the precision with which a codetermination of these parameters can be established. We even give that a name, the uncertainty principle.

What leads us to this rather strange action is a very real need for some kind of math that has a time-like and a frequency-like (and a space-like, and so on) set of properties which can all be used in the same description. Up to now our tool box of math relationships has only contained the parameters related by Fourier transformation. So we've been stuck.

And by the way don't think that this is a problem unique to audio. Other disciplines face a similar dilemma.

But audio has a driving force which other disciplines do not. Audio has people who listen, and listening is what audio is all about, no matter how much chrome plate we use on our equipment. The listening experience implies not only that there are coexistent parameters, but there are more than just two of them.

History of the Term

So much for the problem. Now for a little bit of history. In 1862, Helmholtz completed one of the finest texts on music and sound ever written. Highly successful, "On the Sensations of Tone as a Physiological Basis for the Theory of Music" was translated into English in 1885 and remains, even today, one of the finest discussions of the topic. It is still in print. To my knowledge, this is one of the first books to use Fourier series as a basis for analyzing complicated periodic signals.

The English translation used the phrase "vibration number" in the first edition to identify the number of vi-

Geometry Of Fourier Transformation

The appearance of anything depends upon the frame of reference we use to observe it. Geometrically, the Fourier transform is nothing more than a method of changing the frame of reference in such a way as to keep the number of dimensions the same but invert the units of measurement.

The Fourier transform is used in audio as the basis for converting time response to frequency response. In this case, the two frames of reference are one-dimensional. The unit of measurement of time is the second and the unit of measurement of frequency is the Hertz, which is an inverse time measurement.

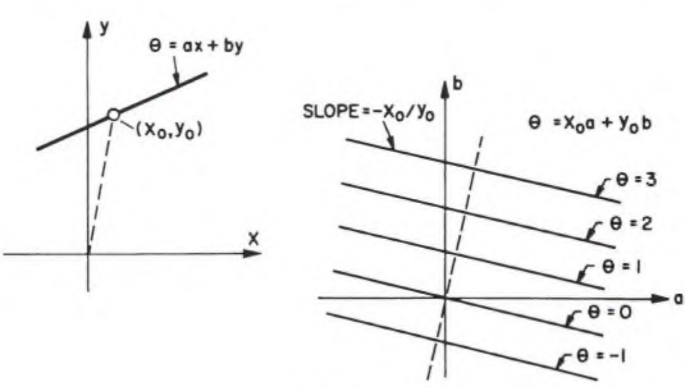


FIGURE 1

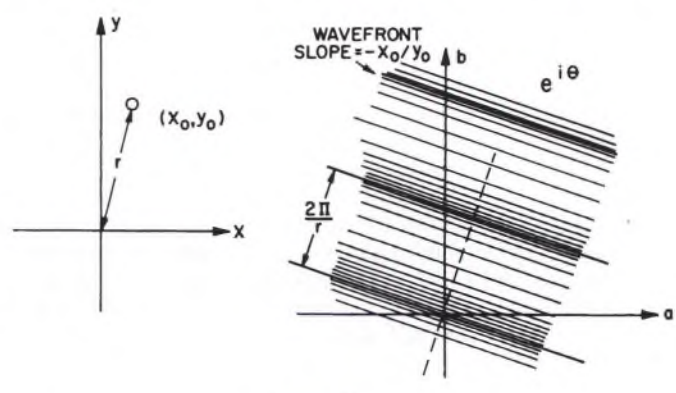


FIGURE 2

This novel geometric approach to the meaning of Fourier transformation can be more readily visualized in a two-dimensional example, as shown in these figures. In this example, a two-dimensional system, shown with coordinates a and b, is a Fourier transformed version of the two-dimensional system with coordinates x and y.

The requirement that the units of a-b and x-y be the inverse of each other shows up as the equation of a straight line, illustrated in Fig. 1. The parameter θ acts to spread the value of any particular line passing through a point in x-y, say x_0-y_0, into a specific line in a-b.

The parameter θ thus acts as a spreading operator that doesn't govern "how much" but does govern "where." If we want to find out how the point x_0-y_0 in system x-y appears to someone using the a-b system, we can pass a straight line through x_0-y_0 and rotate it like a propeller. This will sweep out all possible points in x-y, but only the common point x_0-y_0 will build up to the highest possible contribution in the a-b system when we add everything up.

When we do that, we find that a point in the x-y system appears as the wave $e^{i\theta}$ in the a-b system. This is shown in Fig. 2.

The geometric requirement shows no partiality. The x-y system and the a-b system are duals of each other. So a point in a-b will appear as a wave in x-y.

The x-y system and a-b system are different ways of looking at the same thing. Each part of a thing as described in the x-y frame of reference will appear everywhere as waves to a person looking at it in the a-b frame of reference. That is not magic, but a result of the way we defined the a-b alternative view of x-y. If we say that something appears precisely at a single place along the x axis, we cannot then turn around and insist that it also be located at a precise position along the a axis.

Everything involving Fourier transformation must submit to this point-wave duality. It makes no difference whether we started out defining things in terms of Fourier transformation, or discovered well along the road of other analysis that some of our parameters were Fourier transforms of each other. The fact remains that if Fourier transformation is involved, we will find that some of our parameters cannot be precisely codetermined. When this happens, and when other experience tells us that such parameters should be codeterminable, or appear to be codeterminable under other conditions, then we probably made an improper identification. The parameters are not what we thought they were. That is true of what we call time and frequency, as well as some other mysterious victims of the uncertainty relation.

brations a sound completes in a fixed period of time. The second edition changed that to "pitch number" so as to align it with the sensation of pitch as a numerical quantity. Fourier series were stated in terms of pitch number. The pitch number was also called "frequency" by the translator in that second edition, ". . .as it is much used by acousticians. . .".

Prior to that translation, 100 years ago in 1877, Rayleigh completed volume one of his equally famous *The Theory of Sound.* Two giant contributions to the knowledge of sound. Helmholtz preceded Rayleigh like a flash of lightning precedes the roll of thunder.

Rayleigh also needed a word to denote the number of vibrations executed in a unit of time. So Rayleigh called it frequency, stating that this word had been used for this purpose by Young and Everett. It is clear that Rayleigh equated the concepts of pitch and frequency, at least on a numerical scale.

Thus, while Helmholtz only used the term pitch number, his translator introduced the terminology "frequency". And since the translation occurred after the publication of Rayleigh's *The Theory of Sound* (which cited Helmholtz' German text in a number of places), it is possible that it was Rayleigh who really got this word started as applied to sound.

So what's wrong? Isn't it possible for a tone to change pitch with time? Of course, pitch can change with relative time. But frequency cannot!

The Fourier Transform

Now, let's do a wild thing. Let's use geometry to *derive* the mathematical relationship known as the Fourier transform. Then, from this geometric base, let's determine what the word "frequency" really means. And you won't find this in text books, at least not yet.

Let us begin to look at things geometrically. Suppose we want to measure something. How do we start?

In my opinion, the best advice on this matter was given by Albert Einstein who said, "It is the theory which decides what we can observe."

For one thing, it is the theory that determines the frame of reference we are going to use for the observation. A typical frame of reference for audio measurements is the passage of time, measured in seconds.

Having established this frame of reference we can set up instruments responsive in that system. An oscilloscope might be considered such an instrument. So we make oscilloscope measurements.

This next step is a big one. There is an infinity of frames of reference we can use. Each frame of reference is complete in itself and is a legitimate alternative for the description of an event. I call that the Principle of Alternatives.

If the passage of time is a legitimate frame of reference, then it is only one of an infinite number of alternatives. What might we be able to say about some of these alternatives?

In order to answer that, we need to take an even bigger mental step. We need to accept the fact that the alternatives may differ in the number of dimensions as well as the way in which the units are measured.

There is an infinity of frames of reference we can use. Each frame of reference is complete in itself and is a legitimate alternative for the description of an event. I call that the Principle of Alternatives.

Dimension? Yes. Consider the conventional waveform presentation of the signal coming out of an amplifier, volts as a function of time. Time in this sense generates what is geometrically called a "one-dimensional manifold." Each place in the dimension of time has a signal value associated with it. The distance between two places in time is measured in units we call seconds.

Suppose we want to change our frame of reference to come up with some alternate system of measurement. There are rules for changing the form of presentation from one frame of reference to another. The process of doing this is called a transformation.

If we transform in such a way that we do not change the number of dimensions, but have a new reference system measured in units which are the inverse of what we came from, then this very special transform is called the Fourier transformation. So it should be possible to transform our one-dimensional time measurement into a one-dimensional thing measured in inverse time, somethings per second. If we perform a measurement in this new frame of reference, we will call it the frequency response measured in Hertz.

For those who feel I am trying to pull the wool over their eyes, let us now actually derive the mathematical expression of the Fourier transform from these first principles of geometry.

I like to use pictures, so let me show how to derive the equation from considering the problem for some two-dimensional frame of reference.

In Fig. 1 let us assume we have a two-dimensional coordinate system, shown as x and y. This two-dimensional frame of reference is complete in characterizing something of importance. For example, it may be the reference system for a photograph with the distance between coordinate points measured in units of millimeters.

The Fourier transform of this will be another two-dimensional system in which the distance between two points corresponds to inverse millimeters. This is the a-b system.

The question is, how do we go from x-y to a-b?

We know the units are such that their product is a "dimensionless" value. (Millimeters times constant per millimeter is constant.) So let us say that the axis x will bear a special relationship to the axis a such that if we mark off some distance along x we will find that the thing that happens along a is a corresponding distance such that,

$$x \cdot a = \text{constant}.$$

And the same thing will happen between y and b.

What we have required is that the relationship between x-y and a-b be dimensionally reciprocal such that,

$$\theta = ax + by$$

The Greek letter θ stands for a fixed number, and it can be any number we choose it to be. I use the symbol θ because we are going to make that equal to the angle of something.

Look at this equation as some geometric curve in the x-y system. This is the equation of a straight line. The coefficients a and b in that equation determine the angle which the straight line makes with the x-y axes, and the constant θ determines where the line cuts across the axes.

There is a deep geometric significance to this relationship. The need for not changing dimension, but inverting measurements, leads to a zero curvature surface having one less dimension than the space in which it is imbedded. In two dimensions, this is a straight line. In one dimension, it is a point, and in three dimensions, it is a plane. Since most of our geometric thinking is done in three dimensions, this type of surface is called a plane when we are in three dimensions, and a hyperplane when we are in other dimensions. A straight line is a hyperplane in a two-dimensional system.

The general equation of a hyperplane is always the sum of products of coefficients and coordinates as we have written down. In three dimensions, there are three terms equal to θ. In one dimension, there is only one term equal to θ.

When we are comfortably seated in any frame of reference, the way we see the coordinate axes of the alternate Fourier transform view is as coefficients of hyperplane surfaces passing through our space. After all, the Fourier-transformed view is another way of looking at the same thing we observe, so we should be able to see the structure of the other frame of reference as something in our view.

Now, let's go back to our two-dimensional example and ask how we could take any place in the x-y system, x_o-y_o for example, and find out how it is distributed in the a-b reference system.

The relationship is in terms of straight lines (hyperplanes) passing through x_o-y_o. Each line passing through x_o-y_o tells what a and b coordinate locations will contain the information of all x and y values along that line. A neat thing happens. No matter what the angle the line makes in the x-y system as it passes through x_o-y_o, the result will be a straight line in the a-b system which has a constant slope.

If we want to find out how x_o-y_o (and only x_o-y_o) appears in the a-b system, there is only one thing we can do to the straight lines passing through x_o-y_o—we can rotate them around x_o y_o like a propeller about its shaft. And that's where we find the angle θ. We take the value of the signal at the point x-y and multiply it times the angle of all lines passing through that point to find out how that point is smeared over the a-b system.

If we write down the most commomly used words which we all use to describe what we hear, we find that there is a definite structure to those words. We can arrange the descriptive terminology into categories which are reminiscent of a geometric framework.

The mathematical expression for this is,

$$e^{i\theta}$$

If we write that out and see what it corresponds to in the a-b system, we find a startling fact. Each point in the x-y system is represented by a wave uniform over the whole of the a-b system. The period of this wave is the reciprocal of the distance from the point to its origin, and the angle of the wave in the a-b system is such that the wavefront is perpendicular to the angle the original point has with respect to its x-y axes. This is shown in Fig. 2.

I hope this rings a few bells, if not setting off sirens. The geometric relationship inherent in Fourier transformation is such that a point (particle) in one frame of reference will be manifest as a wave in the alternate frame of reference, and conversely.

Therefore (underline, exclamation point, big arrow), Fourier transformation is a local-to-global map, in which each point in one becomes everywhere in the other.

Now suppose we try a dumb-dumb and attempt to describe the same thing in terms of the x-y *and* the a-b system. Here is what happens. We can codetermine the location of a point in x and y, *or* in a and b, *or* along x and along b, *or* along y and along a. But we are going to run smack up against our own definition if we attempt codetermination along x and a or along y and b. Not because nature stepped in and pulled a curtain over our results. But because we are trying to violate the very conditions we set down to derive this particular transformation.

What form will that codetermination be stymied at? The form is determined by the equation of the hyperplane (which is another way of saying the equation of a wave) and is,

$$\Delta x \cdot \Delta a \geqq \text{number}$$
$$\Delta y \cdot \Delta b \geqq \text{number}$$

where the triangle means the extent of the range of parameter where most of the value of *the same thing* is concentrated.

Oh yes, the equation of the Fourier transformation.

We add up the contributions of each point in x and y, which is called integration. In two dimensions this becomes,

$$g(a, b) = \iint_{-\infty}^{\infty} f(x, y)\, e^{i\theta}\, dx\, dy$$

If you're not into math, don't worry about this equation. The equation is not important. The ideas that led us to the equations are what are important. And the principal idea, that can never be repeated too often, is that expressions joined by transformation are nothing more than *different ways* of describing the *same thing*.

The Meaning of Frequency

Now! What the devil does frequency mean? Frequency and time are *alternate* coordinate systems for describing the *same* thing. Frequency *cannot* change with time because frequency and time are different ways of describing the same thing.

In our haste to match sense experience with some *existing* mathematics, we have found a thing called frequency which has a pitch-like behavior, and we found another thing which has a time-like behavior and we use them. The greatest majority of the cases we encounter in audio have number values such that the interrelationship between frequency and this time-like parameter does not cause any trouble. And that is a soporific because we have lulled ourselves into the belief that there could not be anything else needed, or available, to handle any problem.

The concept of harmony, the agreeable combination of sounds, got its first mathematical treatment in the days of ancient Greece when the Pythagoreans observed certain numerical relationships in musical sounds. Two equally taut plucked strings harmonize only when their lengths are in certain ratios to each other. The musical intervals of unison, octave, fourth, and fifth are related to the numbers 1, 2, 3, and 4.

When Helmholtz and Rayleigh analyzed sound, they did so in an age-old frame of reference that tied sound to the passage of time. Fourier's theorem that any repetitive function could be generated by proper combination of sine waves, the shape of the purest tones in music, made everything fall into place. Nothing could be more natural than to use this mathematics for the analysis of complex sounds.

I do not believe that either Helmholtz or Rayleigh had visions of replacing the parameter of time with frequency. Frequency was a convenient expression that made a lot of sense in the analysis of tones.

Helmholtz and Rayleigh, and almost everyone after them, used some ready-made mathematics as a model that fit perception pretty well. We experience a thing we call time. We give it a symbol, t, and write equations using t. Juggling the equations produces a new parameter, which we call frequency. If we do not look too hard, this parameter called frequency seems to behave analogous to another thing we perceive, which we call pitch.

Here is the catch. The parameter t is not the time of our perception. Nor is the parameter ω the pitch of our perception. t and ω are mathematical entities that are different versions of each other. The theory decides the observation. If we set up an observation in the parameter t, we will get measurements in the parameter t. We can transform the mathematics in t to a mathematics in ω. If we set up observations in the parameter ω, we will get measurements in the parameter ω.

We can transform the mathematics in t to a mathematics using four parameters if we choose. And if we set up observations in those four parameters we will get measurements in those four parameters. That is the significance of the Principle of Alternatives.

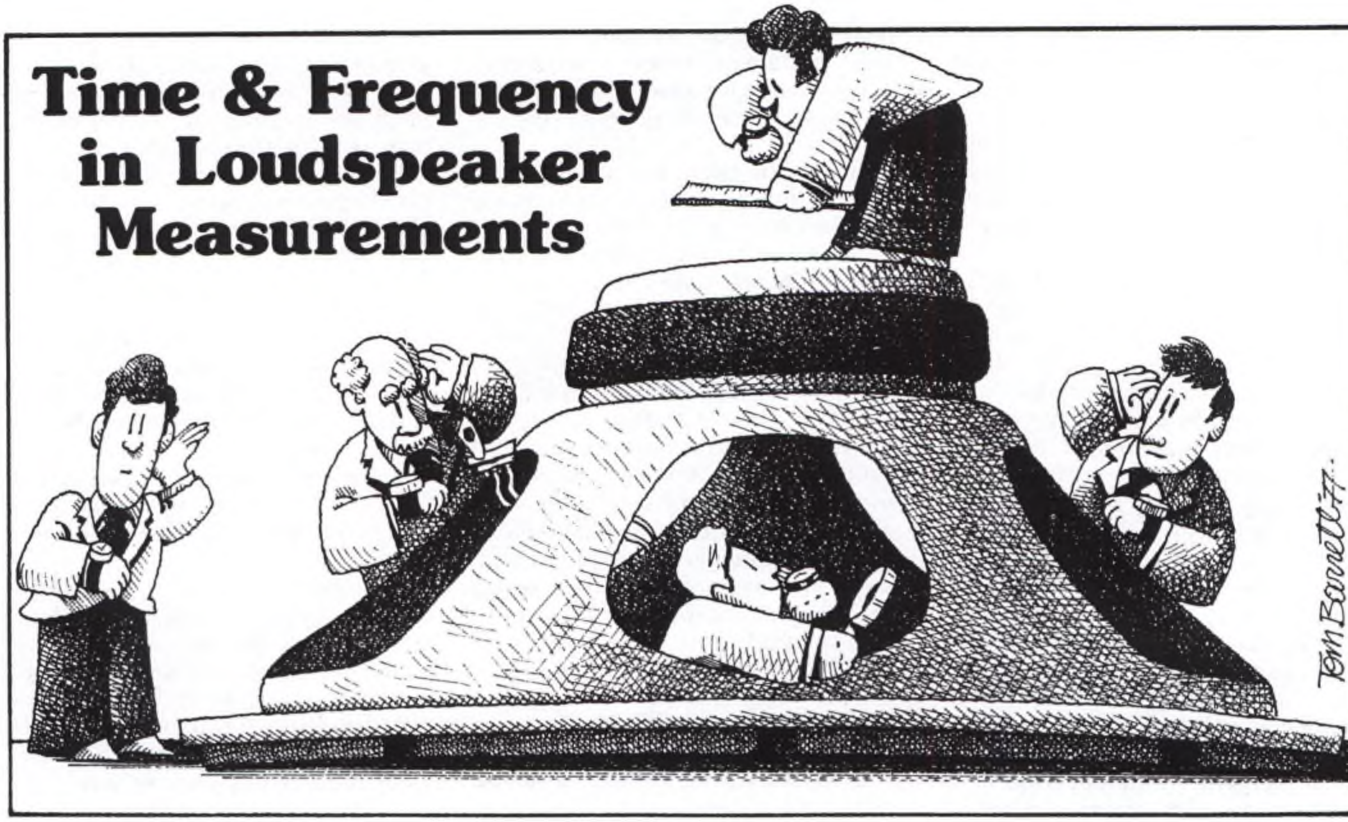
The fact that we can break out of the t-to-ω-to-t loop, which we call Fourier transform, is what is brand new in this theory.

It happens that the representations in t and the representations in ω do a pretty good job of modeling most of the things we need to analyze in audio. There are higher-dimensional versions of the t and ω representation, an infinity of them. Some of these versions have coexistent time-like and pitch-like parameters. The difference between the representation of a signal using these higher-dimensional parameters and what we get from glueing together a t and an ω axis to pretend we have higher-dimensionality is lost in the noise for most of what we do. For that reason, we might as well continue using the impulse response and steady state frequency response for loudspeakers, amplifiers, and the like. After all, the impulse response and the frequency response do have a meaning and they are legitimate measurements. It just happens that in detail the meaning is not what we thought it was.

But where we need to recognize the limitations of t and ω representations is when we get involved in the interpretation of these measurements with perception, which *has* a higher-dimensionality. It is then that the geometry is important.

Let me put this another way. You out there, Golden Ears, the person who couldn't care less about present technical measurements but thinks of sound in gestalt terms as a holistic experience. You're right, you know.

Time & Frequency in Loudspeaker Measurements



Richard C. Heyser

I claim that it should be possible to measure audio systems and have those measurements correlate with what we hear out of those systems. We are not doing that now. Our measurements are more precise than ever, but our understanding of what those measurements mean to the way a system "sounds" is still hazy.

I further assert that we are locked into that dilemma because we do not truly understand the meaning of those technical concepts which we now use. I don't think I can be more blunt about the matter.

OK. So having shot my mouth off, what am I going to do about it? Well, what I would like to do is present the readers of *Audio* with my personal view of the meaning of some of the more important terms we use in audio. These are some of the results from my own continuing research into the problem of finding out how to bring subjective and objective audio together. What I present here is my own work. I'm laying it out, in a put up or shut up fashion.

But I am not asking you to accept these things blindly. Question it, think about it, because what we really need to do is dig down to these underlying principles, the philosophy of the problem. In these discussions we go below the equation's mechanical formalism and question what the meaning is behind the equations. Then when we come up to the equations of audio we find that, while there may be no change in form, we often have a completely new perspective on just what they mean, not only to the pedestrian task of measuring components, but to the possible link with subjective perception.

In a previous article, I started out at ground zero and gave my interpretation of the meaning behind a technical term that is commonplace in audio, the term we call frequency. In this article I would like to carry this point further and apply it to the interpretation of certain loudspeaker measurements.

But before I get technical, let me put one thing into perspective. The end product of this whole multi-billion dollar audio industry is the listening experience. It is what we "hear," in the abstract sense of this word, that is important.

It is not the oscilloscope pattern but the listener's perception that is paramount. This does not mean that we should reject technology... quite the contrary. We know that most persons have the same general impressions of the realism and quality of a performance when listening to identical sound reproduction. There is something that is used by all of us in making our judgment, and that something is tied to the ingredients making up the reproduced sound. If this something is there but not specifically outlined in our present technical measurements, then we need to get even more technical and find out why. We need a Renaissance out of what may prove to be the "middle ages" of audio. The winner, if there is to be a winner, would be the listener, for we would know how to make his enjoyment of sound far better.

In my last article, I pointed out that when we do become very technical and poke around at the precise meaning of terms, a startling fact emerges. Even as fundamental a term as *frequency* turns out to have a meaning quite different from that which most of us employ in audio.

It is a subtle thing, but sometimes subtle things topple kingdoms. Let me recap. We know that at present there are two major ways of describing an audio signal. There is a time-domain representation and there is a frequency-domain representation. The time-domain representation and the frequency-domain representation are Fourier transforms of each other.

Now what the heck is a Fourier transform? A conventional textbook answer to that question is to write out a certain hairy integral equation and state..."that is a Fourier transform."

Simply writing down some equation, as though it were a Machine of the Gods, doesn't answer anything. Nature does not solve equations, people solve equations. Nature works in spite of us, and at best the equation is some sort of model for the way in which nature works.

In the previous article, therefore, I suggested a different approach. Suppose we have a signal which we agree is a legitimate time-domain representation. And suppose we ask ourselves what form that signal will take if it is observed by a being who uses some other coordinate instead of time. In particular, what would the form of that signal be if it has the same dimensionality but is somehow measured in units that are the reciprocal of the units of time we use?

Remember, we would both be seeing the same signal, but would be using different frames of reference.

Pursuing the point further, we asked what recipe we could use to take our time-domain view and see it within the framework of this other being's coordinate system. We derived the recipe, which turned out to be the Fourier transform. And the coordinate system which this other being uses turned out to be the parameter we call frequency. Exactly the same equation you will find in a textbook, but with a totally new interpretation.

The thing we call time in audio measurements and the thing we call frequency are different coordinates for describing precisely the same signal.

Subjective Descriptions

Oh, yes... ho hum, technicalia. But if we begin to think what this means to audio it gets a bit exciting, because this means that frequency and time are only two out of an infinite number of coordinate systems we can use to characterize a signal. We don't have to go just from time to frequency, we can go from time to some other coordinate. And even more stunning is that since we can have *either* time *or* frequency, but never both together in a meaningful description, this means that those properties of sound which we perceive and relate to the words "time" and "frequency" are *not those parameters at all.*

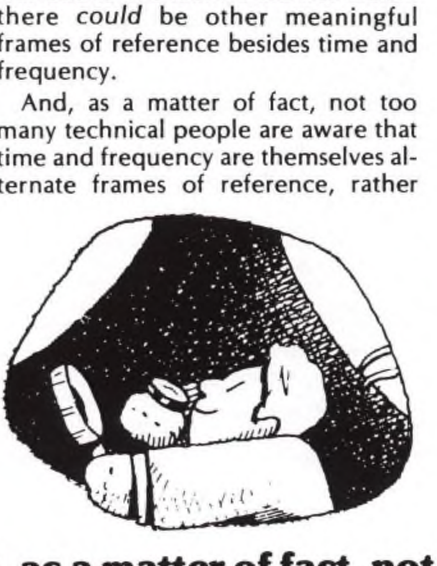
Now, think for a moment about those words we often use to characterize the sound of imperfect reproduction. Words such as "grainy" and "forward." These words do not seem to fit in with either an exclusive time description or frequency description. Is it possible that these words belong to some other, as yet unrecognized, coordinate system which is a legitimate mathematical alternative to time and frequency? I claim the answer to this question is yes.

Putting it in blunt language, if we measure the frequency response of a system, and do it correctly, then we know everything about the response of that system. We have all the technical information needed to describe how that system will "sound." But the information we have is not in a system of coordinates that will be recognizable by a subjectively oriented listener. Everything is there, but the language is wrong.

That is the root cause of the continuing fight between subjective and objective audio. It is *not* that *either* is more correct than the other... rather it is due to the fact they do not speak the *same language*. And when I say language, I do not mean just the descriptive words, but the very frame of reference upon which these words are based.

Sticking my neck out further, I assert that the reason technical people (and I am one of them) did not recognize the root cause of this problem was due to the fact we did *not* realize there *could* be other meaningful frames of reference besides time and frequency.

And, as a matter of fact, not too many technical people are aware that time and frequency are themselves alternate frames of reference, rather than just two terms to be applied haphazardly to measurement.



"And, as a matter of fact, not too many technical people are aware that time and frequency are themselves alternate frames of reference..."

There! How's that for tipping over icons?

Loudspeaker Tests

As a reader of *Audio*, you've probably noticed that our loudspeaker reviews have been a bit more technical than is normal industry practice. There's a reason for this. These tests are a first attempt to relate measurement to subjective perception. The various tests we perform did not just happen; each is in some way related to simple mathematical results in the type of geometric structure which we might use in perception. It is a first attempt, and very crude at that. But somebody's got to start the process, so let it be here.

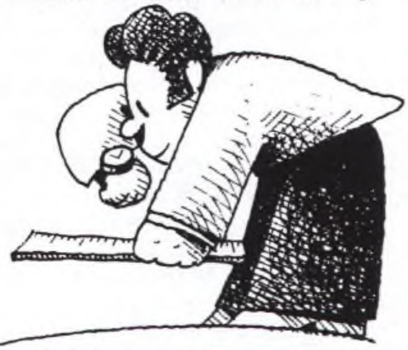
In the remainder of this discussion I would like to explain the technical aspects of spectrum sampling and apodization as they relate to the loudspeaker tests we perform in *Audio*.

Let me begin by recapping a very important concept which I flogged to death in the previous article. That is this mysterious and seemingly sinister thing called the uncertainty principle. There is nothing mysterious about the uncertainty principle at all. It is not something nature does to us, but something we do to ourselves through the definitions we give things.

Here is the point. It makes absolutely no difference whether we start out by defining parameters as being related by the Fourier transform, or somehow discover well along the road that two properties happen to be related through Fourier transformation: when two properties are Fourier transforms of each other, they represent different ways of describing the same thing and hence cannot be thrown together into one common description. The Fourier transform is a map, you see, which converts one coordinate system into another coordinate system.

It is a property of changing from one view to another that each part of one view becomes somehow spread over the entirety of the other view. In particular, the Fourier transformation takes a single coordinate location in

"Nature's clocks always run forward; at least, the most diligent searching has failed to reveal any experimental results to the contrary."



one view and makes it into a very special geometric figure in the other view, a figure which we call a wave and which extends over the entire range of coordinates in the other view. If we try to take a restricted range of coordinates in both views, we cannot do so and be precisely accurate. But what we can do is ask what the minimum ranges of coordinates are in *both* views such that "most" of the same information is contained in each. The form this takes for a popular measure of "mostness" is such that the product of these two ranges is greater than or equal to some number. This is called the uncertainty principle.

Let's see what this means in audio terms. Suppose we are testing a loudspeaker. We kick it with a voltage and the loudspeaker produces some sort of sound. Let's pick that sound up with a microphone and convert it back to voltage. Now let's put a switch in the output of the microphone. Suppose the switch is initially open, so that we do not have any sound signal to analyze. Some time after the loudspeaker puts out a pressure wave, we close the switch for one second and then open the switch.

What do we have? In the coordinate of time we have a signal that only has a sound-related value over a period of one second. We have created a one-second chunk of time. . .a time-domain representation.

Imagine, if you will, how that voltage would appear to some being who does not live in a coordinate called time, but whose frame of reference is something we call frequency.

In fact, if we want to see what he sees, we can convert to his coordinate system by making what *we call* a spectrum analysis. In order to do this, we have to give up the thing we call time. Time will show in this frequency spectrum, but it will be in the form of the relationship of phase and amplitude of waves in the frequency spectrum.

When we look at the frequency representation, we will see that there is some energy spread over the whole of the frequency coordinate. But the effect of having taken a frequency spectrum from a small chunk of time is that the frequency spectrum will be very slightly out of focus. The edges will not be sharp, but somehow smeared. The amount of this smear will be on the order of one Hertz, which is the name we give to the unit of measurement in this other being's coordinate system.

If we had only closed the switch for one-thousandth of a second, and then seen what our frequency-domain friend saw, we would find that the smear was of the order of one thousand Hertz units (I'm only talking in ballpark figures).

That is the manifestation of what is called the uncertainty principle. In performing *Audio*'s loudspeaker tests, I use a 13-millisecond time window to make the three-meter or room test. I want to find out what spectral components are found in that important time period which can establish some measure of timbre or tonal balance of the sound heard from that loudspeaker when placed in a room. This time duration derives from psychoacoustic tests. I cannot legitimately present any frequency measurements focused to an accuracy of better than about 100 Hz, including the range from d.c. to 100 Hz, because of the chunk of time which the data represents. To be safe, therefore, I only give data from about 200 Hz upward.

Apodization

Now there's this problem called apodization, which literally means "the process of removing feet."

When we hack off sharp edges, such as closing and opening a switch on a voltage, the equivalent transformed view will be blurred in a most unpleasant manner. There will be foot-like appendages, or sidelobes, which extend outward from each place where there should be a solitary frequency value standing apart from its neighbors.

Again, I must stress this is not due to some caprice of nature, it is due to our definition. If we hack off edges, and if we take a Fourier transform view, then we will find sidelobes. And I don't give a darn whether we measure the equivalent frequency response with sharp filters or with a computer FFT, our definition requires they be there. The theory determines what we will observe.

In order to minimize (we can never remove) them, it is necessary to do some sort of blurring or defocussing in the hacked-off parameter. The process of removing spectral feet by operating on the original data is called apodization. There are an infinity of apodization processes available, depending upon the type of corresponding blurring we are willing to tolerate in the apodized spectrum. Apodization usually consists of smoothing the sharp edges by using more of what is in the middle of the hacked-off distribution than at the sharp edges. *Audio*'s loudspeaker data is apodized with a nearly raised-cosine weight function when frequency response is plotted, and with a Hamming weight function when time-domain response is plotted.

Time Measurements

Nature's clocks always run forward; at least, the most diligent searching has failed to reveal any experimental results to the contrary. Where we poor humans get into trouble is when we start out from a frequency measurement and compute the corresponding time-domain response. If we have a *chunk* of frequency response, for example if we have no data above 20 kHz, then the time-domain response will be blurred.

In nature, the sharpest edge of all is at "now." A computed time-domain response will therefore spread before and after "now." The computed time-domain response will appear to predict the future... that is not really a prediction, but a blurred edge.

The energy-time loudspeaker measurement we make is a computation from the anechoic frequency response. We band limit from zero frequency to 20 kHz. In order to get the sharpest definition of discrete signal arrivals, such as due to diffraction from the edge of the enclosure, with the least amount of predictive "feet," we use an apodization function called Hamming weighting. Our measured sidelobes are actually down close to 40 dB below the peak giving rise to them. But you will still see what appears to be a predictive risetime prior to extremely sharp pulses.

As a matter of professionalism, we also check the loudspeaker impulse response by using a raised cosine pulse of voltage that has a 10 microsecond half-width. The loudspeaker impulse response is viewed on an oscilloscope and compared against the computed energy-time response to make sure all is kosher.

The reason for this belts-and-suspenders approach is due to a fact of apodization that, unfortunately, very few professional people seem to be aware of. Apodization, or a weight kernel, or whatever you choose to call it, has all the properties of the data to which it is applied. This includes the properties of amplitude and phase. In fact, we could take a converse view that the data is actually a weight kernel on the apodizing function.

Now, you know what happens when we take a Fourier transform of a product of two functions in frequency. The result is a time-domain convolution of what would have been the time-domain representation of each by themselves. They get all mixed up.

They get tangled up in phase as well as amplitude. And quite often a messy data signal will "unsmooth" even a good apodizing function. In short, this means that sometimes the computed response is lumpier than we think it should be. But, and computer people take note, unless you have a cross check or precise knowledge of the amplitude and phase of the data being transformed, you don't know it happened.

The geometry of this is too lengthy to go into here, but most apodizing functions used in Fourier transform analysis are non-minimum phase. Mostly they change the amplitude without changing phase. This includes Hamming, Hanning, and the rest. Historically, this is because the interest usually lays in the power spectrum (phase, what's that?). That works swell when the data is minimum phase. But when the data (in our case loudspeaker frequency response) has a maverick phase term, it can unsmooth a good apodizing function. Look at it this way, the effect is as though the loudspeaker response was minimum phase and the excess phase term was thrown into the weight kernel.

I realize that such talk might be highly confusing if you're not in the FFT business, but computer people ought to know what I mean. Other than my own comments in technical journals, I don't believe this fact has been pointed out before.

What it boils down to is that *Audio* makes every effort to be technically accurate, even if we are not terribly popular among some manufacturers when we do so.

Wrap Up

Let me wrap up this little discussion with two observations. First, if we really want to bring subjective and objective audio together, we need to get down to the fundamentals which can be highly technical. Second, with the editor's permission, I am trying an experiment with these discussions—in using words rather than mathematical symbolism, but I am not watering down the technical level.

Audio's readership covers the full range of involvement in the sound industry, from listener to researcher. Reader survey cards (yes, we do read them) indicate that many of you want more technical articles. And you like straight talk. All right, this was a trial balloon. Want more?

ALTERNATIVES

Richard C. Heyser

A fascinating and most gratifying thing happened recently. I wrote two articles—highly technical articles—that discussed the proper meaning behind the words "time" and "frequency" as we use those words in audio. I wrote those articles with the intention of stretching the mind, my own as well as the reader's. This mind-stretch seems to have aroused favorable interest, and a great many readers took the time to write and encourage us to continue with more such articles. Very well, I will try.

The Dilemma Of Perception

So there is no misunderstanding, let me state what it is I am doing and how it differs from tradition. I am conducting an exercise in thought by taking a wholly different approach to our present technology. I choose to form no *a priori* conclusion about the truth or falsity of the mathematical relationships which we now use. Instead, I am attempting to come up to those relationships from a different level of conceptualization. I feel that we should try to find out "why" as well as "how" nature appears to do the things it does. If you wish to call that a philosophical base, then I do not object.

The reason I search for "why" is simply that right now we have an intense dilemma confronting us whenever we attempt to correlate our physical measurements with human subjective perception. It is no laughing matter that what we measure does not always correlate with what we "hear." It is a disgrace which no amount of finger pointing will eliminate.

I start from the conviction that nature does not solve equations. If we are going to use math, then I want to know why that math should work. It is necessary to start from some primal base. The primal base I present in these articles is that through the study of relationships we can infer which exist within any structure of allowable description. The fancy name is abstract geometry.

It obviously derives from my interest in the subjective/objective problem, since one structure of description can be that of perceptual concepts, while another structure can be that of objective measures of the ingredients of that perception. Can we link the structures? That is my interest.

TOM BARRETT

Alternatives

A few years ago I published a theoretical concept which I call the Principle of Alternatives[1, 2, 3]. It introduces alternatives as a geometric concept. Alternatives are defined as those equally valid descriptions expressed in different frames of reference and corresponding to alternate spaces of representation. Geometrically, alternatives are different ways of looking at the same thing, and the set of all alternatives forms a universe of allowable descriptions.

What this principle asserts is that there is no preferred way of describing anything, either from the standpoint of dimensionality or units of expression. And therefore, there are an infinity of alternatives for any description we might make.

Abstract? You bet, but very powerful, because we can consider all known classes of description, including those in which all we can determine is the probability with which assignment of properties can be made within a particular frame of reference ("fuzzy" alternatives).

In the previous *Audio* articles I used this principle to give a new interpretation to the concepts of time and frequency and to the Fourier transform which defines them. Namely: the "time response" is one of the alternatives to the "frequency response" and the Fourier transform is that map which converts from one alternative to another.

In this discussion, I would like to carry this mind-stretching exercise a little further and ask you to think about the inner meaning which this brings to some relationships common to audio as well as other branches of science. Again I state my *caveat:* Do not blindly accept everything I state, think about them, mull these things over in your own mind. I believe it is far more important to convince oneself of the reasonableness of such things than to accept as dogma that which might later be overthrown. Besides, once one becomes accustomed to thinking about such things in the abstract, it is going to be much easier to consider the deeper waters I would eventually like to discuss on the implications this brings to reconciliation of the "meter" and the "ear."

Euclid, Hilbert, and Audio

In the previous article I pointed out that "waveness" and "placeness" are possible alternatives of each other. If, for example, we have set up a description in terms of any frame of reference, then it is possible to recast that description into another alternative frame of reference in which each place in the first becomes a wave extending over the whole of the second. One of the infinite number of ways of performing this conversion is that map which we call the Fourier transform.

Very well, let us turn this around. When two descriptions are related to each other through the Fourier transform, then these descriptions are alternatives of each other.

What special things might we infer about these particular alternatives? One of the first things we can infer is that the number of dimensions will be the same for each of them. There is no way that a three-dimensional description can be related by Fourier transformation to a two-dimensional alternative, for example. There are maps which connect three-dimensional alternatives with two-dimensional alternatives, but the Fourier transform is not one of these.

Let us now consider that special type of geometric framework in which everything obeys the laws of Euclidean geometry. All the postulates of Euclid, including the parallel line postulate, hold. A very little thought about the geometric basis for the Fourier transform, which I gave in the previous article, will reveal that each of the alternatives joined by Fourier transformation are Euclidean in nature. So a Euclidean space is transformed into a Euclidean space.

Let me pause right here and reveal a bit of where we can use this in much-later analysis. I contend that the thing we call distortion in audio, both objective and subjective, can be regarded as a warping of the geometry within a given frame of reference. The effect of distortion is to convert a Euclidean representation into a non-Euclidean representation, for example. This warping may possibly be handled as curvature tensors at some later time. But right now I want to point out that this talk of Euclidean spaces is very important to audio, and it is not part of a "snow job."

If what we are describing has a limit to its total energy, as all practical audio measurements do, then we can state that the proper sum of all energy components is finite. Many of the things we measure are such that their squared value is proportional to energy. Sound pressure is such a parameter; so are air particle velocity, voltage, and current. Not obvious now, unless you are into math, but the appropriate sum of the magnitude of such parameters squared is known as the Lebesgue square measure, denoted by the symbol L^2.

In striving to find some possible deep-seated meaning to properties, whether of perception or physical observation, we are led to search for the most general possible statements about those properties. Statements which are not dependent upon special objects of description, but determined by abstract relations. If we are really successful, our reward is the discovery that we have no words with which to adequately convey those abstract impressions. So we must often double up on the use of certain descriptive terminology which can invoke some appropriate mental analogies. The term geometry, as I use it, in these discussions, is one such word.

each of the alternatives joined by Fourier transformation are Euclidean in nature

Another such word is "space." In the abstract, the word space refers to a set of defined elements together with some agreed upon rules for combining those elements into the analog of a structural configuration. A multidimensional Euclidean space is a readily identified example. In this case, "space" means what we normally mean by the word space.

But there are other ways of defining elements and putting them together to form other "spaces." Another way of saying this is change the frame of reference. An example is what mathematicians call the Hilbert space L^2, the infinite-dimensional analog of Euclidean space.

Thus, one alternative for expressing finite energy signals is a space we can identify as a finite-dimensional Euclidean framework. Volts as a function of time is an example which uses the one- dimensional coordinate measured in units of time with the amount of volts at each moment of time being the number representing the signal at that particular coordinate location. Another alternative is the infinite- dimensional Hilbert space L^2 in which each possible form of signal which has finite energy is one of the coordinates, and "how much" of that signal is the position along that coordinate. Sure, it is abstract, but that is what Shannon brought into engineering and was the start of that very practical endeavor which we now call Information Theory.

As a technical point, we can thus observe that alternatives can be infinite-dimensional as well as finite-dimensional. As a mind-stretch, we should prepare ourselves to grasp the conception of infinite-dimensional spaces. The reason is that in the early parts of this century, mathematicians really began to develop tools for infinite-dimensional representations under the general name of Functional Analysis. There is a great wealth of knowledge to tap here, as Shannon did.

To give you some idea of how we might use it in audio consider this question: What is melody, or even a melodic contour? Stretch the mind a bit. If each allowable tone is assigned as a dimension, then certain groups of tones, bearing particular relations to each other, define subspaces of finite-dimensionality. These subspaces may be combinable in a different manner so as to form characteristic patterns which have *extremum* metric properties relative to subspaces formed from random combinations of tones. That is, the preferred subspaces are more densely packed with less distance separating members of the subspace. I do not know how that would work out on a number cruncher, or whether it may prove to be a silly idea. But the conceptual "distance" between certain notes, and I do not mean where they are on the musical scale but whether they seem to "fit" together, seems to form an attractive way of discussing chords and how they might fit together in the various combinations we might think of as melodies.

And tweak your imagination with this: Might it be possible that such a primal framework, which we could call a gestalt base of analysis, is also tied to other perceptual-observational disciplines, such as psycho-linguistics? Is there an analogy with Noam Chomsky's Theory of Transformational Grammar such that the perception of sound has a deep structure as well as a surface structure?

These are indeed important considerations, but discussion of these things lies well ahead of us. And we must get back to the fundamentals I wish to present in this brief article. With the definition of terms cited above and the appreciation for the geometric role that is involved, we can see that the Fourier transform defines a way of changing one representation into another in a special form-preserving manner. In contemporary mathematical language, the Fourier transform defines an isomorphism of the Hilbert space L^2 onto itself.

The consideration I want to place before you is that whenever we run across two types of description, both of which define a Hilbert space and are linked by Fourier transform, then these are alternate descriptions. They both describe the same thing. The coordinates of these two alternatives are versions of each other and can never be considered completely independent.

Observer-Observed

We may have thought, with deepest conviction, that we were assembling a description of an event (or process, or thing) in which there were two types of parameters, both of which were required for a complete characterization. But, if along the way, we discover that these parameters are linked by Fourier transformation, then nature is telling us that they are alternatives.

Should we persist and try to combine both parameters in a common description, we will discover that there is no way we can codetermine an infinitely accurate "place" on both of them. After all, a "place" on one of them is a "wave" on the other; that is what the Fourier transform means.

Those who believe in the adverse perversity of fate—the butter-side-down philosophy—might point out that somehow our attempt to measure one of them causes us to lose clarity in the other. Whenever we set up an experiment to determine one of them, our apparatus depends upon the other one to such an extent as to blur complete knowledge of both of them. The effect is stated correctly: We cannot measure one without calling the other into play, that is because they are different versions of the same thing. But we should never confuse effect for cause.

In audio we want very much to say that frequency and time are both needed to specify a tone. If we try to measure a complex tonal structure with a narrow bandwidth filter, we find that as the bandwidth gets narrower and narrower, the time response of the filter smears out to such an extent that we can no longer say when that frequency component occurs. To a pessimist it might seem that our very attempt at gaining precision in frequency was befouled by nature so as to lose precision in time. The instrument with which the observation was made seems to react with the signal in such a way as to disturb what we are observing.

In other words, if one were not aware of alternatives, it would be very easy to presume an oberver-observed limitation to our knowledge. Any attempt at disproving such an interpretation would be doomed to failure on any grounds that attempted to show there was, even conceptually, the possibility of a true infinite accuracy of codetermination of the parameters joined by Fourier transformation. After all, we got into the trouble by the definition which we originally gave these terms plus the assumption we made that they were wholly independent. Therefore, every possible counter experiment we might propose that provides indefinitely accurate joint parameter codetermination will get destroyed when properly analyzed.

What happened? What is wrong? Is nature mad at us because we tried to mix time and frequency? No, nature does not give a darn what frame of reference we choose to use. Nature does its thing whether we are looking or not. Based on this principle of alternatives, I offer the following suggestion: In our subjective evaluation we do indeed perceive properties that are frequency-like and time-like, and they do coexist. But the dimensionality of this alternative is higher than that of the alternative we use to model some of our simpler objective evaluations.

This does not in any way mean that time and frequency form subspaces in a higher-dimensional perceptual space. What I mean is that it is possible to map a one-dimensional space upward to a four-dimensional space if we so wish. Nothing appears in one space that does not also appear in some fashion in the other space. They are alternatives of each other.

A discussion of the mathematical relations for changing from one frame of reference to another when they have a different number of dimensions lies far beyond the points I wish to raise in this discussion. We will eventually get to that problem. But right now I offer this as a suggestion of a way out of the observer-observed dilemma when the properties we think should be independent are actually related by the Fourier transform. There may not be anything whatsoever wrong with the frame of reference, except that the properties are not what we think they are, belonging as they do to a lower-dimensional alternative.

Fifty Years Of Uncertainty

I leave you with this important fact to ponder. Exactly 50 years ago, Werner Heisenberg made use of the Fourier transform relationship between descriptions in momentum and descriptions in position, the Dirac-Jordan transformation theory, and he discovered a most puzzling fact. The narrower one made the region of confinement of a description in position, the broader became the region of confinement of a description in momentum. The complete derivation can be found at the bottom of page 180 in his now-famous paper.[4]

This particular relationship, which we now call the uncertainty principle, is, of course, absolutely correct for the reasons we have discussed. Surprisingly, little recognition seems to have ever been taken of the role played by the Fourier transform or of the implications which this brings to the interpretation of the inner meaning of that relationship.[5, 6, 7, 8] The inner meaning that the parameters which this relationship ties together are nothing more than different ways of describing the same thing.

REFERENCES

1. R.C. Heyser, "The Delay Plane, Objective Analysis of Subjective Properties, Parts I and II," *J. Audio Eng. Soc.*, Vol. 21, pp. 690-701 (Nov., 1973); pp. 786-791 (Dec., 1973).

2. ———, "Geometrical Considerations of Subjective Audio," *J. Audio Eng. Soc.*, Vol. 22, pp. 674-682 (Nov., 1974).

3. ———, "Perspectives in Audio Analysis: Changing the Frame of Reference, Parts I and II," *J. Audio Eng. Soc.*, Vol. 24, pp. 660-667 (Oct., 1976); pp. 742-751 (Nov., 1976).

4. W. Heisenberg, "Uber den Anschaulichen Inhalt der quantentheoretischen Kinematik und Mechanik," *Zeitschrift fur Physik*, Vol. 43, pp. 172-198 (1927).

5. M. Jammer, *The Philosophy of Quantum Mechanics*, John Wiley & Sons, New York, 1974.

6. E. Scheibe, *The Logical Analysis of Quantum Mechanics*, Pergamon Press, Oxford, 1973.

7. D. Bohm, *Quantum Theory*, Prentice-Hall Inc., Englewood Cliffs, N.J., 1951.

8. W.C. Price & S.S. Chissick, eds., *The Uncertainty Principle and Foundations of Quantum Mechanics, A Fifty Years' Survey*, John Wiley & Sons, New York, 1977.

Reprinted from *JAES*, vol. 26, no. 3, pp. 151–153, 1978 March.

Fuzzy Alternatives

RICHARD C. HEYSER

Tujunga, CA 91042

Previous papers introduced alternatives as a geometric concept [1]–[3]. Alternatives are defined as equally valid descriptions which are expressed in different frames of reference. Geometrically, alternatives are different ways of looking at the same thing, and the set of all alternatives forms a universe of allowable descriptions. In this communication we wish to expand on that concept and introduce the condition of probability of assignment as a classification of alternatives.

We suggest that the process of mapping, or transforming, a description from one frame of reference to another frame of reference can be considered as a transferral of probability.

When a frame of reference, or space, is such that the description within that space is totally deterministic, then that description will be said to be a *sharp* alternative. This means that each part of the description can be assigned with a probability of unity to some corresponding portion of the frame of reference. If the description has measure, then it will be said to be sharp except over sets of measure zero.

The mapping from one sharp alternative to another sharp alternative will be defined as a *jump* transform. This means that the probability of assignment is transferred in a discontinuous manner from one space to the other with no part of the description having a probability of assignment other than zero or unity. The term *jump transform* is used to identify the discontinuous transition in probability.

When the transition of probability can flow continuously, this process will be called a *blur transform*. Unlike a jump transform, the blur transform can be terminated between sharp alternatives. This forms a class of alternatives which one could consider to "fill in" all possible versions between sharp alternatives. This will be called the class of *fuzzy* alternatives. The term *fuzzy* is used here to designate the condition that the description cannot be made in a completely deterministic manner within its specified frame of reference. The concept of fuzzy subsets, due to Zadeh [4], [5], and their use is now firmly established [6] – [8]. We believe that the term as we use it here is consonant with the original intent and complements contemporary usage.

If a description has finite measure, then that can be used as a condition of unit total probability of containment within a frame of reference. For a sharp alternative, the total probability of finding the descriptive values in terms of that frame of reference is unity, and it is zero elsewhere. A jump transform transfers unit probability to another sharp alternative as zero probability is transferred into the vacated sharp alternative.

Heuristically we can think of this in the following manner. Imagine that the universe of allowable descriptions is composed of densely packed tesselations. Each tesselation defines a boundary for a sharp alternative description. If we have an event, and that event is describable in terms of a specific set of coordinates, the description must be assigned over this universe in such a manner that all of that description is contained within some definable boundary. If the boundary is that of one and only one tesselation, then the description is sharp. A fuzzy alternative may range over all or part of the allowable sharp alternatives, much as a unit volume of fluid spilled on densely packed floor tiles.

If we have established a sharp description in terms of a particular set of coordinates, then the description exists with a probability of 1 wholly within that particular tesselation and exists with a probability of 0 elsewhere. When we map a description from one alternative to another, we transfer unit probability of occupancy from the source tesselation to the destination tesselation; and we transfer zero probability from the destination tesselation into the vacated source tesselation. All other tesselations have zero probability in this operation.

From the standpoint of an observer who moves with the jump transform, he stands at one moment in a land with sharply focused edges and in the wink of an eye is transported to another land that looks different, but still has sharply focused edges. In the jump transform the description moves from one tesselation to another like a chess piece which is lifted off the board and placed back in a new location.

An observer who moves with a blur transform between sharp alternatives will see things about him go out of focus in terms of the source frame of reference as the focus is sharpened in terms of the destination frame of reference. When an observer is in a fuzzy alternative his blurred vision now contains a part of the source system as well as the destination system. Because his total probability of containment extends beyond the boundaries of any single sharp alternative, the probability of confinement to one or more of the allowed sharp coordinates is less than unity — the description is fuzzy — some edges are blurred.

An important example of the use of this theory occurs when the fuzzy alternative is distributed solely over the union of two sharp alternatives. This can be illustrated with a common example drawn from linear audio circuits. The finite-energy frequency representation of a signal is a sharp alternative. The time representation of that same signal is another sharp alternative. The Fourier transform, which defines an isomorphism of the Hilbert space L^2 onto itself, is the jump transform between these structured sharp alternatives. We can blur a description of that signal over the union of these sharp alternatives so as to combine the coordinates of both time and frequency in a common manner. This is a fuzzy alternative, which means that all we can specify, in terms of time and frequency, is a probability of joint assignment, with a total probability of unity when all time and frequency is considered.

The mapping function, defined in the earlier paper [3], which ranges over the union of both spaces, determines the assignment of probabilities. In the case involving Fourier transformation, the uncertainty relationship is an expression of that probabilistic situation. This was pointed out in the earlier paper by the observation that the uncertainty relation is an interspace mapping rule and is itself a special case of a more general interspace relationship. The interspaces referred to were the set of all fuzzy spaces between sharp spaces, and the mapping rule is one expression of the assignment of probabilities for the fuzzy alternatives involved.

Defining mapping as an assignment of probability, and from this the concept of fuzzy and sharp alternatives, provides a formal structure for many of the procedures we now use. The engineer's use of the word "frequency," for example, can be identified with a fuzzy parameter defined only in a probabilistic sense when other parameters are involved, but which approaches the sharp parameter when the probability of involvement with the other parameters approaches zero.

The blur transform allows us to blend from one frame of reference to another with absolutely no discontinuities encountered during the transition. This is true whether we pass upward, downward, or remain unchanged in dimensionality. Properties perceived about us while we are in a fuzzy alternative will combine or coalesce in such a manner as to give the appropriate dimensionality as we proceed in this transition.

I must caution, however, that when we use this concept of alternatives we must give up the idea that certain properties are invariants of any description. Among the properties that may change when we convert from one alternative to another are dimensionality, units of expression, metrizability, continuity, and "betweenness."

A description which is completely continuous and "well behaved" in an alternative at a particular level of dimensionality may become quantized with allowed and unallowed states in an alternative using a different level of dimensionality. Parts of a description may seem to disappear at certain places in such an alternative and reappear at other places with zero or low probability of being found at intervening locations. This seemingly bizarre behavior is due solely to the choice of frame of reference and implies only that observations made within that frame of reference will yield such results.

It is entirely conceivable that even in a familiar well-behaved frame of reference this discontinuous property may be identified under certain conditions. One such condition is that modification of the geometric structural rules which gives rise to the deformation we identify as distortion. A pure tone reproduced at ever increasing levels through a nonlinear process will give rise to distortion products. These distortion products, in the simplest case, may be found only at harmonic locations with any reasonable probability. Furthermore, the signal energy may appear to pass from fundamental to third harmonic with no measured second harmonic content as the level is increased. Admittedly, this is an oversimplified example, but does illustrate the situation.

This communication is a summary update of work in progress toward developing an analytical structure suitable for modeling both subjective perception and physical observation. The terminology of subjective audio often contains what can now be formally identified as fuzzy and imprecise descriptors. These descriptors, syntactically structured and appropriately identified, can form a rudimentary language capable of conveying meaning at an experiential level of a higher dimensional alternative. The geometry of alternatives can thus serve as a metalanguage for translating between subjective and objective descriptions of the same event.

REFERENCES

[1] R. C. Heyser, "The Delay Plane, Objective Analysis of Subjective Properties, pts. I and II," *J. Audio Eng. Soc.*, vol. 21, pp. 690-701 (Nov. 1973); pp. 786-791 (Dec. 1973).

[2] R. C. Heyser, "Geometrical Considerations of Subjective Audio," *J. Audio Eng. Soc.*, vol. 22, pp. 674-682 (Nov. 1974).

[3] R. C. Heyser, "Perspectives in Audio Analysis: Changing the Frame of Reference, pts. I and II," *J. Audio Eng. Soc. (Communications)*, vol. 24, pp. 660-667 (Oct.

1976); pp. 742-751 (Nov. 1976).

[4] L. Zadeh *et al.*, Eds., *Fuzzy Sets and Their Applications to Cognitive and Decision Processes* (Academic Press, New York, 1975).

[5] A. Kaufmann, *Theory of Fuzzy Subsets* (Academic Press, New York, 1975).

[6] S. Ali and G. Emch, "Fuzzy Observables in Quantum Mechanics," *J. Math. Phys.*, vol. 15, pp. 176-182 (Feb. 1974).

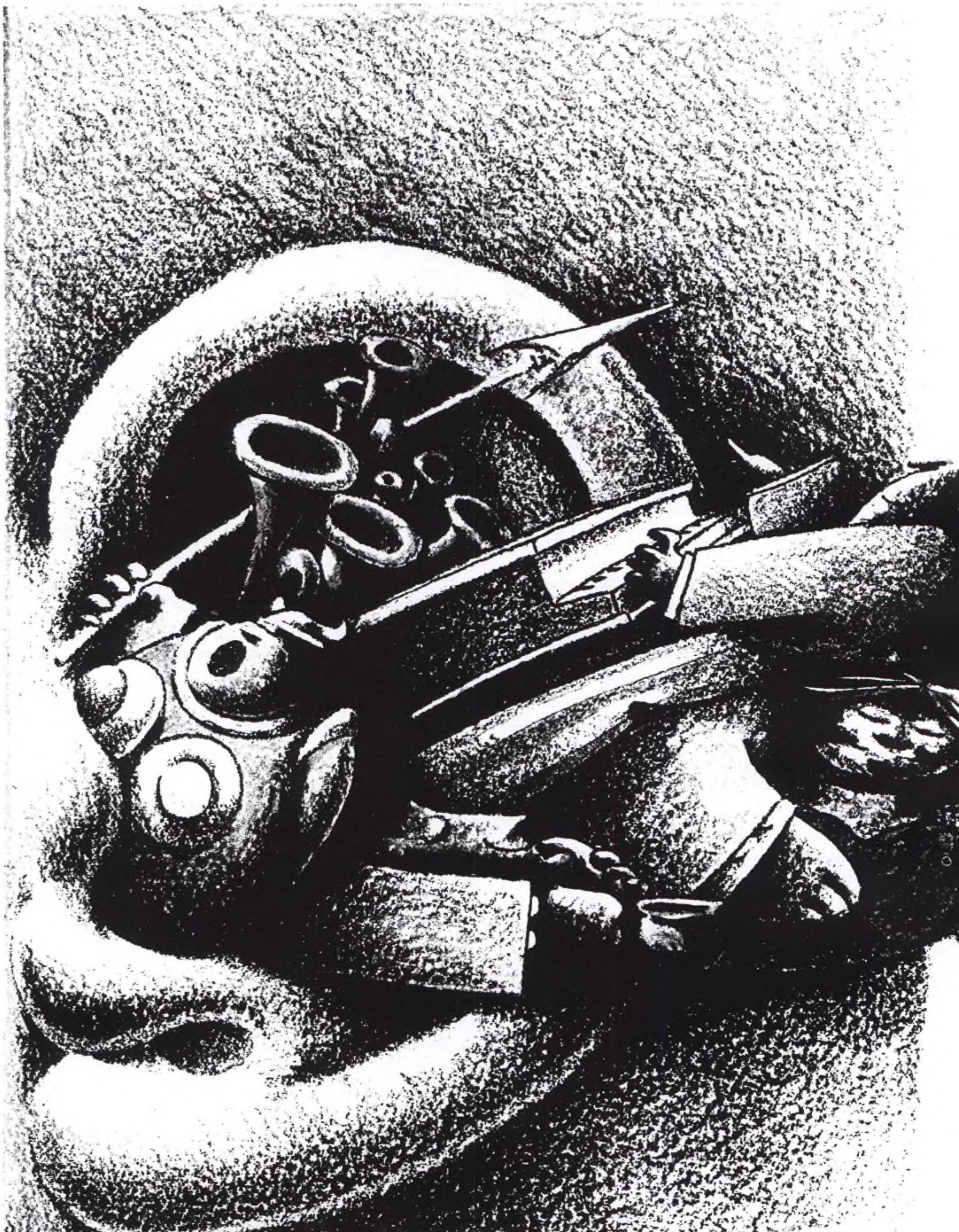
[7] E. Prugovecki, "Probability Measures on Fuzzy Events in Phase Space," *J. Math. Phys.*, vol. 17, pp. 517-523 (Jul. 1976).

[8] S. Ali and H. Doebner, "On the Equivalence of Nonrelativistic Quantum Mechanics based upon Sharp and Fuzzy Measurements," *J. Math. Phys.*, vol. 17, pp. 1105-1111 (Jul. 1976).

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, and the Acoustical Society of America. He also holds membership in the Institute of Electrical and Electronics Engineers, as well as Tau Beta Pi, Pi Mu Epsilon. Phi Kappa Phi, Sigma Pi Sigma, Sigma Xi, and Theta Tau.



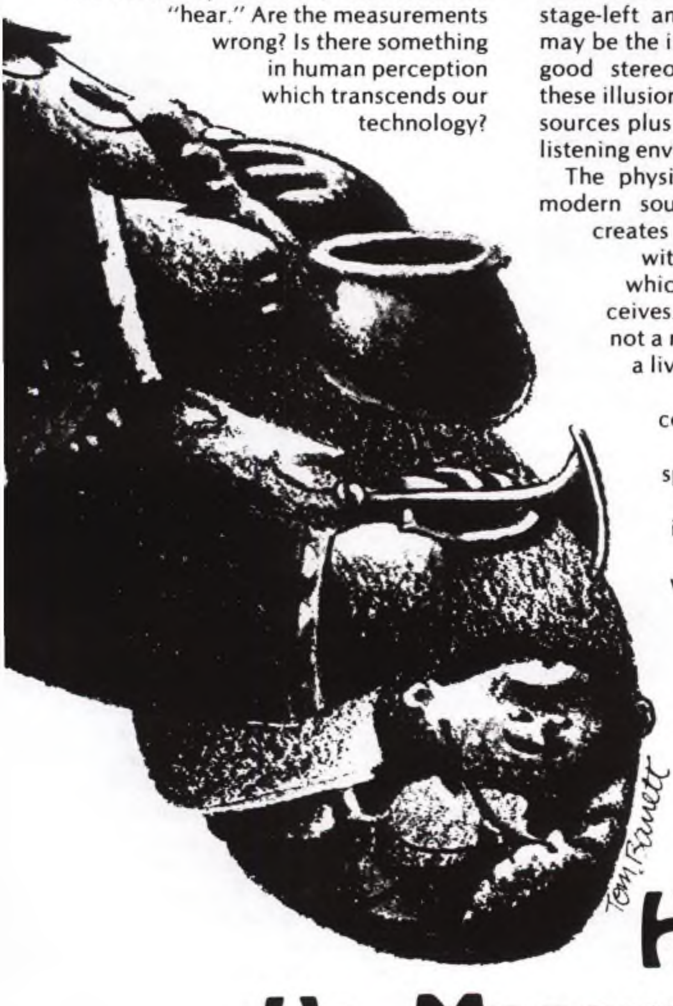
Hearing Vs. Measurement

Richard C. Heyser

Surely the end product of audio technology is the listening experience. We must never lose sight of this fact. No matter how exotic our instrumentation, no matter how impressive our mathematics, it is what we hear, not what we measure or compute, that is the final arbiter of audio quality.

But this does not mean that we should turn away from technology when attempting to assess or improve audio equipment. It means that we should become more aware of the proper role that is played by instruments and mathematics. For it is still the sole dominion of technology to give us objective and repeatable measures of our gradual climb toward perfecting audio systems. And until that day when we can quantify human experience and emotion, it is still our standard of improvement.

Yet we face a dilemma in modern audio technology: Our measurements do not always correlate with what we "hear." Are the measurements wrong? Is there something in human perception which transcends our technology? Are there "hidden variables" that we overlook? Or are we fooling ourselves by creating a mystique of the golden ear? Whatever your personal views on this matter, there is one thought I would like you to ponder. . .the effect that modern sound reproduction strives to achieve is the creation of an acceptable illusion in the mind of the listener.



Illusions

It takes no small amount of intestinal fortitude to stand up and tell an industry striving for technological perfection that what we are really trying to do is create an illusion. Yet that is the inescapable conclusion to be drawn from analysis of our present situation.

Almost without exception, the physical sound field in a listening environment could not in any way be created by actual sound sources located where we perceive them to be. There can be no stage-center vocalist located between our stereo loudspeakers and 10 feet behind a back wall. There can be no string section stage-left and 30 feet back. Yet that may be the illusion we perceive from a good stereo reproduction. We fuse these illusions from two discrete sound sources plus internal reflections in our listening environment.

The physical sound field which a modern sound reproduction system creates is definitely not congruent with the apparent sound field which we hope the listener perceives. What a listener "hears" is not a reconstructed hologram of a live performance. Instead he is subjected to a carefully contrived sound field which is intended to stimulate a specific type of perception. The listener is not a dupe in this circumstance, but is a willing participant who will often knowingly reject interfering sensory cues that would otherwise damage the illusion. The enhancement of this illusion, as a commercial enterprise, involves art as well as science. . .psychology as well as physiology.

Ingredients of Listening

Consider the ingredients of this listening experience. Let me define perception as the awareness of the world about us which we gain principally through sensory experience. While the sensory stimulus may be the result of independent processes, the perceptual image which we fuse from these senses is combined within the higher levels of the mind into an interrelated structure. Sensory imagery involves a multi-dimensional structuring in which ordinary physical space comprises only part of the dimensional frame of reference against which we form perception. This perceptual structuring is based on physical and emotional experience and is such as to align the majority of sensory experience at any moment with a consistent world-picture in our minds. The perception of sound involves more than just what we "hear." It is a holistic experience that involves not only the other senses, but past experience and present emotional state as well.

Not all of the sensory stimuli or prior experience need necessarily agree in order for us to form a perceptual image. Consider the art of a ventriloquist. Through manipulation of acoustic, visual, and associative relationships, a ventriloquist can project an illusion quite inconsistent with reality. I suspect that a ventriloquist would find it difficult to confuse a blind person.

This deception, which we find so entertaining, indicates a deep structural compatibility within human perception. Not only can we cope in a world which presents us with a continuing barrage of sensory stimuli, some of which can be misleading, but we can willingly "shut out" certain cues in order to enhance our perception. These are things that I, as a reviewer, must recognize whenever I form a value judgment on the listening qualities of a product.

On a survival level, the structuring of perceptual cues should relate to physical reality. A cave man would have been easy prey for a tiger which was seen on the left, heard on the right, but could bite him from the rear. We align our perceptual cues into a metaframework which I have referred to as a "rightness of perception." But on a more leisurely level, otherwise significant structural cues can be slotted into a lower hierarchical level of importance to perception. The world-image which we fuse in our perception may seem quiet real to us, but it does not necessarily coincide with ingredients of a physical reality.

Not A Hologram

The often overlooked art of recording lies in knowing how to structure the acoustic cues so as to enhance either the illusion of reality or the evoked emotional experience. Simply sticking microphones in a place where recordings are to be made will not do it if we want the proper listening experience from our present reproduction technology.

I must point out that it is theoretically possible to record a dynamic diffraction pattern of an acoustic performance—a hologram. Some day we will do that, but it is *not what we now record.* We do not record a hologram; we do not even pretend to record a hologram. Nor do we even pretend to play back a reconstructed holographic sound field. Yet, I submit, much of the hoopla of present audio component measurement technology is based on the assumption that we listen to a reconstructed hologram.

The component designer who, in good faith believes he is thoroughly measuring the performance of his product, tacitly assumes that perfection is a reconstructed hologram. He then compounds this problem by using distortion measurements which are based on linear mathematics (I will have more to say on this important matter at another time). When a nontechnical listener hears this product as part of a modern sound reproduction system, he may perceive an unpleasant warping of the illusion. It is distorted as a perceived experience. The designer is enraged that his product—which measures double 0 nothing per cent "distortion"—can be perceived right through the much higher measured distortion of other components. Obviously, in the technologist's eye, the non-technical listener is a freak to be ignored. Particularly, since this listener uses weird words such as *furry* to complain about the imperfections he perceives. The result is an industry split right down the middle, with math and fancy test instruments on one side and dissatisfaction expressed in flowery rhetoric on the other side.

Both sides of this controversy have been squared off against each other for at least 50 years, and neither will give an inch to the other.

It is my opinion that before we can try to answer the question "why can't we measure what we hear?" If we do not know what we are trying to do, then how can we expect to know how to do it better?

> **We align our perceptual cues into a metaframework which I have referred to as a "rightness of perception."**

I submit that what we are trying to do *in today's technology* is provide a particular type of listening experience under the limitations imposed by our ability to recreate a physical sound field. A great many years from now we will be able to record and reproduce an acoustic hologram, assuming that this is what the listener wants.

Meta-Language

Once we recognize that the actual sound field in a listening enviroment is not identical to the sound field which we may perceive, we get a whole new perspective on the problem of being able to measure what we hear. It is the illusion of reality, not the reality itself, that we must measure.

Now, I know that such a statement may turn a lot of people off, but do not be misled by any emotional reaction to my observation that the listening experience involves the structuring of an acceptable illusion by means of artfully contrived sensory cues. This does not mean, as I pointed out earlier, that we must abandon technology. It does not mean that at all. As a matter of fact it directs us back toward technology of a much higher level.

Consider this: Our entire multibillion dollar sound reproduction industry depends, in one way or another, on the observation that most persons will experience the same type of illusion if subjected to the same type of stimulus. Stereo would have been a total flop if the illusion of lateralization and depth were a random occurrence among the listeners. In other words, there is a commonality of structuring which shows the promise of being analyzed by a higher level of technology than that which we now use.

This higher level of technology might serve as a meta-language which we can use to translate between certain objective and subjective descriptions of the same event. And isn't that really what we want to do if we are to correlate what we measure with what we hear?

Language of Perception

Let me pursue that particular point a little farther. If human perception is structured in the manner I indicated earlier, then any attempt to convey information about personal impressions of a perceived experience might use terminology dependent upon that structure. The language of perception may depend upon inter-sensory analogies of form. We might describe our impressions of a sound in terms of shared experiences of sight, touch, taste, or smell, as well as sound.

A language capable of conveying information about our perception may be syntactically structured to evoke the appropriate sensory imagery. Seen in this light, the symbolic, often flowery, terminology of subjective audio begins to make a bit more sense (no pun intended). There is a language here, and words such as *sharp, bright,* and *furry* do convey meaning at an experiential level.

But if this language of perception is based on structural rules derived from, or consistent with, physical experience, then there is a conceptual link with objective measures of the ingredients of that physical experience. But, let me come at this from another direction. If, as Einstein cautioned, it is the theory which decides what we can observe, then the frame of reference establishes the form which that theory will take. When two observations are related to the same event but use different frames of reference (such as our perception of a measurable sound field) then there is a conceptual link between these frames of reference if the observations are internally self-consistent.

If we "hear" the same sort of thing every time we listen to the same set of physical stimuli, then, somehow, the measurements are related to what we hear. But that relationship is never a congruence when the frames of reference are not congruent. It is a foolish person who will draw conclusions about the "audibility" of certain technical flaws in the physical reproduction based on limited "listening" tests and ignorance of the possible differences in the frames of reference.

The first step we must take in quantifying perception is to learn the cipher of its language. But we must do more than just compile a dictionary of terms, because such a list of terms will remain a book of "seven seals" unless we try to understand the structure to which this language is applied—the frame of reference.

Altered Awareness

There is a final point I would like to address in this brief discussion. The illusion which we strive to achieve in the mind of the listener does not have to be an illusion of physical reality. The illusion can be that of an emotional experience based on a frame of reference in which the ingredients of physical reality are of minor importance.

No two persons need necessarily have identical frames of reference for perception. Indeed, our individual frame of reference can evolve and change from one time to another and from one situation to the next.

This altered awareness may be the result of a deliberate act on the part of the observer, or it may evolve quite subconsciously as a result of experience, training, or even emotional state.

I present this conclusion with no intent of becoming embroiled in philosophical discussions of: "What is reality?" or "How do I know that you hear a C major chord as I hear a C major chord?" Instead, I am sticking my neck out and presenting certain technical interpretations drawn from a transformational geometry based on the concept of frame of reference.

This structure of perception or conscious awareness, or whatever you choose to call it, is all too frequently overlooked when we consider the superficial technical aspects of audio. But these things are there when we really strive to understand what it is that we are attempting to do in audio—when we realize that the end product is the listening experience. ▲

preprint no. 1350 (F-1)

IMPRECISE DESCRIPTORS

By

Richard C. Heyser
California Institute of Technology
Jet Propulsion Laboratory
Pasadena, California

**presented at the 60th Convention
May 2-5, 1978
Los Angeles**

AN AUDIO ENGINEERING SOCIETY PREPRINT

IMPRECISE DESCRIPTORS

Richard C. Heyser
California Institute of Technology Jet Propulsion Laboratory

ABSTRACT

The fuzziness and imprecision of terminology used in subjective audio has often been the subject of ridicule by the objectively oriented engineer, who will cite the sharpness of mathematical description as a justification for abandoning subjective methods. It is suggested in this paper that the converse prevails: sharpness of description indicates that only a restricted class of alternative description is being utilized. A new class of descriptor is introduced in which only the probability of assignment of properties can be made, and sharp descriptors are considered limiting cases in which probabilities of assignment approach unity or zero. This expands our analytical capability to include the type of description which subjective audio has been using all along.

INTRODUCTION

Words such as "grainy", "bright", and "velvet" punctuate the language of subjective audio. This has been a source of irritation to the technical person who wants to quantify audio systems and put a number on each measure of performance.

It is true that some of the terminology of subjective audio has held a meaning only for the author who created that terminology. But there is reasonable evidence that subjective descriptions do convey meaning at an experiential level.

It is one object of this paper to propose a mathematical model for domains of valid representation which, by nature, use imprecise descriptors for characterization of signals within those domains. It will be brought out that even the most exact mathematical characterizations which we now use can be considered special limiting cases of these more general inexact representations. This introduces probability of assignment as a new property encountered when descriptors are mapped from one domain of representation to another, and suggests that programs which are directed toward objective measures of subjective properties may require the introduction of probability and likelihood estimators.

- 1 -

BACKGROUND

LIKE A FLY'S EYE

The concept of alternatives (1) is key to understanding the relationships which we are about to discuss. An alternative is literally "another" way. The Principle of Alternatives recognizes that there is never just "one" way of describing or characterizing any situation. There are many ways. In fact there are an infinity of ways and each is a valid alternative for all the rest.

When we set up a frame of reference for the process of making an observation, we have determined which of the infinity of alternatives will be used for that observation. The coordinates of that alternative become the coordinates of our observation.

It is quite difficult for us to visualize this abstract concept of alternatives since it transcends our normal experience of the meaning of coordinate systems. But what we can do is draw crude analogies with certain geometric shapes with which we are familiar. In that way certain relationships which exist between alternatives can be partially illustrated. One such analogy is illustrated in figure 1. We can liken the universe of allowable alternatives to a large mosaic made up of densely packed tesselations. Each tesselation represents a valid alternative, and each alternative represents a different view of the same scene. In a symbolic sense we can liken the mosaic of tesselations to a fly's eye. The same scene is viewed in each of the many tesselations but each view uses a different geometric frame of reference.

It is easy for us to comprehend the process of a fly's eye since each of the facets of the eye (tesselation) uses the same number of visual dimensions and has nearly the same coordinate system. But in the abstract model which we need to use for the purpose of gaining a deeper understanding of audio processes, imagine that each tesselation may have its own individual dimensionality and frame of reference. Again, each tesselation represents an alternative view of the same scene, but as observed from alien coordinate systems. The infinite mosaic of tesselations is now too abstract to visualize, but we can symbolize the concept by imagining it as a multi-dimensional extension of the fly's eye.

- 2 -

COORDINATES

When we write a precise mathematical description of a process, we erect a symbolic structure within one of the tesselations. The very process of writing a description requires that we presume some coordinate system in which that description may be expressed. This is so fundamental to the process that we rarely stop to analyze what it means. We are usually so busy trying to write equations that we forget the implied meaning of the coordinate system in which these equations will be written.

Failure to recognize the role of the coordinate system, and of the particular frame of reference on which the system depends, usually leads to a terrible mistake. If we put nature to the test by comparing the results of an experiment with what our equations predict must happen for that experiment, we will get positive results if our equations have been properly written. The fact that nature does what we think it should do for our chosen coordinate system then leads us to assume that the coordinate system we used is the coordinate system of nature.

Nothing could be farther from the truth. All it means is that this is the way nature behaves when observed from that particular alternative view. If we were to erect a totally different structure within a different tesselation we would find that nature would also conform to laws as expressed within the framework of that tesselation.

Figure 2 symbolizes one possibility we can encounter in audio. For simplicity, everything is assumed to be characterized by linear mathematics, although the concept we are proposing is not limited to linear systems. Three tesselations are shown: a one-dimensional time domain, a one-dimensional frequency domain, and a four-dimensional domain of unspecified name. Each tesselation bounds what will be referred to as a sharp alternative. This means that exact equations can be written using the coordinates defined within that tesselation. A being who erects a framework of analysis within the "time domain tesselation" will always get correct answers. To that being, signals are always functions of time. Another being erecting a framework within the "frequency domain" will also always get correct answers and must assume that signals are always functions of frequency. A third being uses a four-dimensional domain and, since signals always appear four-dimensional, naturally assumes that nature is four-dimensional.

Which one is correct? The answer is that they all are correct as far as measurements they can perform within their own tesselation.

Where confusion can arise is under the circumstance when two or more of these beings can overlap their views and rigidly insist that, since nature conforms to structured laws within their tesselation, their particular tesselation governs the overlap. IT DOES NOT. Furthermore, the mathematics of any particular tesselation cannot be used to predict the behavior of nature when the observations are performed in a frameword that extends beyond the boundaries of that tesselation.

The basic problem is that unless the observers recognize the existence of alternative tesselations, they may haggle and argue in endless fashion concerning who is right and who is wrong about descriptions of nature.

This fact has been repeatedly pointed out by this author who has attempted to identify the problem by geometrizing the situation and introducing the concept of alternatives (1)(2)(3). It was asserted that subjective perception uses a higher level of dimensionality than most of our objective mathematics, and this view is now becoming accepted (4).

It is the purpose of this paper to introduce a new interpretation to the fundamental mathematical process known as transforming (or mapping). This will have no effect whatsoever on the equations we now use. But what it will do is introduce an entirely new type of mathematical tool. And, hopefully, bring subjective and objective audio a little closer.

TRANSFORMING PROBABILITY

Consider the following symbolic relationship,

$$\text{signal} = f(x).$$

Just what does that mean?

It means that the signal is expressed in terms of the coordinate x and not in terms of anything else. While in this form, that is, while expressed in terms of the coordinate system x, we can state precisely what value that signal takes for each point of x.

Even if the signal derives from a random process, it will have completely defined values for each value of x. The probability with which values can be assigned for each x is unity.

The probability with which signal values can be assigned to any other coordinate system is zero unless we know the rule by which each coordinate x is changed into coordinates of that other system. While in the system of x's we know things only in terms of x.

All other possible descriptive coordinates do not explicitly appear. We cannot state with any nonzero probability what the signal values are in terms of some y unless we know how y is expressed in terms of x. And that knowledge is imparted only when we have a defined coordinate transformation.

Therefore, the role of coordinate transformation can, from this point of view, be considered to be that of transferring probability of assignment from one set of coordinates to another set of coordinates.

If we change the description of a signal to a new form,

$$\text{signal} = G(y),$$

we have converted to a new coordinate y. We have gone from a description in x to a description in y. This is symbolized in figure 3.

When we transfer from the coordinate system of x to the coordinate system of y, we move from tesselation x to tesselation y. We are looking at the same signal, but from a different perspective. We now see things in terms of y, whereas before we saw things in terms of x. The transform, the map, that took us from the domain of x to the domain of y transferred unit probability of description to y, and leaves us nothing in terms of x.

This mapping is a jump from the domain of x to the domain of y. Probability of exact description is transferred discontinuously from x to y. This is a jump transform between sharp alternatives.

Now I want to introduce a new concept - the blur transform. What happens if instead of jumping discontinuously from one domain to the other we slowly go out of focus in terms of the coordinates of one domain as we come into focus in terms of the coordinates of the other domain. The answer is that the probabilities do not discontinuously jump but smoothly transition between zero and unity.

This can be visualized as in figure 4. We transfer probability of precise description from x to y. But in the process we move through regions of imprecise description. We cannot make any certain statements about the signal in terms of x or y. All we can give is the probability associated with that particular part of the blur transform.

As we move through this fuzzy, out of focus, region we find that we are in a region with higher dimensionality than the domain we started from or the domain toward which we are moving. This is symbolized in figure 5.

From the standpoint of an exact description these are pseudo-dimensions, because the moment we try to form an exact description (that is, unit probability) in terms of any of them, we find ourselves back in one or the other sharp domains and not in the region we wanted to be. This will happen every time we attempt to form descriptions over a domain which overlaps one or more sharp alternatives.

The boundaries of the tesselations which define sharp alternatives are rigidly fixed. They are solid fences which enclose a patch of land with a unit volume of probability of occupancy. The boundaries of fuzzy alternatives also enclose a patch of unit volume, but these boundaries are elastic and may transgress the boundaries of a number of sharp alternatives.

We can imagine this as the situation where each tesselation is a floor tile with unit area and the tiles are densely packed. The signal is a flud of unit volume which is spilled on the tiles. If the fluid lies within the area of one tile, then the descriptions we can form of that signal are sharp descriptions. If the fluid spills over one edge of two or more tiles, then the height of the fluid must be lower than unity in some places and the descriptions we can form of that signal are fuzzy. It is still the same signal, but the descriptors are mathematically imprecise in terms of the exact descriptors of any single tile. This is symbolized in figure 6.

What we have, then, is a situation in which descriptions can, by nature, be imprecise. A perfectly valid description which is formed from boundaries that overlap sharp alternatives will use imprecise descriptors in terms of any of the defined properties within the sharp alternatives.. The term "imprecise" must not be interpreted as anything indicating lack of validity. This term relates to the probabilities which can be associated with a description.

In terms of the framework of allowable sharp tesselations, the alternatives which overlap one or more such tesselations will be called fuzzy alternatives.

The foregoing discussion can be summarized in the following manner, which relates quite obviously to audio: The fact that one can write an exact equation in terms of a particular set of coordinates does not make that coordinate something special. Nor does one's inability to articulate a precise meaning to a descriptor, in terms of any exact coordinate, destroy the validity of that descriptor.

CONTEMPORARY TECHNOLOGY

The point of this abstract mathematical approach can now be presented as it relates to contemporary audio technology. Many technical persons will reject, and often ridicule, the descriptive terminology used by persons attempting to convey subjective impressions of defects which they perceive in sound reproduction. The terminology is rejected for being vague and imprecise and not at all able to be quantified, as are the descriptors of technical analysis.

This paper serves warning that this situation is not as one sided as it is presumed to be. I caution my technical peers that precise descriptors are special limiting cases of a general class of imprecise descriptor. Thus, the technical person must accept the possibility that descriptors of a higher-dimensional process, such as sound perception, may have to be expressed in probabilistic terms when presented in an otherwise precise lower-dimensional alternative.

This does not defend the indiscriminate use of flowery language for expressing esoteric concepts in perception. But it does seem to verify something which many persons have long suspected. Namely, there is a language of subjective terminology and it can convey meaning at an experiential level.

Another byproduct of this concept of imprecise descriptors and fuzzy alternatives is a means of coming to grips with something which technical persons have struggled with for some length of time. Even the most mathematically oriented signal analyst struggles to maintain some semblance of joint time-frequency parameterization. Even those who bristle at the use of the nonsense phrase "RMS watt", try to speak of real-time spectral analysis or of the short term frequency of a tone.

We know such things are mathematically wrong. Yet we use them. Why?

I submit that this is a clear example of an imprecise descriptor used in a fuzzy alternative formed over the union of two sharp alternatives. We use it because it makes sense. We use it because we are subconsciously trying to express a genuine higher-dimensional perception in the language of lower-dimensional alternatives. The technical person is, in this case, doing exactly what the nontechnical person does who uses imprecise descriptors. The only difference is that the technical person seldom realizes why he is forced into mixing otherwise precise terminology.

- 7 -

The technical person who defends the use of a joint time-frequency description (because it makes sense) is also aware that there is a limitation on the ultimate precision with which a joint determination can be made of frequency and time. For formal, historical reasons this is called an uncertainty principle. The more exact one becomes in the determination of frequency, the less exact is the knowledge of time. That is the tradeoff in probability for this particular imprecise descriptor.

I submit that it is the increased pseudo-dimensionality which drives those of us who are technical to use such an imprecise descriptor. Of course it makes sense! Our sensory perceptions support a higher-dimensional frame of reference. We need this in order to begin to align what we measure with what we perceive, and the enhanced pseudo-dimensionality allows us to come closer to our dimensionality of perception than either sharp alternative would allow by itself.

FUTURE DIRECTIONS

It is interesting to contemplate what direction this concept of imprecise descriptors may take us in our quest to develop perception-related audio measurements. Principally, what we have done is introduce probability and likelihood as a technical property deriving from the process of mapping (or changing from one system of coordinates to another).

One means by which it may prove possible to relate subjective and objective properties can be understood by referring to figure 6. We might presume that the choice of descriptors used in subjective terminology relates to some, as yet undefined, perceptual space which has a higher dimensionality than any sharp alternative which we may use for objective measurement. This perceptual space is a fuzzy space of undefined dimensionality.

A lower-dimensional, precise characterization of attributes identified by the higher-dimensional imprecise descriptor will involve geometric shapes over that lower dimensional space. In other words, the lower-dimensional sharp alternative may contain everything of significance which is to be known about the signal, but the properties which the imprecise descriptors refer to are manifest as complicated geometric patterns over that sharp alternative. There may well be a seven-dimensional sharp alternative which effectively characterizes some person's perception. But what that person may do is try to erect a pseudo seven-dimensional fuzzy alternative spread over the parameters of selected lower-dimensional sharp alternatives with which that person is familiar.

- 8 -

The context of the use of an imprecise descriptor must be taken into account before the geometric pattern can be identified in the lower-dimensional space. The imprecise descriptor "grainy" can mean different things (in terms of a time domain representation) depending upon the dimensionality of the fuzzy alternative in which that descriptor conveys experiential meaning.

A contextural boundary needs to be established for the fuzzy alternative and determination must be made of the possible sharp alternatives which may be contained within that boundary. Then an assessment needs to be made of the probability relationships for the sharp alternatives that are involved. This may, in preliminary stages, be nothing better than the observation that one descriptor is "... something like frequency", or that another descriptor is "... more like space than time, except under specific conditions of tonal balance".

With the full expectation that these descriptors relate to a nonlinear process, it is conceptually possible that we could erect a fuzzy alternative over the union of a number of sharp alternatives which we could measure separately. The process might be analogous to the present technique known as multispectral analysis, with the exception that the subspace projections which sharp alternatives present on this higher-dimensional structure will involve likelihood estimators.

There are a few guidelines which need to be recognized in this process. When transferring from one descriptive space to an alternative descriptive space, we cannot assume that any axiomatic geometric relationship will be preserved. Concepts such as "larger than", "smoothly changing", "in between", and "at one place" will not necessarily hold when we convert from one frame of reference to another. Something that is perfectly well behaved and continuous in terms of one frame of reference can involve bizarre relationships in terms of some other frame of reference.

A thing that occupies a well defined "place" in one frame of reference may have no definable "place" in another frame of reference.

There is nothing strange in this; it only indicates the difference in the manner of describing the same thing - how that thing appears when viewed from different frames of reference. The concept of location, or place, or point, therefore, is not an absolute, but is relative to the frame of reference used for description.

And finally, it is entirely feasible that when we convert from one alternative to another, particularly when those alternatives are of differing dimensionality, quantum conditions may prevail. Properties

which are continuous and well behaved in one sharp alternative may map into discrete behavior in terms of the coordinates of another sharp alternative.

All of these guidelines are the consequence of changing the frame of reference and any further discussion of them would go far beyond the intent of this paper.

As a result of this process, and taking into account all of the guidelines, if we have carefully erected a reasonably correct fuzzy alternative that relates subjectively determined imprecise descriptors to sets of objective measurements, we may begin to say that we are better able to measure what we perceive. This means that instead of one sharp alternative measurement, such as waveshape accuracy in the time domain, we will use several complementary sharp alternative measurements combined in probabilistic fashion and involving geometric patterns over each of the sharp alternative coordinate systems.

In other words, there may never be a machine which produces a meter deflection proportional to subjective quality. Instead we can expect that analysis of selected technical measurements, taken as a whole, can yield the probability of acceptance under specific conditions.

REFERENCES

(1) R. C. Heyser, "Perspectives in Audio Analysis: Changing the Frame of Reference, Parts I and II," J. Aud. Eng. Soc., (Communications), Vol 24, pp. 660-667 (Oct. 1976); pp. 742-751 (Nov. 1976).

(2) R. C. Heyser, "The Delay Plane - Objective Analysis of Subjective Properties, Parts I and II," J. Aud. Eng. Soc., Vol 21, pp. 690-701 (Nov. 1973): pp. 786-791 (Dec. 1973).

(3) R. C. Heyser, "Geometrical Considerations of Subjective Audio," J. Aud. Eng. Soc., Vol 22, pp. 674-682 (Nov. 1974)

(4) M. A. Gerzon, "Mathematics and Sound Perception," J. Aud. Eng. Soc., (Communications), Vol 26, pp. 46-50 (January 1978).

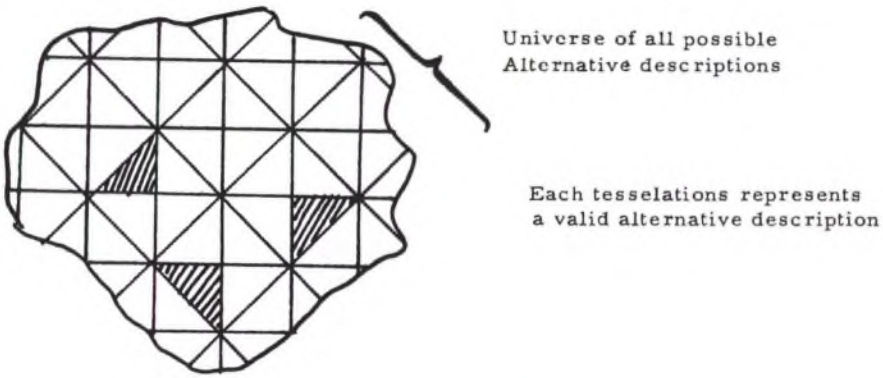


Fig. 1. Alternatives as tesselations over the universe of descriptions

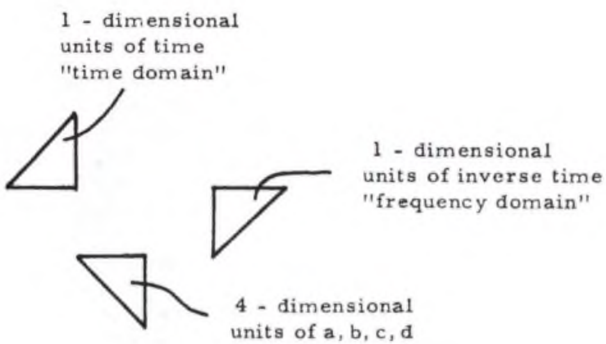


Fig. 2. Selected tesselations

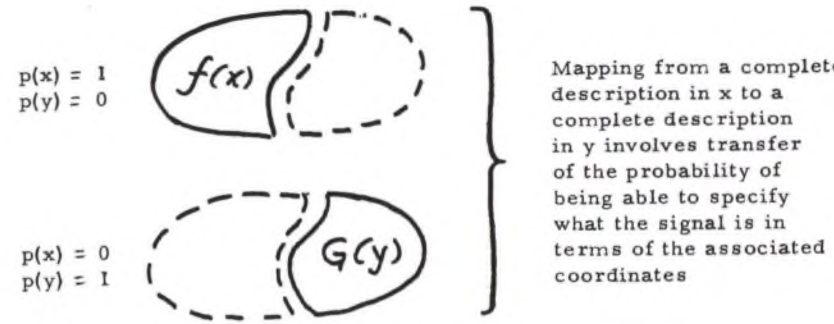


Fig. 3. Illustrating a jump transform.

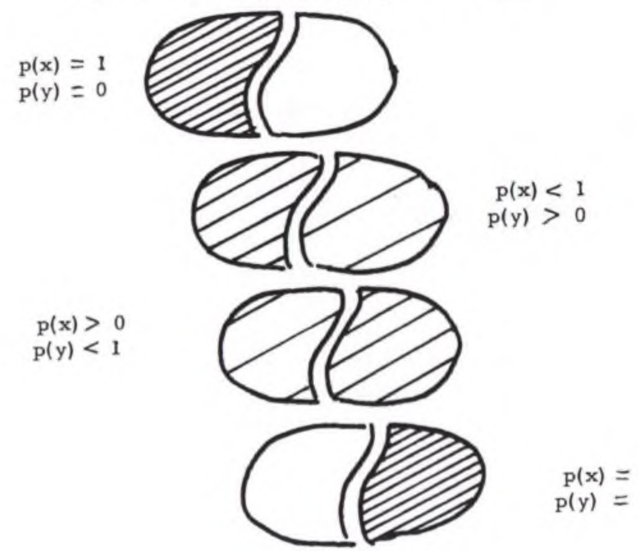


Fig. 4. Illustrating a blur transform.

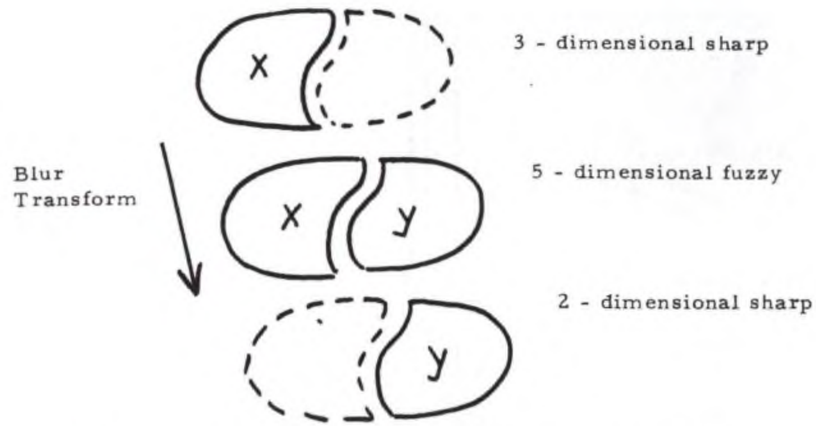


Fig. 5. Illustrating pseudo-dimensionality encountered during a blur transform.

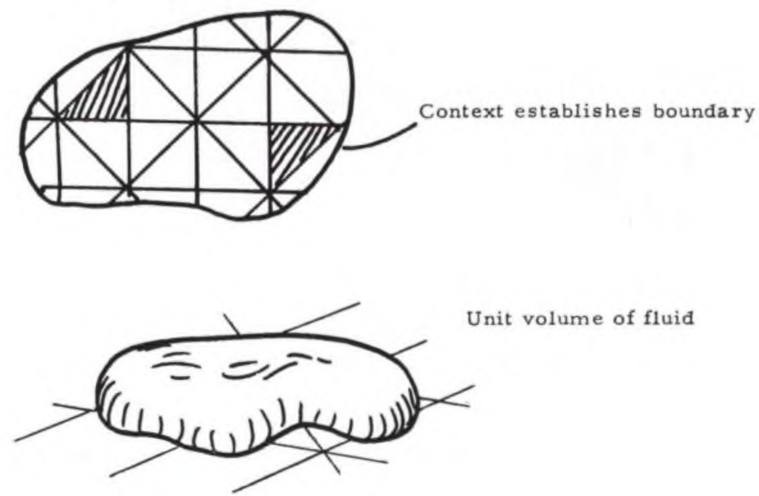
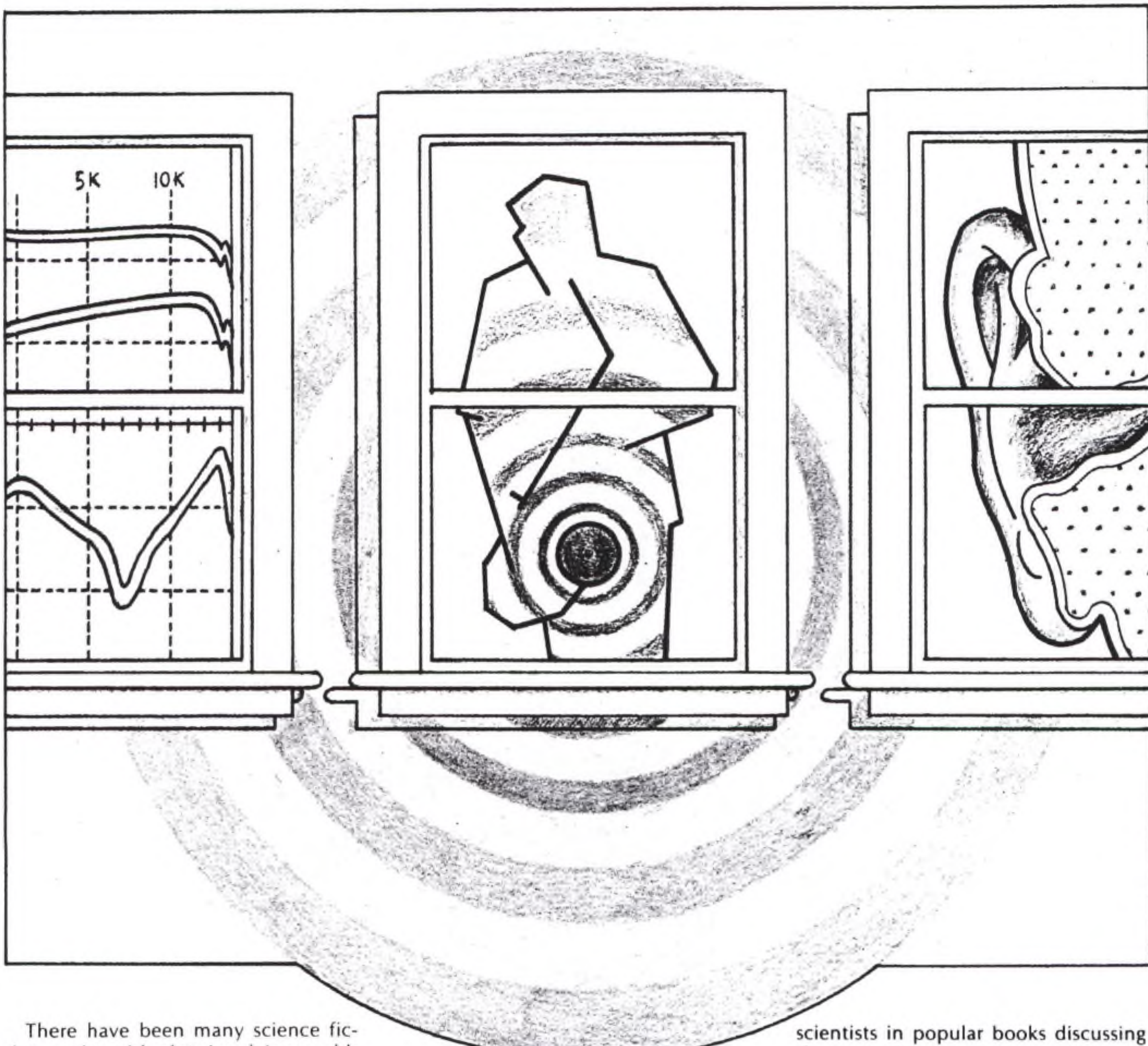


Fig. 6. Fuzzy alternative as a higher-dimensional space formed over a number of sharp spaces likened to fluid spilled on floor tiles.



A VIEW THROUGH DIFFERENT WINDOWS

RICHARD C. HEYSER

There have been many science fiction stories with plots involving worlds of differing dimensionality. My favorite among these is a perennial science-fiction classic called "Flatland, A Romance of Many Dimensions." Written in 1884 by E.A. Abbott, a schoolmaster, it is still available, in its seventh edition in paperback (1). What does that have to do with audio? Well, if you know the story, and accept some of the concepts I am about to present, you might come to agree with me that we now live in an audio flatland.

Written more than 20 years before Einstein's first paper on relativity, "Flatland" is pure fantasy and makes no pretense of application to human affairs. Yet in an abstract sense, the subject matter of Abbott's book anticipated some of the conceptual problems which three-dimensional humans might encounter in dealing with the four-dimensional universe of relativity. This has been pointed out by several scientists in popular books discussing the space-time continuum. (2, 3). The central difficulty is in demonstrating to people, whose total conceptual structure is geared to one level of dimensionality, that higher dimensionality exists and is capable of being understood.

One Dimension, Two Dimension, Three Dimension, Four...

The allegories are numerous, and Abbott's "Flatland" is one of the better known of these. . .I will not spoil the story by telling the plot. It is a delightful little book and, like its subject matter, can be read at several levels of conceptualization.

But I will impose on you the following mind-expanding thought. We can quite easily grasp the idea of a two-dimensional subworld that is imbedded in our three-dimensional world. In mathematical language, a subspace is an easy thing to imagine because it is usually generated by denying one or more of a higher dimensional set of attributes to the lower dimensional space.

As an example, this piece of paper on which these words are printed can be thought of as a two-dimensional subspace. Left-right and up-down are the dimensions of the printed page, but in-out is not.

The concept of in-out is denied to any being who "lived" in the two-dimensional world of this page. The third dimension, which we call height, is denied this being because he cannot stay completely within the confines of his space and touch all parts of our three-dimensional space, *so long as his is a subspace*. That is easy to understand.

But suppose his was not a subspace. Suppose his space was our space, but the difference was that he *saw* it as two-dimensional while we saw it as three-dimensional. In order to visualize, as a three-dimensional person, how this might come about, suppose the sheet of paper was enormously large and we began folding it back and forth to compress the paper to some reasonable volume in our space.

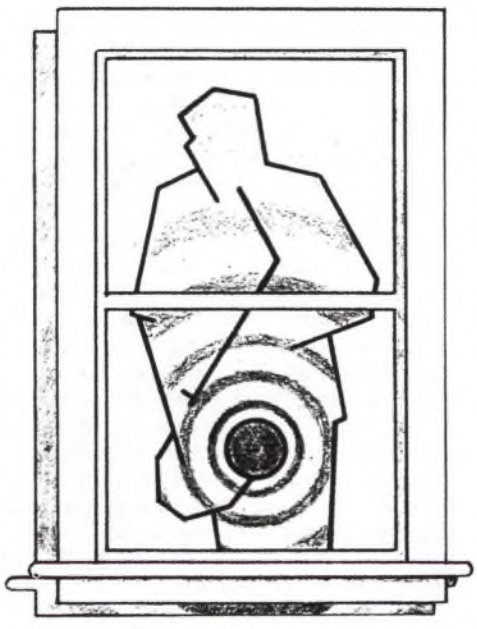
The math is sloppy, but it is the thought I want to get across. If all we did was fold and compress the paper, we would never alter the fact that it was a subspace, but we can begin to see that the property we call height would in fact begin appearing as an attribute within the two-dimensional subspace. Depending on the way we folded the paper, things that moved up, as we saw them, might appear to shift in lateral position in our two-dimensional friend's point of view.

The being who lived on the paper would now see the third dimension. But he would not recognize it as a dimension unless we told him what it was. He would, instead, observe it as some weird property joining relationships in his two-dimensional existence.

Now, let's do it again. Only this time we "superior" three-dimensional beings see the plight of a one-dimensional being. He lives on a string and does not even know what up-down means. All he knows is what we call left-right.

But his string world is so very long that we begin rolling it up like a ball of twine to concentrate it in a small region of our space for observation. Finally, even though this one-dimensional space has no width, when it is so crammed together that no part of our three-dimensional space is farther than some previously agreed upon small distance from some part of the string, we can agree that he "touches" everything we do. He actually sees three dimensions. But he does not observe the properties as a thing he might call "dimension." Instead he sees it as certain relationships among his one-dimensional observations. What to us is a vertical line might appear to him as an array of disjointed coordinate locations.

Now the mind expanding part . . . the part that is new. Two beings can observe and describe everything that happens, but do so from the viewpoints of different dimensionality.



I do not mean to imply that there are four-dimensional beings or two-dimensional beings among us. The point I wish to convey is that there is no preferred frame of reference for any observation, either for number of dimensions or units of measurement.

There is an additional consideration to this geometric concept of Alternatives, one that has far reaching consequences. Although the folding of the two-dimensional paper and twining of the one-dimensional string were allegorical, they do illustrate that the concepts of continuity and "betweenness" do not have to be preserved when we change dimensionality.

A trajectory of smoothly continuous values in a higher dimensional alternative, for example, might show up in a lower dimensional alternative as a discontinuous set which may disappear at certain places and reappear elsewhere without being found at intervening locations.

A being who is accustomed to sensations perceived in a particular dimensionality and frame of reference might form certain concepts about that situation which "make sense." Any attempt to convey to such a being the possibility of a higher dimensional version of his universe will be a most difficult chore because it cannot make sense if one tries to explain it in terms of the coordinates of that lower dimensional space.

This was the conceptual problem in Abbott's "Flatland," except that now we are not talking about a lower dimensional subspace imbedded in a higher dimensional space. We are confronting the problem of a lower dimensional Alternative to a higher dimensional space. Things can disappear from a subspace and not be found anywhere in it, but that is not the case with Alternatives. Things do not disappear in Alternatives, they show up as other geometric configurations.

The point I wish to convey is that if we discover some seemingly bizarre behavior that does not seem to make sense in our otherwise orderly view, we might be able to reconcile such behavior by converting to an alternative system of coordinates — possibly at a different level of dimensionality.

These are new ideas and take getting used to. And yes, by George, they do have application to audio and subjective perception. But allow me to continue with some of the fundamental concepts before swinging into sound.

Windows

Suppose we have a good technical description of something. We have a mathematical description nailed down. There are no hidden variables. Our description will involve certain cause and effect relationships among parameters. If we now set up a physical observation in those parameters and if we have not left anything out, then nature will oblige us by operating in consonance with those parameters. This does not mean that nature prefers these parameters. Nature does not give a darn what parameters we choose. All it means is that we were consistent in setting up a model.

Suppose that we wish to take another view using other parameters. We wish to see the world through a different window. How many windows are available? As many as we care to find. That is what I call the Principle of Alternatives.

We already discussed two of the windows in previous articles. One window is measured in units called time, while the other window is measured in units called frequency. It makes no difference whether we look through the time window or the frequency window, we see all there is to see of the same thing. Only the way it appears is different.

A Five-Space

Now let's see what relevance this *technofreak* talk has in sound reproduction.

Did you ever hook an oscilloscope onto the output terminals of a power amplifier and watch the voltage waveform while listening to the sound coming out of the loudspeaker connected to the same terminals? Or hook a 'scope to the output of a microphone while you listened? What you see on the oscilloscope is a representation of a one-dimensional signal. It is volts as a function of time. But when you listened to that signal, what did you perceive in the way of dimensions? More than one, I will wager, if you aligned that perception with prior experience of the way things sound.

Sound has a "where." That is, it is located in physical space with respect to us. That is at least three dimensions right there. Sound also has a property which I will simply call "tone." Pitch, timbre, and the things we measure in units of pitch are expressions of tone. "Where" a sound source originates and what "tone" that source has are independent properties. Whether a clarinet is stage left, stage center, or stage right does not dictate what musical notes are going to be playable on that clarinet. So "tone" is somehow independent of "where" and can rank as *at least* an additional dimension.

Then there is "when" a sound occurs. Think about it a bit. That is another possible dimension. Then, there is a "how much," or intensity, which is not a property precisely dependent on the other things which I call dimension.

All in all, if we add up those properties generally agreed to be independent attributes, we find that the *least* number of dimensions we can get away with in a *subjective* description of sound is *five*.

And where is all (or most) of that higher dimensional information in that silly one-dimensional waveform we view on an oscilloscope? It is there. But just as the three dimensions of space viewed from a one-dimensional ball of twine, the higher dimensionality of perception is encoded as the relationships existing between properties in the one-dimensional waveform.

Yes, yes, I know about two channels for stereo, four for quadraphonic, and all that. But right now we are on the ground floor and trying to tie certain properties of subjective perception with other properties we now measure in objective analysis. I imagine many of us at one time or another have had the experience of finding that due to a technical error we had been listening to a two-channel mono feed when we thought we were hearing stereo. And like the optical illusion which, once recognized, never looks the same again, we find the subjective dimensionality collapsed to an "obvious" mono program when the deception is discovered. But we had been fooled . . . that one-dimensional program had supported a whole stereo illusion.

Back To Flatland

The audio technologist who measures things in the frequency domain resides in a linear world of one dimension. He is a Flatlander. There is nothing wrong with that. So long as the device is linear, or essentially linear, the audio Flatlander sees everything there is to see. His window happens to look out onto a one-dimensional universe.

The prime audio problem ("how can we measure what we hear?") arises when this Flatlander tries to convey measures of fidelity to the being who sees things through a higher dimensional window. Neither one sees something the other does not.

But that which appears essentially perfect to a Flatlander, may or may not be essentially perfect when viewed in a higher dimension. The reason for this, as we have discussed, lies in the fact that if these are genuine Alternatives (different ways of describing the same thing), then the map between them is a geometric transformation. Distributions are mapped into distributions. A simple, elemental place in one Alternative will appear as some geometric distribution (or configuration) in the other Alternative.

As an aside, I must point out that the concept of "place" is perhaps better understood in terms of this idea of geometric distribution, or configuration, or figure. The concept of "point," or the concept of "line," or of any special type of "figure" is not fundamental to the establishment of a valid geometry. That is a very difficult thing to recognize, accustomed as we are to the strong arm methods of teaching mathematics to generations who couldn't care less about mathematics. Only recently have we begun to explore distributions as a Theory of Generalized Functions, in order to bring light to a badly illuminated part of our understanding. The Dirac delta is the prime example of a distribution that can begin giving meaning to the "place" where something can be found. Unfortunately, some popular discussions about the so-called delta "function" as applied to audio have fallen into the trap of trying to explain it in heuristic terms, such as . . . "existing only at a point but having unit area." Such an explanation is no explanation at all, because generalized functions cannot be assigned values at isolated points.

The audio Flatlander, viewing the frequency response of a loudspeaker, amplifier, cartridge, or whatever, cannot possibly make "sense" out of the protestations of a higher dimensional being that the Flatland ranking of distortion does not always correlate with that being's ranking of distortion. It does not make sense to the Flatlander as long as he uses his own coordinate system.

Unless it is disclosed to him, the Flatlander has no way of knowing that higher dimensional "places" show up as geometric properties in Flatland. The converse is also true; each place in Flatland may appear as some spread of values in the higher dimensionality.

The problem becomes enormously compounded when either the Flatlander or the higher dimensional being set up test figures to check for the possibility of geometric warping. The test figure is usually chosen to be that which is easiest to measure *within a particular frame of reference.* The test figure which a Flatlander might choose does not necessarily correspond to a test figure which might be chosen for any other dimensionality.

So our audio Flatlander might set up a test figure which represents a perfect concentration of a unit volume of material at a well-defined "place." We might call it a delta function corresponding to unit energy at one frequency. (Another audio Flatlander, living in the one-dimensional time domain, would perceive that particular test figure as a "wave" extending over the whole of the time domain. He would call it a sine wave.)

The frequency domain Flatlander then checks for geometric warping by determining how much material appears at other "places" when the test figure is processed. The Flatlander might set up a standard of warping, such as: If no more than one thousandth of one percent of the volume of the test figure can be found at any place other than the original location, the geometric warping will be essentially nil. It would seem logical to presume that if one found two percent of the volume out of place, the warping would be greater than if one only found two-hundredths percent out of place. So the audio Flatlander can go about checking for geometric warping by placing test figures at various locations in his space.

But what might a higher dimensional being perceive? First of all, the Flatlander's test figure might have no particular significance in the view through the higher dimensional window. How might you feel if you looked out of your living room window and saw colored lights flashing across the sky like a giant aurora borealis display? If you asked what was going on, you would get a reply that a one-dimensional being was testing the universe for linearity. "See there," he would say, "that's a perfect signal at coordinate location 47." And you would see a steady green glow with ripples of red slowly moving across the sky.

The Flatlander's test figures are things that he can understand. In the Flatlander's world, a test figure corresponding to something of significance to a higher dimensional being might appear hopelessly complicated. If, in looking out of your window, you had said, "Hey, knock off the silly lights; if you want to check for geometric warping, use this meter rod." And you hold the meter rod up. After a brief pause the Flatlander would reply, "You're nuts, all I see is a blurring of edges and a purple glow."

Silly example? Perhaps, but the math, or more properly the geometry, is reasonably illustrated by such an example. One-dimension and five-dimension beings cannot agree on the subtler aspects of scene distortion because each sees the view through different windows.

Views

If you remember our discussion of the end product of audio, the query "how do we measure what we hear?" translates to "how do we measure the illusion we perceive?" This brief discussion has been our deepest penetration so far into the abstract. But the problem it addresses is of the utmost practicality. Perhaps we cannot measure the illusion of sound, but we might be able one day to grasp some of its structural properties, as a perceived higher dimensional experience.

In this case I am attempting to convey a reason why conventional distortion measurements, such as harmonic and intermodulation, need not necessarily correlate with our subjective impressions of distortion. Geometric warping of a perceived illusion and geometric warping of a lower dimensional test signal are both distortions. But they are distortions expressed in terms of the framework of the reference system against which they are measured. If those frames of reference are not the same, whether they differ in dimensionality or some other way, we cannot automatically rank them as equivalent. It may happen in a gross measurements that *three* is greater than *two,* and *two* is greater than *one* for certain types of distortion as expressed in either reference system. But somewhere along the line, we are going to get into difficulty quantifying subtler forms of distortion with such gross equivalents.

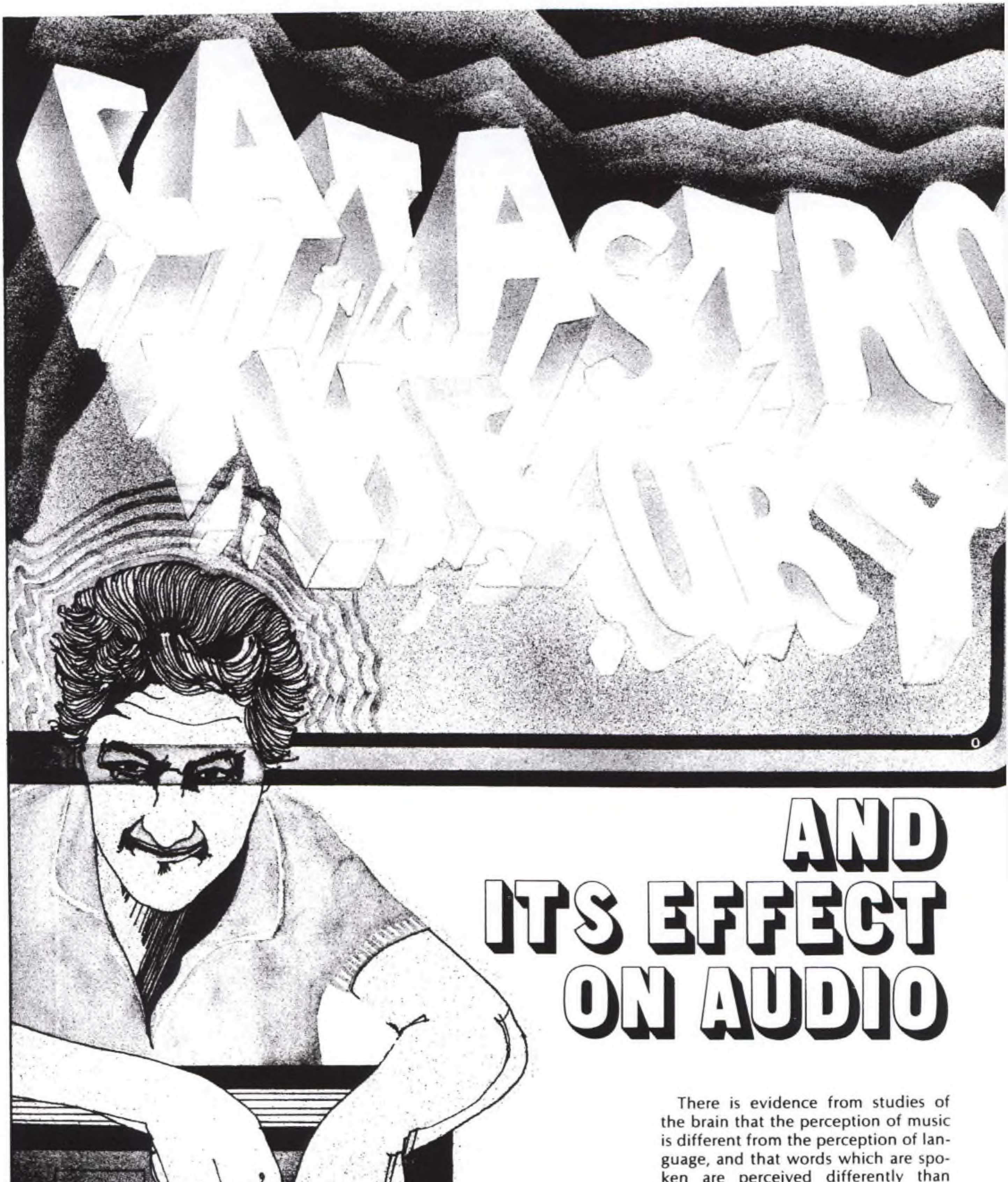
But whether we are talking about a distortionless situation or one that is badly distorted, the deeper geometric property I want you to consider is that of the possible dimensionality which is involved in any particular situation. We have only begun to touch on this subject and I hope to expand on it in later discussions, but the next time you hear an argument between a technologist and a golden ear about the audibility of certain types of distortion, think of this: Is the technologist really a Flatlander, and is it possible they do not agree because they each have a view through a different window? A

References

1. E. A. Abbott, "Flatland, A Romance of Many Dimensions," Dover Publications, New York, 1952.

2. A. Eddington, "Space, Time, and Gravitation," Cambridge at the University Press, 1968 (p.57).

3. C. Lanczos, "Space Through the Ages," Academic Press, New York, 1970 (p.88).



AND ITS EFFECT ON AUDIO

There is evidence from studies of the brain that the perception of music is different from the perception of language, and that words which are spoken are perceived differently than words which are sung. There is also evidence that the way in which we perceive certain natural sounds, whether as music or language, may be related to cultural differences and learning experience. If these, and many other such things in our perception of sound, be true, then where in our audio technology do we address such factors? If not, then why not?

In an earlier discussion I broached the issue of the end product of audio. The end product is the listening experience. The end product is not advertising specs, it is not meter readings, or wiggles on an oscilloscope, or piles of charts and graphs. The end product is that very private and personal experience we have when listening to reproduced sound.

If we are ever going to put a number on the quality of that experience, then it is clear that we must do more than specify the cosmetic perfection of a waveform or pursue an endless quest of reducing measurable distortions on laboratory signals which may have little bearing on the process of perception of sound.

Somehow in our technical considerations of audio we must also recognize the role played by human emotion. Aggression, paradox, strength of opinion, and conflict of interest may not be considered as control variables by an audio designer, but they can be very important in determining the success of the product which he designs.

Under The Lamppost

Achieving a satisfactory illusion of reality is a goal of present audio technology. But understanding how to achieve this goal demands that we consider a great deal more thdn some of us may be willing to admit. I know from firsthand experience the feelings of frustration that can result when an audio component measures well but sounds bad. Fancy mathematics and precision test equipment tend to lose their charm when they disagree with our ears.

The problem lies not in our science but in ourselves. We misapply the science. Nowhere is this more evident than in the way in which we use concepts drawn from linear mathematics to develop models of distortion. Today's best distortion measurements are based on linear theory. Need we wonder why it does not always work?

A suitable parable for this misapplied attitude can be found in an old theater routine. The curtain opens to show a man, obviously well in his cups, muttering incoherently and scrambling around on his hands and knees under a lamppost. Soon a policeman strolls by, looks at the drunk's frantic motions, and asks the cause of his actions. The drunk replies that he dropped his last coin and must find it since it was his carfare home. So the kindly peace officer gets down on his own hands and knees and commences to help look for the coin under the lamppost. After a while, the policeman asks if the man is sure he dropped the coin in this place. "No," the drunk replies, "I dropped it over there," pointing to the dark bushes away from the post. "What," the policeman bellows, "are you doing looking for it here?" "The light is better here," answers the drunk.

That corny old stage routine is a good analogy for many of the things we tend to do in the name of science. We solve the linear math problem because the light is better there. And, just like the man in his cups who can justify his deeds because he can see what he is doing, we can produce dazzling mathematical chicken tracks to defend our choice of mathematics.

But it is not the mathematics that is wrong; the folly lies in its misapplication. Nor can we excuse our use of linear concepts by assuming that a so-called "piecewise linear" approximation will always work for nonlinear situations. It *won't* always work. And where it fails the worst is in the most interesting type of distortion of all distortion giving rise to instability of form.

Instability is a most important consideration in audio, particularly as it relates to our perception. When an amplifier is unstable, we consider it a bad amplifier. Our perception of sound quality can also be unstable, but unstable perception is not considered bad perception. We accept that a sound system can sound great for some type of material and terrible for others. Our opinion of a sound system can change dramatically, but we seldom think there might be some underlying pattern of behavior which, if better understood, could help us understand *why* it sounded good one minute and bad the next. Nor could our preoccupation with linear lamppost math ever lead us to believe that there could be a branch of mathematics which could be applied to problems of perception *and* to problems in equipment design.

All of this is a lead into the subject I would like to present for your consideration.

Jumping

Let us think in terms of what I shall call *factors* and *response*. Factors control response. As the controlling factors of a process smoothly change we may find that the response to those factors suddenly changes. There is a jump in response, and there may even be a jump to a new type of response. Jumping is a property that shows up for certain types of distortion and nonlinearity, and is not something handled by our present linear audio math, no matter how impressive the pedigree of that math may seem. Jumping is a manifestation of instability of form.

The split which we get when we jump from one response to another is called *divergence*. A relaxation of the controlling factors back to the values they had before the jump will not generally produce a backward jump in response. Usually the factors must be substantially reduced before the backward response jump occurs. This means that if there is a jump — if there is an instability in the nonlinear process — three additional properties will appear. These are the properties of *hysteresis, bi-modality,* and *inaccessibility.*

Hysteresis is the name given to the lag in response under cyclic changes in factors which control that response. In the hysteresis region between jumps the response has an "either-or" nature — it is *bi-modal.*

The response jumps from one state to another state and there is no possibility of finding a response between these two end states — the region is *inaccessible.*

These properties of *jump, divergence, bi-modality, hysteresis,* and *inaccessibility* are interrelated such that the appearance of one of them usually means that the others are around.

Nor are these the only properties of a nonlinear process which are not revealed when we attempt to use linear math (because the light is better there). *Irreversibility* is one such property. Once we jump, or cross some decision threshold, it may never be possible to get back to the original response, no matter how the factors change.

Splitting is another nonlinear property. Splitting is an ambivalence in response under certain combinations of conflicting factors. It is a coexistent response state.

Indeterminacy is a property like splitting, but more diffuse in nature. Not an "either-or" state, indeterminacy could be characterized as a "maybe" state. Indeterminacy is an amorphous response.

Catastrophe Theory

Evolution and change. Genesis. Structural stability with preservation of form, then sudden catstrophic change. I doubt if there is any aspect of human endeavor that does not involve the development or unfolding of circumstance. We develop rules and expectations concerning the outcome of an evolving process; and then, suddenly, there may be a surprise. Perhaps the surprise is a part of a grander set of rules which we had not anticipated, or possibly it is a sudden switch to a new set of rules — like a train being shunted onto a different set of tracks.

The concept of stability has concerned mathematicians for a long time. But it was not until the 1960s when the brilliant mathematician, Rene Thom, perceived that sudden changes in form — catastrophes — could be classified in a finite number of ways. Up until that time there seemed to be no way to put a handle on the problem. Thom was able to show that a stable unfolding of a process near a point where change can occur, can have any change that does occur categorized as one of a few basic types. Thom called these *elementary catastrophes.*

Even in its elementary form, Catastrophe Theory stunned applied mathematics. Suddenly (a catastrophe in its own right) a nonlinear theory was available which correctly modelled processes evolving in the four dimensions of space-time. Whereas a truly original mathematical concept may remain hidden for decades until a need is found for it, Catastrophe Theory was an instant success and has been applied to disciplines as diverse as biology, economics, human behavior, and mechanical structures.

In fact the fantastic success of Catastrophe Theory has almost killed it. In a manner well known in high fidelity circles, a new idea can be picked up by overzealous proponents and trumpeted as the final breakthrough of breakthroughs and grandly applied to everything from toenails to tweeters. The voice of Thom has almost been drowned out by those who would take parts of this still-evolving theory and apply it indiscriminately; then reject the whole thing when it may not seem to work in certain applications.

Catastrophe Theory can, with caution, be applied to certain fundamental problems of audio and our perception of audio quality. Being a genuine mathematics of nonlinear processes, we can expect to apply it not only to physical equipment but also to our perception. But PLEASE. Catastrophe Theory is only one of several evolving mathematical concepts. While it can explain certain things that we know happen in audio, but which seem to make no sense in terms of our linear math, and it can do this in stunningly simple fashion, Catastrophe Theory is not going to be the end-all for understanding audio. Let us not make it an ad copy "zip phrase." We have quite enough of that nonsense going on now without making it worse.

In this brief discussion I want to introduce the concept of Catastrophe Theory to audio. All I can present in the short space available for such a discussion is a simple, almost naive, look at how it can be applied. My intent, as always, is to stimulate thought. In what follows I will attempt to explain the mathematical basis in terms which I hope will be understandable.

Factors — Response

Factors control response. In discussing the nature of a response (also called a behavior or reaction) we want to know those conditions under which the response has stabilized when the controlling factors are steady. We want to try to understand the response that has no tendency to drift when the controlling factors are held constant. This will occur when the response is in those stable locations in which there is no attraction capable of pulling it away. Translated into mathematical language, the behavior pattern experiences no gradient in response when each possible control factor is held constant.

The behavior lies at a stationary point (either a minimum or point of inflection) in some sort of a response potential. The potential can be expressed as an equation in which the response is the independent variable and the control factors are coefficients. If, for example, there are two controlling factors and one type of response, the potential will be an equation in one variable with two coefficients. The condition that the gradient of this potential be zero means that the slope of this equation, with respect to the response, be set equal to zero. The set of relationships such that the gradient of the response potential is zero will define a special type of topological surface called a *manifold.* This manifold defines the location of all possible stationary responses.

For two factors and one response we have a three-dimensional behavior space. The manifold will be a two-dimensional surface that folds and curves through this three-dimensional behavior space.

While this may sound highly complicated, it is actually a reasonable way of conceptualizing the interaction of response and control factors. Normally, the math would stop here because we might think that there are a hopelessly large number of manifolds which could correspond to all possible situations. But Rene Thom proved a brilliant theorem that broke this problem wide open. He proved that the only possible potentials were derived from a universal unfolding of a finite number of forms. These forms are what the mathematicians call "germs of singularities." It wasn't an infinite number of surfaces after all, but a very few; and these were of known type. The concept of germ and of unfolding is much too complicated to go into in this discussion, but the results can certainly be appreciated.

If the number of control factors is five or less, then there are only a few types of potential which will determine the response to those factors. The dimension of the response manifold, derived from these elementary potentials, is always equal to the number of control factors. The way in which the response will be observed in terms of the set of control factors is a special type of mathematical map from the manifold's surface onto the space of control factors. This map is called a *projection*. In naive terms we could think of the projection as the *shadow* which the manifold response curve casts on the control space. This map, induced by the projection on the control factor space, is called the *catastrophe map* of the behavior potential.

Thom proved that any singularity (wild change) in the catastrophe map is equivalent to one of a finite number of types which he called *elementary catastrophes*. In any dynamic system there are a precise number of topologically distinct discontinuities which can occur. The number of elementary catastrophes depends *only on the number of control factors*, when there are five or less. For five factors, there are 11 types of elementary catastrophe. For four factors, there are seven types; for three factors, there are five types; for two factors, there are two types; and for one factor there is only one type of elementary catastrophe which can occur. For six and above, there are an infinite number of catastrophes.

If we stop, for a moment, to think what this might mean in audio terms it gets pretty interesting. Do we like the sound of a certain loudspeaker or don't we like the sound; that, of course, is a response. What are the conflicting factors which might be involved in creating that response?

Here, you can put in your own set, but suppose there are only two conflicting factors: how much we listen to live music, and how much we listen to music reproduced from this loudspeaker. The solution of this audio problem will involve the *Cusp Catastrophe*.

Cusp Catastrophe

If that is the ball game — two factors and one response — then there are only two kinds of elementary behavior we can expect. The names given to these are the *fold catastrophe* and the *cusp catastrophe*. Of the two, the cusp catastrophe is the more interesting from the standpoint of the behavior pattern which it predicts. In order to understand how it can be applied to this audio problem, it is necessary to continue a bit more with the basic math discussion.

Let us take the case of two factors and one response and show how the cusp catastrophe develops. The potential for this particular case is the universal unfolding of a germ which is the response coordinate raised to the fourth power (that fact is not obvious

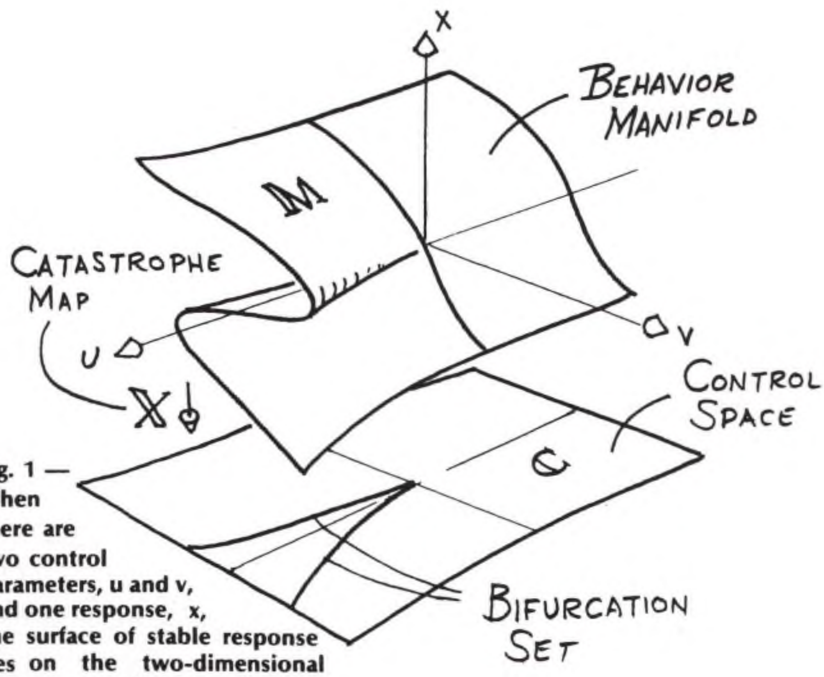


Fig. 1 — When there are two control parameters, u and v, and one response, x, the surface of stable response lies on the two-dimensional behavior manifold, M.

nor does it follow from anything in our discussion, but is included for completeness). This potential has the form of a fourth-order equation in response with two parameters. The parameters are the coordinates of the control factors. The equation of this potential, which we shall call P, is:

$$P = (1/4)x^4 + (1/2)\,u\,x^2 + v\,x$$

where x is the response coordinate and u and v are the control factors giving rise to the response.

Catastrophe Map . . . shrunk down to stylized chicken tracks pregnant with meaning.

The response will be stationary for those values of x, u, and v where the derivative of P with respect to x is zero. The response, in that case, will be at a stable point of the behavior potential. This occurs when:

$$x^3 + u\,x + v = 0$$

The two-dimensional manifold, which we will symbolize by the letter M, is then that surface in the three-dimensional x, u, and v space which is described by the equation we have just developed.

If we put this manifold in the three-dimensional x, u, and v space we will get the folded surface shown in Fig. 1. This is perhaps the most widely publicized example used to describe Catastrophe Theory. The reason is because, being a two-dimensional surface in a three-dimensional space, it can be sketched and its geometry readily grasped. Like everyone else, I am guilty of showing this because it is both easy to draw and understand. There is no convenient way of sketching a three-dimensional swallowtail catastrophe in a four-dimensional space, or any of the other higher-dimensional geometric catastrophe sheets.

In Fig. 1 the first thing we note about the manifold M is that for values of control parameter u above a certain level, the sheet becomes folded. The projection of this fold onto the two-dimensional control space, shown as the surface labelled C, is a sharp pointed curve which forms a *cusp*. The projection of values found on M onto the plane C is the *catastrophe map* of the behavior potential. The catastrophe map is symbolized here by the capital X. Capital X, the catastrophe map, denotes the operation of dropping perpendicular projections of what is happening on M onto the plane C.

All highly symbolic, and, in typical math fashion, is shrunk down to stylized chicken tracks pregnant with meaning. But don't get hung up on the symbols or the fancy names. Think of the actions which give rise to those things. The surface M is the hypothetical manifestation of the position of unchanging response, x, under controlling factors u and v. We, who attempt to observe the response in terms of controlling factors, cannot see the surface M. All we can observe is what happens in terms of those controlling factors. We must observe what happens on the surface C. We see the projection of M onto C. We see the shadow of the bird flying overhead, but not the bird itself.

The projection of the fold onto C forms two intersecting segments called the *bifurcation set*. The term bifurcation relates to two-forked or dual transition behavior. The cusp lines show the thresholds for sudden and catastrophic response changes which occur under action of the controlling factors. This is what gives rise to the name *cusp catastrophe* for the kind of behavior change we will observe.

When we change the control factors and a response is induced . . .

When we change the control factors and a response is induced, this is mathematically equivalent to our moving from one place to another *on the manifold* M. But because the point on M is controlled by changes in coordinates u and v, the point must jump whenever the increment in control passes a cusp boundary corresponding to the passage of a fold. This can be visualized by referring to the simple sketch of Fig. 2. The trajectory of response induced by a certain change in control factors is shown as curve R. The place on surface M where the response occurs is found by projecting a perpendicular, upward from the corresponding u and v coordinate location, to the place where it intersects the surface M. When the control locus passes the bifurcation line at point (a), the point projected on M must jump from the lower sheet to the upper sheet as shown in this sketch. Looking at the three-dimensional geometry, it is obvious what happens: In order to remain on the stable surface M, the response point must jump the gap whenever the control factors go past the fold. But, looking at the result only in terms of the control plane C, we would see a seemingly bizarre behavior: The response was continuous and well-behaved as the factors were changed, then all of a sudden without warning

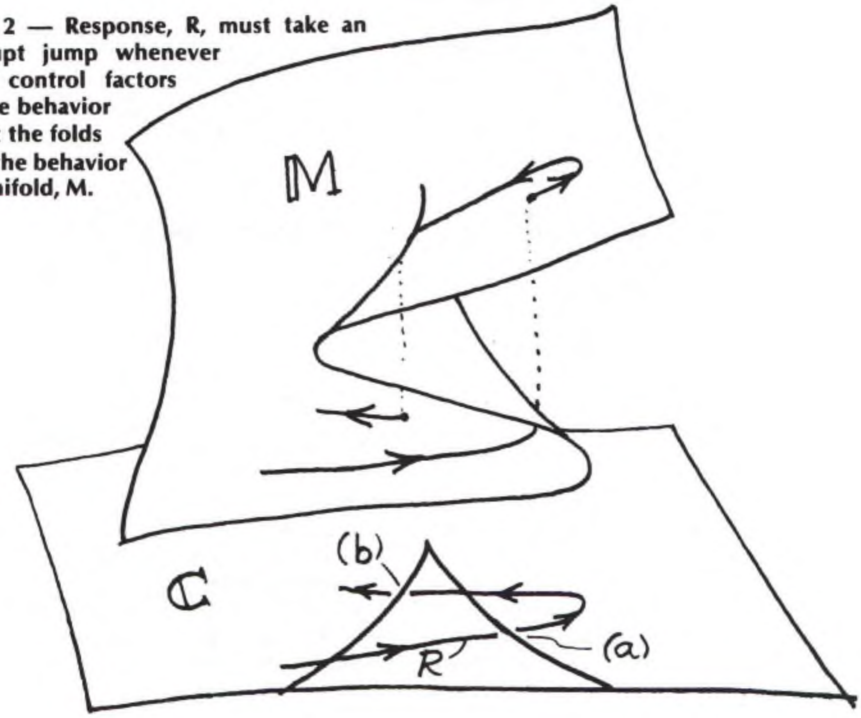


Fig. 2 — Response, R, must take an abrupt jump whenever the control factors drive behavior past the folds in the behavior manifold, M.

the response dramatically jumped to a new behavior.

If we try to restore the original response by relaxing the control factors back to the value they had before the jump, we would not see the response come back to its former value. Instead, we would have to continue reducing the control factors back to the place where they cross the bifurcation line at (b). Then all of a sudden the response pops back to its former value. Of course, what is happening is that the trajectory on M comes back over the upper sheet until it falls over the edge of the fold, and then it is back on the lower sheet.

Flatland

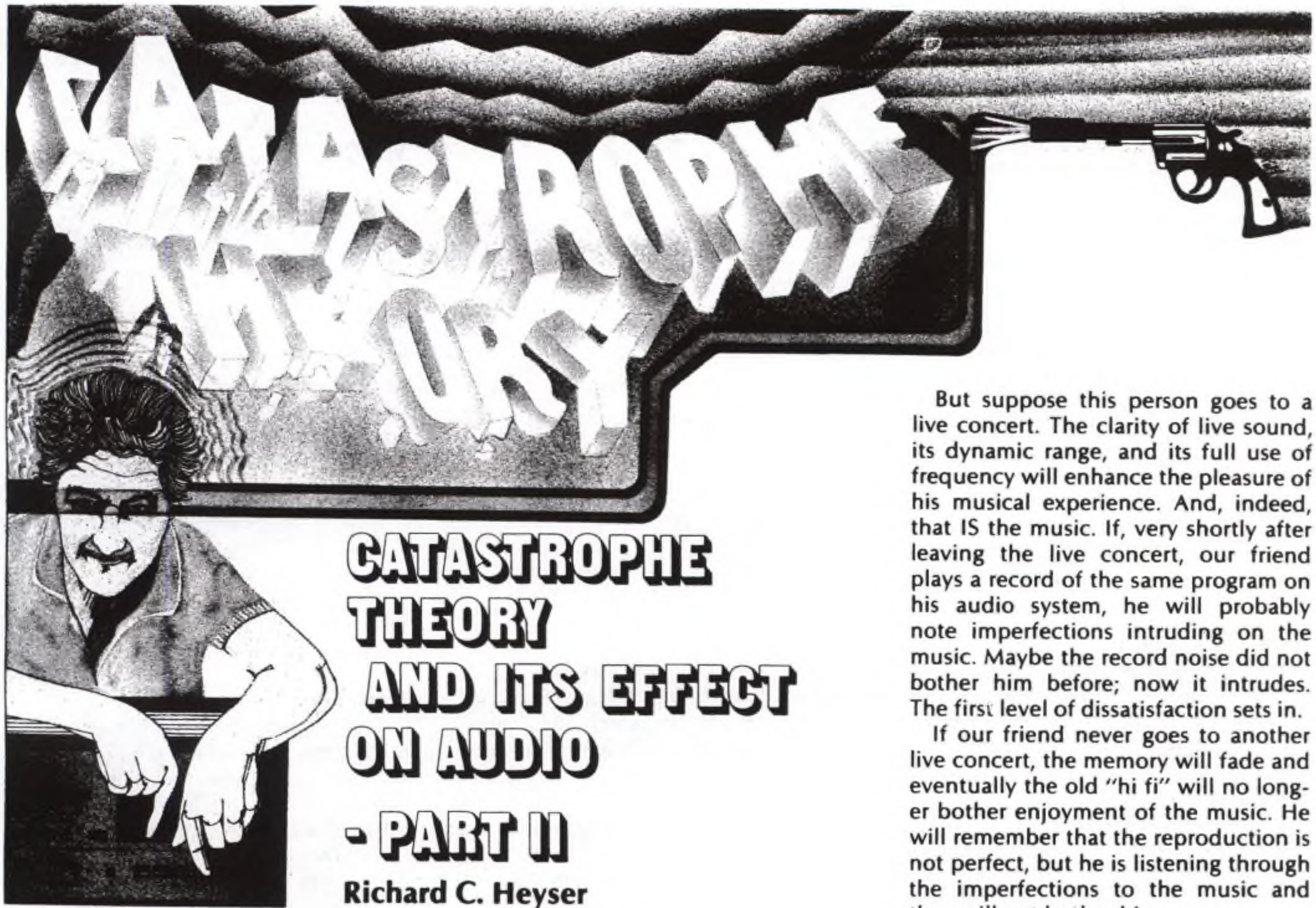
If you remember one of our earlier discussions (*Audio*, Feb., 1979, "A View Through Different Windows"), surface C is like a Flatland. A flatlander, living on C, knows only that there are magic lines which change the way people act when passed in one direction, but which cause no change when passed in the opposite direction. A higher dimensional being can observe the surface M and comprehend why a flatlander observes a seemingly magic boundary. But any attempt to convey this fact to a flatlander will be fruitless unless the flatlander is willing to accept the reality of dimensions above those of the world which he sees.

And there is a lesson here for us. For we, too, are flatlanders who sense patterns of response under conditions of varying factors. We can "see" the external factors and their relevance to the situation at hand. And we can observe how a dynamic system responds to those factors. But the hypersurface of behavior control is invisible to us. We sense the effect, but not the control. What Rene Thom has done is bring a higher dimensional concept to we flatlanders who could observe catastrophic changes in response, but could not reason why they should occur.

If we look at the behavior predicted by even this simple cusp catastrophe it is obvious that the properties we discussed earlier (jump, divergence, bimodality, hysteresis, and inaccessibility) are handily explained. In our next discussion, let us apply this new theory to audio.

References

1. R. Thom, "Structural Stability and Morphogenesis" (W. A. Benjamin, Reading, Mass., 1975).
2. Th. Bröcker & L. Lander, "Differentiable Germs and Catastrophes" (Cambridge University Press, Cambridge, 1975).
3. Yung-Chen Lu, "Singularity Theory and an Introduction to Catastrophe Theory" (Springer Verlag, New York, 1976).
4. E. C. Zeeman, "Catastrophe Theory, Selected Papers" (Addison-Wesley Publishing Co., London, 1977).
5. A. Woodcock & M. Davis, "Catastrophe Theory" (E. P. Dutton, New York, 1978).
6. T. Poston & I. Stewart, "Catastrophe Theory and its Applications" (Pitman, London, 1978).



CATASTROPHE THEORY AND ITS EFFECT ON AUDIO - PART II

Richard C. Heyser

Away, you scullion! you rampallion! you fustilarian! I'll tickle your catastrophe.

King Henry IV, Part II, II, i.

In the previous article, we discussed the elementary basis of Catastrophe Theory and suggested that it may be applicable to problems in the perception of sound. In this discussion, I would like to present a simple example to show how this can be done.

Catastrophe theory, if you remember, is a mathematical basis for modelling certain simple patterns of response that can be expected under the influence of conflicting drives. This is a general nonlinear theory which can be applied to the analysis of equipment and to the study of human behavior.

The theory gets its name from the fact that sudden and dramatic alterations in behavior, response catastrophes, can be predicted within its framework. What makes this attractive, from the standpoint of our perception of sound quality, is the structured analysis which it brings to bear on problems involving the emotional reaction of the listener.

Suppose we now consider a very simple and straightforward problem in audio: How might one's opinion of the quality of sound reproduction from an existing audio system be modified under the influence of two factors, the amount of live music one hears and the amount of reproduced music one listens to from this system?

First, let us postulate a scenario that draws only on our observations of human nature. Presumably, if the owner of an audio system likes music, he will continue to indulge himself by acquiring new records and listening to reproduced sound. If there are no interfering factors which can reveal imperfections in quality, there is no drive to modify the opinion of the present audio system. If the listener never goes to a live concert, he is probably satisfied with the music heard at home. The listener probably never thinks about the audio system and would be perfectly satisfied with music heard from a table-model radio.

But suppose this person goes to a live concert. The clarity of live sound, its dynamic range, and its full use of frequency will enhance the pleasure of his musical experience. And, indeed, that IS the music. If, very shortly after leaving the live concert, our friend plays a record of the same program on his audio system, he will probably note imperfections intruding on the music. Maybe the record noise did not bother him before; now it intrudes. The first level of dissatisfaction sets in.

If our friend never goes to another live concert, the memory will fade and eventually the old "hi fi" will no longer bother enjoyment of the music. He will remember that the reproduction is not perfect, but he is listening through the imperfections to the music and they will not bother him.

If, on the other hand, there is a larger percentage of time spent on listening to live music, there is a good chance that one night, when he comes home and puts a record on the turntable, it will suddenly dawn on him how lousy the sound really is. He no longer likes the sound of music played on his system. From that point on, the degree of discontent will grow in proportion to the ratio of live vs. reproduced sound that is heard. If he mostly attends live concerts and only rarely plays records at home, his knowledge of what live music sounds like will increase the discontent he has with the quality of his reproduced sound.

If he drops off in the amount of live concert attendance, but maintains a small, steady diet of listening to reproduced music, his discontent will, at first, slowly diminish. But without his awareness, there will suddenly come a time when he is so caught up in the music that he never once thinks about the record scratch that bothered him so much just a few evenings before. His opinion switched from moderate dislike to moderate like. If the ratio of live to reproduced sound continues to diminish, he will again resort to a condition of satisfaction with reproduced music.

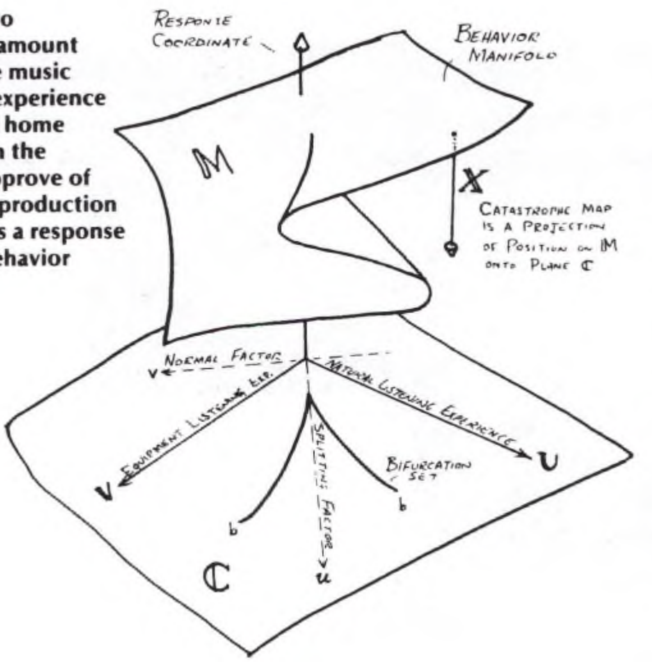


Fig. 1 — If there are two control factors, 1) the amount of experience with live music and 2) the amount of experience with listening to one's home audio equipment, then the degree to which we approve of the quality of audio reproduction from that equipment is a response that must lie on the behavior Manifold M.

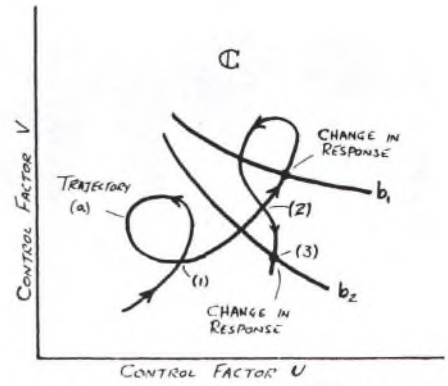


Fig. 2 — A flatlander, living on the control space C, will encounter mysterious, unmarked boundaries, b_1 and b_2 which will cause him to experience abrupt changes in behavior when passed in one direction, but not when passed in the opposite direction.

Cusp Catastrophe

Admittedly this scenario is quite simplistic. But the behavior is not out of line with human reaction. Let us once again set up the same problem, but this time use catastrophe theory to anticipate behavior.

There are two control factors: Amount of experience with live music and amount of experience with reproduced sound. There is only one response we wish to consider, the degree to which the listener approves of the quality of audio reproduction.

Two factors and one response define a three-dimensional behavior space. This three-dimensional behavior space is sketched in Fig. 1. From our previous discussion we know that the behavior manifold, the location of all stable responses under unchanging factors, will be a subspace with the same number of dimensions as there are control factors. The manifold M, is a two-dimensional surface.

This surface, as we discussed last time, forms a folded shape of the type shown here. The horizontal plane, C, represents the given coordinates of live vs. reproduced listening. We use the letter C because this represents the Control subspace within the higher-dimensional behavior space. It is also referred to as the Parameter space in some mathematical literature. In this figure, I have dropped the position of plane C down below the behavior manifold for illustrative clarity. It makes no difference where the plane C is located relative to M because our interest lies in the projected "shadow" of M onto the control space. By separating M from C, we can readily observe what goes on.

The orientation of coordinate axes on the plane C depends upon the nature of the factors which they represent. The derivation which we presented in our prior discussion developed the concept from a behavior potential which would give coordinates u and v, shown dotted in Fig. 1. In contemporary literature, the axes u and v are referred to as the splitting factor and the normal factor, respectively. These coordinates would be used for situations in which the response under consideration is normally influenced by a single factor in a smooth, continuous manner, while, above a certain threshold, the action of the second factor is to set up a trigger condition where slight changes in the normal factor precipitates larger than normal changes in response. The start-up conditions in a free-running multivibrator are examples of this; a perfectly balanced circuit could not oscillate when voltage is applied, but offset symmetry — the splitting factor — can allow circuit noise above a certain level to start oscillations that build up to full-limit cycles.

When there are conflicting factors which pretty much compete in their contributions to response, then there is a little bit of control and a little bit of splitting in each of them. These axes are then rotated relative to u and v, as shown in Fig. 1 by the solid lines marked by the capital letters U and V. In the case of audio listening, I will assume that the conflicting factors of live music experience and reproduced music experience are of this latter type. This does not mean that they are rotated 45 degrees with respect to normal and splitting, but that they have some amount of rotation.

Flatland

The projection of the behavior fold in M onto the surface C is called the Bifurcation Set, and this curve is symbolized here by the letter b. It is called bifurcation, or two-pronged fork, because two different kinds of behavior occur when we move our location away from this line. This is the boundary of precipitous behavior in terms of the control factors. This curve has a sharp point which forms a cusp, and that is why the particular type of behavior pattern associated with this type of precipitous response is called a Cusp Catastrophe.

In order to understand how our listening emotions enter the picture, refer to Fig. 2. Imagine that we are flatlanders living on the surface C. We are moved about our flatworld under the influence of two factors, and our position within flatland is marked by the coordinates U and V. Our emotional feelings alter with our position in flatland. As we move along the trajectory marked (a), our feelings smoothly and continuously change with our coordinate location. When, in our wanderings, we cross back to the coordinate location shown here as (1), we duplicate the emotions we previously experienced when passing this same place. But when our trajectory crosses the magic boundary b_1 we suffer a dramatic and sudden change in emotion. Just before we got to this boundary we were content and liked our state. At the moment we touched this boundary, our state flipped to that of a strong dislike. Our emotions suffered a catastrophic change.

Seeking to restore our status we loop back to position (2) which we had just before encountering this magic boundary and which we knew was a position of content. But our emotions do not come back to what they were. Now, at position (2) we still have a feeling of strong discontent.

Baffled, we retrace our path until suddenly, at another magic boundary, b_2, we catastrophically jump back in emotion state to our previous condition. We had previously passed this second magic boundary going in another direction and nothing had happened; now, coming back across it, our emotions are dramatically altered.

Living life as a flatlander, unable to comprehend forces outside our world, we would probably attribute this mischief to divine influence and might develop some interesting theories to explain why these things should happen to us.

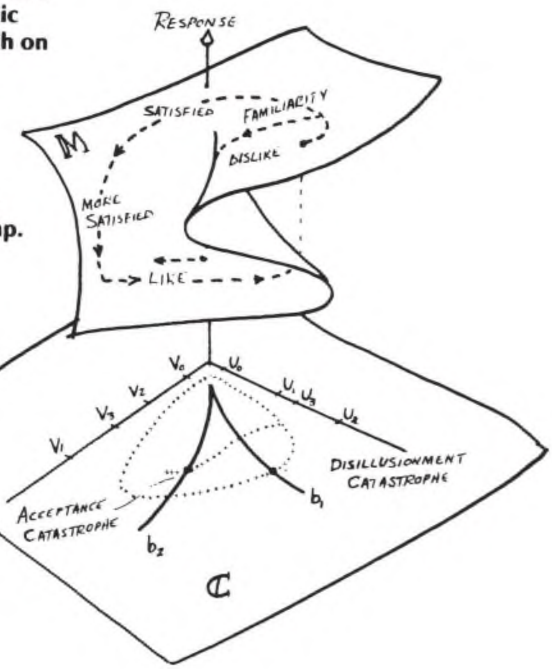
We might even devlop a *technocult* of flatland surveyors who, through ever finer instruments and more glorious linear mathematics, seek to quantify the measure of the geometry of flatland. Of course, these technocultists might be so burdened down with the weight of their precision apparatus that they cannot stray far from low-curvature regions where no catastrophic changes occur. Rumors of catastrophes might reach their ears, but no technocultist would ever accept the existence of such magic nonsense, since it was not only inconsistent with their linear mathematics, but could not be discerned by their survey instruments. In order to ease the fears of the *perceptofreaks*, who believe in such magic nonsense, the survey instruments are constantly being improved to measure imperfections to an ever finer resolution.

But a flatlander falling off a cliff takes little comfort from knowing that the science minds of flatland, who do not believe in the existence of cliffs, had developed a new flat measuring rod capable of resolving a nanometer.

It is not that the science minds are wrong; they just do not happen to be where the action is. They are under the wrong lamppost. If this situation sounds a little bit like our own problems in audio flatland, the resemblance is not coincidental.

What no one in flatland can realize is that his fate depends on higher dimensional influences. Let us go back to Fig. 1. The dramatic change called a catastrophe is a symbolic falling off a cliff in a higher dimensional space. The *catastrophe map*, shown here by the capital X, is the process of projecting the shadow of the actual position of response on the behavior surface M onto an apparent position in terms of the control space C.

Fig. 3 — As our exposure to live music and reproduced music changes, we trace out a path on the behavior Manifold M. Whenever our experience takes us past folds in this surface, our opinion of the sound quality which our audio system provides will take a catastrophic jump.



Journey on a Manifold

Let us take a journey on the manifold M. This journey is shown by the dashed line in Fig. 3 and starts out at the place marked "satisfied." Our *altitude* marks our *attitude*. Our height above the plane C (flatland) is a measure of response to the control parameters. The *higher* we are on M, the more we *dislike* the sound of the audio system. We are driven up and down this surface by the control parameters. We start this journey at "satisfied," the position of which is determined by control coordinates U_o and V_o.

At "satisfied" we have V_0 units of listening to our audio system and U_o units of listening to live music. Out of enjoyment of music, we begin to listen to more reproduced sound and begin our journey on the manifold M.

The more we listen to reproduced sound, the more that sound becomes our standard of performance. This drives our location on manifold M to a lower height, which means we become more satisfied with the sound of our audio system . . . or, looked at another way, the less we think about the quality of reproduced sound.

Then, around coordinates U_1 and V_1 we being to go to more like concerts. Our trajectory now takes a sharp change of direction back up the manifold. With increasing live music experience, our opinion of the old "hi fi" begins to drop, until somewhere around coordinates U_2 and V_2 we cross the magic boundary b_1. All of a sudden we experience a disillusionment catastrophe . . . our opinion changes from "like" to "dislike." The reason for this is that in order to remain on the surface of stable response, M, we had to jump from the lower sheet to the upper sheet where our trajectory took us past the fold. Under small changes in factors, we must take a big jump in response in order to stay on the manifold of stable response. When we approach a fold under smooth progressive drives, there is no way we can find ourselves on the inner sheet of M.

If we slack off on the ratio of live to reproduced listening, our opinion of reproduced sound quality will not snap back until we cross the boundary b_2 at position U_3 and V_3. Then, as we cross this boundary, our opinion will fall off the cliff, and we will suffer what I have referred to here as an acceptance catastrophe. We are back on the original trajectory and must accumulate a bit more live-listening experience before again experiencing a disillusionment catastrophe.

If, on the other hand, we simply slack off in both live and reproduced listening experience, we are passing along the path called here as "familiarity." The old habit patterns slowly take hold and we again will find ourselves at a "satisfied" status, back where we started.

Buy By

It is rather startling to contemplate the richness of emotional reaction which is revealed by even this naive catastrophe model. Each of us, I am sure, would like to believe that he is master of his own behavior under conflicting factors. But the trend in behavior pattern which is disclosed by Thom's theory reveals the existence of an inexorable machine which we ought to be aware of. Knowledge of this machine introduces a new factor into the game and raises the dimensionality to a higher level. It is a case of forewarned being forearmed; once we know that participation in a situation with two factors and one response yields a cusp catastrophe, we can introduce that knowledge as a new control factor and avoid the cusp. But our cleverness could also be our undoing since we may have changed the situation to one of higher dimensionality.

One of the situations in which knowledge of this elementary catastrophe can be of value is in the purchase of audio equipment. If instead of a live listening versus reproduced listening, we were to label the control factors: Listening to Brand A versus listening to Brand B, we can sense how a clever salesperson could walk an unsuspecting customer up the manifold to sufficient strength of opinion to trigger purchase of a component.

Suppose you had decided Brand B sounded pretty good and was an excellent match to your bank account. About the time you show signs of being ready to purchase Brand B, the clever salesperson lets you hear just a brief bit of sound from a more expensive Brand A. By this time you had disclosed which kind of music you like and had expressed satisfaction with the way Brand B reproduced that sound. So, quite by "accident," Brand A is punched up on that music.

You are at point U_1 and V_1 and suddenly the introduction of a better sound stops your downward plunge on the opinion manifold and pulls you in a new direction of upward motion. You like the music and your curiosity makes you want to hear a bit more. The smallest dissatisfaction with Brand B starts to set in, and a clever salesperson knows that if you can be persuaded to listen long enough you will get "hooked" on the better sound of Brand A. A good salesperson will not *force* you to listen to Brand A; you said what your purchase limit was and Brand B was at that limit. So the trap is sprung to let you sell yourself.

A simple A-B comparison switch is all it takes, with you the unsuspecting driver of the machine when you are allowed to switch the music back and forth between the two competing systems. Any increase in relative exposure to the sound of A versus B will inexorably drive you upward on the manifold. If you trigger a disillusionment catastrophe, the deed is all but done. Once you are on the uppermost sheet of the manifold, it is likely that you will subconsciously place the switch in the A position for an increasingly longer time than in the B position. You are driving yourself higher on the manifold. By that time the salesperson is mentally computing his commission on the sale of Brand A.

Maybe you do not dig math. Maybe the idea of topological manifolds in a behavior space does nothing to you. But just knowing of the existence of such things can save your wallet from a needless onslaught the next time you go shopping for audio equipment. At the very least, you can be aware of emotional forces which can be set into motion to present you with tempting bait. Once you grab such bait, the hook is sunk, and it is your own struggle which sets the barb in deeper and pulls you into the purchase of a component you did not previously want to buy.

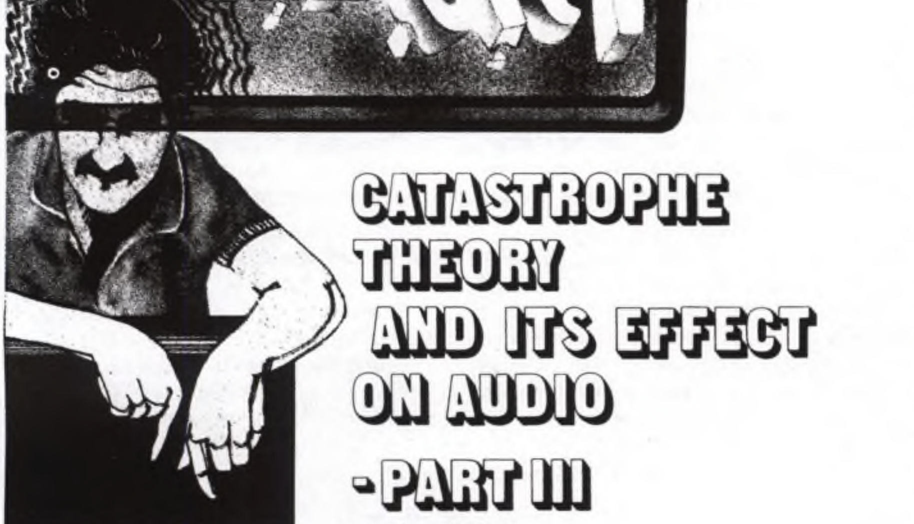
All Is Not Gold That Listens

Quite obviously, the more interesting situations arise when there are a multiplicity of factors, some conflicting and others of a splitting nature. We all recognize that emotional bias definitely plays a role in the reaction we have to conflicting circumstances.

Our individual perception of quality involves a delicate balance of conflicting factors, including our own involvement with one or more of those factors. The designer of a particular audio product may be a poor judge, from the standpoint of detached objectivity, of the relative merits of that product. And it must be admitted that the ratio of lead to gold in the ear of the listener is somehow related to the personal involvement which that listener has with the product being heard. This apparent rupture of objectivity, as perceived in the frame of reference of others, may occur without conscious awareness of the occurrence.

It is also possible that even in the presence of an emotional bias something can happen which will "change our mind" and alter the response we have to a given situation. Beauty, it is said, is in the eye of the beholder. But we all know that events can occur which catastrophically alter our perceptions even in the presence of prior strength of opinion.

In our next discussion we will consider another common audio situation which involves a higher dimensional catastrophe.



CATASTROPHE THEORY AND ITS EFFECT ON AUDIO - PART III

Richard C. Heyser

The existence of bimodality of response is a direct clue that the perception of quality is a nonlinear process. It should also be a clue that any objective measurements which are based upon linear theory will not be worth a hill of beans when we try to correlate those measurements with subjective value judgments.

After bimodality, the next most prevalent response characteristic which we note in human behavior is trimodality. Patterns of stable compromise may emerge between equally stable, but opposing, strengths of opinion. In some cases this may show up as an evolutionary transition between what Thomas Kuhn has called Paradigms.[1] A middle ground may be taken between the tradition of old ideas and the promise of new ideas. When this happens, the press of time or of evolving evidence will tend to resolve the compromise position to one paradigm or the other.

Three regions of stable response may occur under a number of circumstances. When a formerly clear-cut situation splits into discernible camps, those who occupy the outermost camps, which represent the extremes of opinion, will tend to assign rubrics to themselves and to their most extreme counterpart. They are generally self flattering and of a nature which denigrates the opposing view. These are the heraldic flags of strong opinion and the field of audio bears many such flags.

A compromise opinion or reaction to control factors will seldom be identified with strong labels. This is because the trimodal position is less frequently occupied then either of the two extreme positions which surround it. This does not detract from the fact that such a compromise position can be legitimately taken and possess high stability.

The Butterfly

It might normally be thought that since trimodality is the next degree of behavior complexity from bimodality, it would arise in the next higher level of control dimensionality. But this is not so. Two factors and one response yields the cusp catastrophe. Three factors and one response yields the swallowtail catastrophe; however, the swallowtail catastrophe does not exhibit much in the way of stable behavior, let alone trimodality.

Trimodality shows up when there are four factors and one response. And, sure enough, when we attempt to analyze those situations leading to a compromise behavior we find that four, not three, factors are involved. These factors have been given the names: Splitting factor, normal factor, bias factor, and butterfly factor.

The manifold of stable response is a hypersurface in the five-dimensional behavior space. The bifurcation set — places where a change in behavior occurs — lies in the four-dimensional control space. Obviously we cannot sketch a hypersurface in five dimensions, but the math game can be played without such limitations. The catastrophe map thus formed has been given the name Butterfly Catastrophe, in recognition of a certain abstract shape which appears when lower-dimensional "slices" of the bifurcation set are sketched.

In order to provide an audio example of this higher dimensional catastrophe, let us again consider the case of a music listener who has an audio reproducing system. Let me assume that the control space is characterized as follows: Music enjoyment is a normal factor, time spent listening is a splitting factor, cost of new equipment is a bias factor, and product awareness is a butterfly factor.

These are the four factors which will control the following behavior, desire to purchase a new audio reproducing system.

Butterfly Bifurcation

The first two control factors, normal and splitting, are identical to those of the two-dimensional example which we discussed earlier. The response is also the same if we presume that a sufficient dislike of one's existing audio system will relate to the desire to replace that system with one which sounds better.

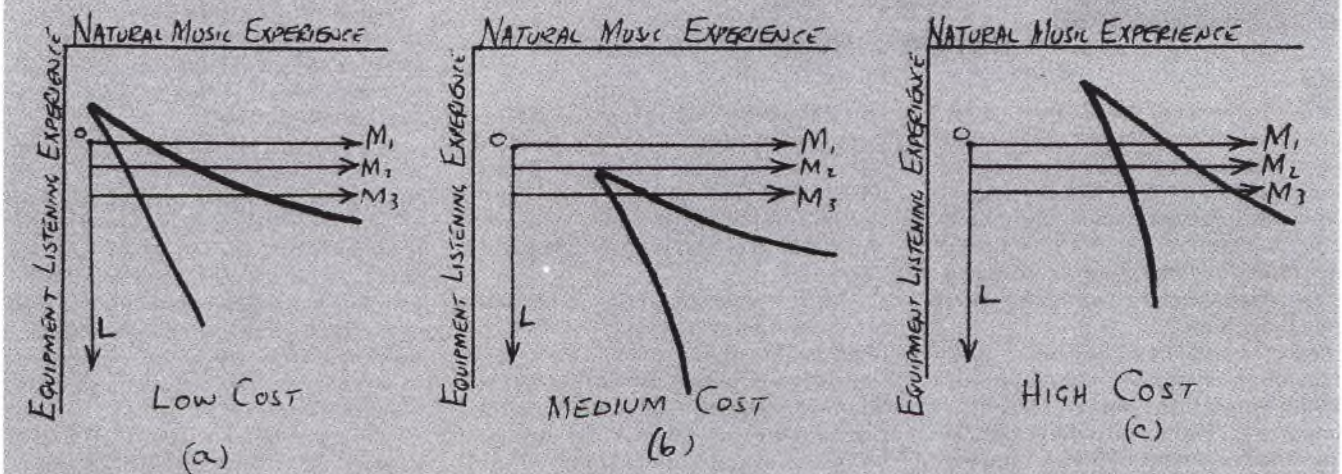
The terms which are new in this example are the bias factor and the butterfly factor. Bias is that factor which, if all else stays put, tends to multiply the response. I have presumed that one's bias to purchase something is inversely related to the price of purchase. Hence cost is a bias factor. The butterfly factor has the effect of an almost inexorable buildup in pressure which tries to upset the status quo. In this case I have assumed that the listener also reads magazine articles on sound reproduction and cannot help but see advertisements and product claims. Thus, product awareness becomes a butterfly factor.

It is a feature of Thom's theory that all of the lower dimensional forms of abrupt behavior can be found in the higher dimensional catastrophes. Progressing upward in dimensionality adds new types of catastrophe to the inventory. The fold catastrophe (which we did not discuss, but is the either-or hysteresis jump available in a one-dimensional control space), the cusp catastrophe (control dimension two), and the swallowtail catastrophe (control dimension three) can be found under certain conditions when there are four control factors. Obviously I cannot go into any sort of detail in this brief discussion, so I will present what I believe may be the more important patterns for audio.

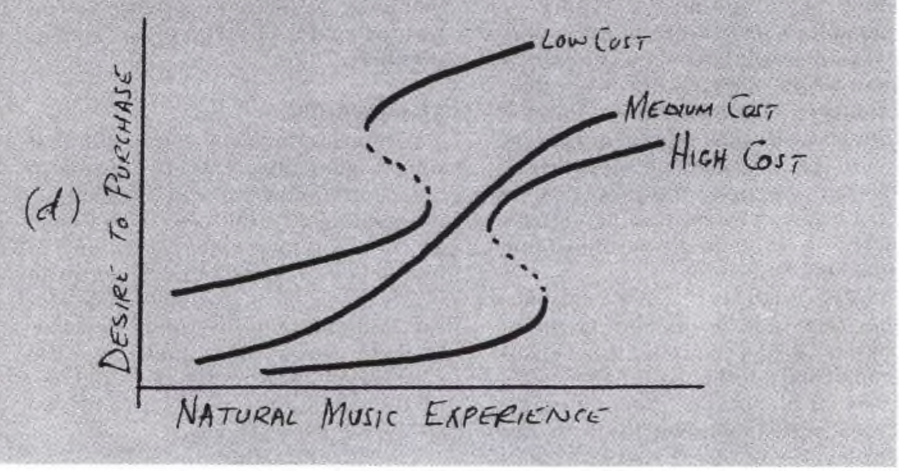
We cannot show four dimensions, so let us concentrate on the bifurcation set (places of abrupt decision) as it appears for the two-dimensional slices with the coordinates of music enjoyment and time spent listening. We will follow the pattern for various cost and product awareness situations. In Fig. 1, I have plotted the bifurcation set as a function of decreasing cost of a new system and with almost no product awareness. The behavior is cusp-like with the cusp swung farther toward the direction of live music enjoyment as the cost of a new audio system increases. If we trace the journey marked M (for music enjoyment), our desire to purchase a new system will increase as we increase our enjoyment of that music. If we trace the journey L (for listening to our existing audio system), the desire to purchase will diminish with increased listening. This is the same sort of situation discussed in the previous example of the cusp catastrophe.

As an additional set of curves, I also plot the desire to purchase (height of the manifold) as a function of cost. Now we can see a situation emerge which we might not have anticipated. At low levels of listening to reproduced sound, the curve of increased live music experience (M_1) cuts the bifurcation set at low cost and high cost, but misses it for medium cost. For a given small amount of listening to reproduced sound, there are response

Fig. 1—Cost, as a control factor in our desire to purchase a new audio component, has the effect shown here when there is very little product awareness concerning components that are available for sale. Three cost situations are sketched: (a) low cost, (b) medium cost, and (c) high cost. Under these conditions the desire to purchase a new audio component will depend upon the relative amount of experience that we have with natural sound as compared to the amount of time we spend listening to the sound reproduced from our present audio system. If, at the point of relative experience shown here as O, we substantially increase our exposure to reproduced sound, without changing the amount of time spent attending live concerts, then we move out on the trajectory L. Travelling on L tends to reduce the desire to purchase a new system. Increasing the exposure to natural sound will move us in the direction M.



Of particular interest is the situation that arises if we travel the trajectory M_1. Figure 1(d) is a sketch of what happens when we move along M_1. As we begin to attend more and more live concerts, our general desire to purchase a new audio system will increase. This increase will take on catastrophic jumps at low cost and high cost situations, but not at medium cost. The low cost catastrophe corresponds to impulse buying, while the high cost situations may be a prestige reaction.



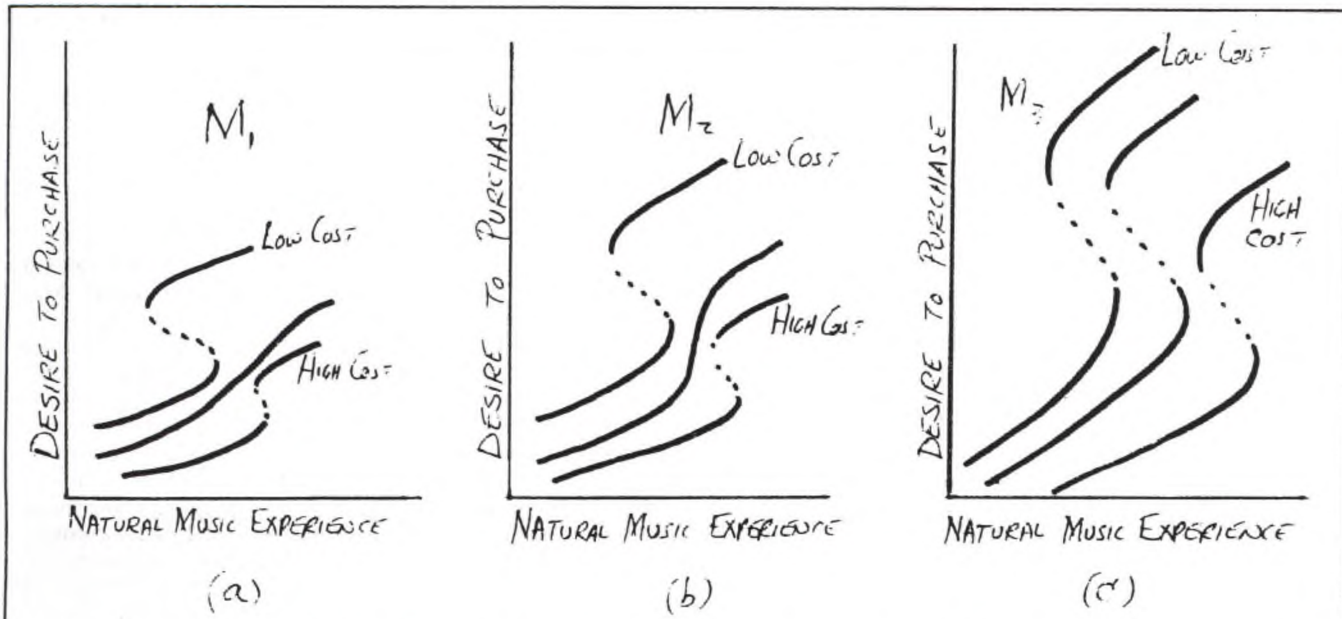


Fig. 2—The result of moving along trajectories M_1, M_2, and M_3 in Fig. 1 are sketched in 2(a), 2(b), and 2(c). This illustrates the effect that increasing amounts of home music listening have when we suddenly begin to go to more live concerts. Although the initial desire to purchase a new audio system is lower when our background is that of a great deal of home music listening, 2(c), and although it takes more live music experience to trigger a disillusionment with our present system, when the catastrophe does come it is much more intense than had we started from a more modest background of home music listening.

catastrophes at low cost and at high cost situations, but there is a region of cost in which purchase desire is continuous with no jumps. In all cases the desire to purchase goes up as the cost of a new system comes down, but there is a certain range of time spent listening where our opinions take a jump.

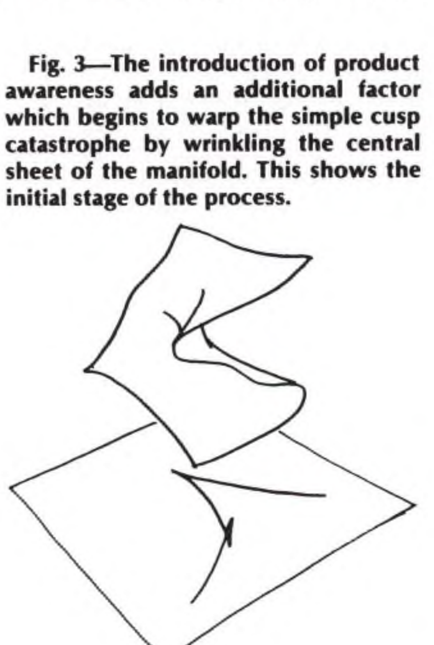
This jump at low cost is the sort of behavior which can lead to impulse buying. We all recognize the situation; one day we run across a bargain too good to pass up. We did not really intend to buy a new cartridge or "whatever," but the cost was "right" and the impulse hit us.

The tendency toward impulse buying, according to this butterfly catastrophe situation, will fade as the price rises. Impulse will give way to a smooth curve of deliberation of worth versus cost. But as the price continues to rise we will again enter a region in which our increasing enjoyment of music will cause a sudden jump in purchase desire. This does not mean we will buy the higher priced product, but our desire will "gain ground" faster than we might anticipate as our enjoyment of good sound increases.

The effect which a greater amount of time spent listening to our present system has on the desire to purchase is to offset cost. The trade-off between enjoyment of music, cost, and more listening, is shown in Fig. 2. The interesting fact which emerges from this situation is that while the general desire to purchase is diminished by more listening to our present audio system, the potential for impulse purchase is greater. Not only do the low-cost impulse catastrophe and the higher cost catastrophe merge to eliminate a smooth change in desire, but the magnitude of behavior jump is much greater due to the increased listening experience.

If we believe the mathematics, the person most likely to whip out his checkbook and make a surprise purchase (even to himself) is the one who does a lot of listening to reproduced sound at home and who has recently become more interested in the enjoyment of live music. That does not seem so surprising, but the magnitude of the desire catastrophe is a surprise.

Fig. 3—The introduction of product awareness adds an additional factor which begins to warp the simple cusp catastrophe by wrinkling the central sheet of the manifold. This shows the initial stage of the process.



The Effect of Advertising

Now let us look at the emotional effect of product awareness. What role does advertising and product chest thumping have on the desire to purchase?

In order to visualize how product awareness (the butterfly factor) can precipitate trimodal behavior, Figs. 3 and 4 sketch the way in which the cusp catastrophe becomes modified with the introduction of the butterfly factor. As product awareness begins to increase, Fig. 3, the simple fold of the cusp catastrophe begins to warp and convolute. Continued increase, Fig. 4, puts a third sheet in the manifold and converts the bifurcation set (the projection of the "edges" of the fold where opinions must jump) into a complicated multi-cusped pattern. This pattern has been likened to an abstract sketch of a butterfly, and is the basis for the name given to this particular response catastrophe.

If product awareness were to increase, the middle sheet would continue to produce a pattern like Fig. 3, but with the line segment break in the bifurcation on the other part of the major cusp line.

Figure 5 shows sketches of the bifurcation set and purchase desire plotted for increasing amount of product awareness, but at fixed cost for a new system. A trajectory of increasing music enjoyment for a critical range of listening is shown by the line N. At a certain level of awareness and at a certain level of cost, the desire to purchase will experience a trimodal behavior. This third mode represents a compromise struck between a strong desire to purchase and a weak desire to purchase. The feeling one might have is "I sure would like to have that system, but it is just more than I can afford."

Trimodal behavior will disappear if any of the four control factors change moderately, but it will disappear the quickest with a change in the butterfly factor — product awareness. There is, in other words, a critical threshold of advertising which is required to cause the greatest increase in desire to purchase. If the listener can be bumped up to a compromise reaction, it is easier to push him upward to a purchase

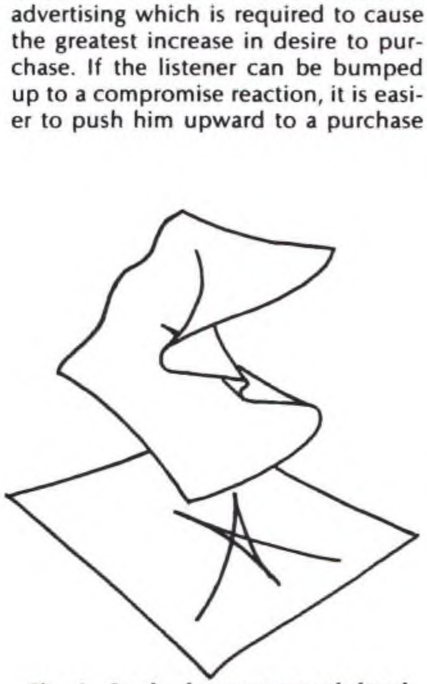


Fig. 4—In the later stages of development of the butterfly factor, the manifold develops a third inner sheet. The bifurcation set which this produces has the shape indicated in this sketch. The term "butterfly" is taken from the shape of this bifurcation set, which has been likened to that of an abstract butterfly.

with a modest increase in advertising than with a proportionately larger drop in price.

As we pointed out, the shape of the bifurcation set at the place where trimodal response sets in is the basis for the name Butterfly Catastrophe. The central region of this set — the body of the butterfly — is often called the pocket of compromise. The effect of the butterfly factor is to drive the bifurcation set from a single cusp shape toward and through this butterfly shape. The effect of the bias factor is to magnify the response at any given set of conditions.

By opening up an intermediate level of stability between an otherwise large jump, the butterfly factor is a trigger mechanism. If we were at place (a) in the response shown in Fig. 5, the size of the jump we would take if we ever got to the fold in the behavior manifold would be quite large if the bias factor (cost) were sufficiently strong. But we do not have to change our listening habits if the butterfly factor is now increased. Symbolically, we are standing on a floor with the ceiling well above us, and the butterfly factor now ripples our floor and bumps us up to an intermediate shelf level between floor and ceiling. Position (a) now changes to position (b) as increasing product awareness opens a pocket. It is now much easier for a modest increase in normal factor to take us to a jump point where we are not at the ceiling — a ceiling we might never have reached without that assistance.

I have called product awareness the butterfly factor because that is the drive which advertising, equipment reviews, and sales claims provide. We are constantly bombarded with advertising and product claims. Most of the time this has no effect on us, and we tend to wonder why advertisers spend so much money and time. The power of advertising, according to this butterfly factor behavior model, comes into play when we become interested in possible purchase of a new product.

Months can go by with the same ad appearing month after month and we barely notice it; then, through music enjoyment and time spent listening, we start to pick up an interest in possible purchase. We start to notice the ads (product awareness begins to increase), and we begin to compare prices (cost factor entering) even though a few months prior we paid no attention to prices or what was new.

This is quite consistent with human behavior, and there would be little reason for mentioning something so obvious to us all. But the butterfly model indicates a condition which we should be aware of. The butterfly factor plays such a dominant role in precipitating a large catastrophe that it "comes on strong" after we have achieved a certain threshold in desire to purchase.

Once we have been pushed high enough on the behavior manifold, there is little time left for rational analysis leading toward a purchase. At that critical stage we are likely to be triggered into purchase by almost any product claim, either in print or verbally by a salesperson. Geometrically, the gradient of the response manifold has its steepest value just before a catastrophe, and the steepest gradient of all will generally be due to product awareness in the audio situation we are considering. All it might take to precipitate a catastrophe is some small increase in desire to purchase, such as perceived cosmetic improvement over competition or an exaggerated product claim. At that crucial stage we, the audio purchaser, are likely to accept product reviews or hearsay comments or advertising claims which we might normally reject as pap or worthless. Wild claims and come-ons are not intended for the person who has no desire for purchase, but are geared for the last stage of purchase intent when we are most vulnerable.

A good salesman is much like a shepherd who, through artful means, keeps the prospective purchaser on a path which will lead to a commitment to purchase. Mathematically, it is the salesperson's task to keep the customer heading in the upward slope of the response manifold. Skillful employment of product awareness can steepen the slope to the place where a customer can be triggered by an otherwise minor increment in product claim. "Would you prefer this model in walnut or mahogany?" is one such technique which directs the customer upward on the manifold by concentrating his attention on alternative positive features of the product and diverting consideration away from the customer's natural contemplative act, namely, whether he wanted to acquire this product at all. Any situation that offers the opportunity for the customer to consider "not buying" as an alternative, is heading the customer the wrong way on the response manifold (from the salesperson's view) and is to be avoided. This is particularly important prior to the point where the customer can be triggered to a desire catastrophe, but, as we shall see, is still important after the catastrophe occurs.

Cooling Off

It is not what it is, but what it appears to be, that precipitates a response catastrophe. Once we are triggered, we have jumped to a higher sheet on the response manifold, and once there we are at a place of much higher desire and of lower gradient in response. Product awareness and cost have done their deed and may now be modestly altered without triggering a disillusionment catastrophe. The sales slip can add up to a bit more than we thought it would (". . . of course, there are shipping charges and . . ."), and product awareness can show modest negative factors (". . . the color scheme you have chosen is not in immediate stock and will take a few weeks for delivery"), but we are hooked at the moment the catastrophe occurs and will tend to drive toward a purchase anyway.

Generally, product awareness still has the highest gradient among the control factors, even after we have been triggered. That is why a "cooling off" period, a time lag, is so important before we sign a commitment to purchase. What has happened to us is something familiar in human affairs. A desire catastrophe is a response similar to "falling in love." Once we have "fallen" for something, it is often said that we are blinded against negative factors. That is not completely true, according to the catastrophe model. What has happened is that the gradient of response is significantly reduced when we jump. If we jumped from a place with a large enough splitting factor and normal factor, the increment in desire is so large that it could be likened to a transition to blinded love — the change in gradient is that large. But if the continued press of product awareness is now directed in a negative manner, the gradient will tend to increase downward. Depending on the interaction of the bias and butterfly factors, the response may stabilize or it may be triggered to a disillusionment. A "cooling off" period may then be said to allow us to "come to our senses" in the rational consideration of our desire (or need) for the audio component.

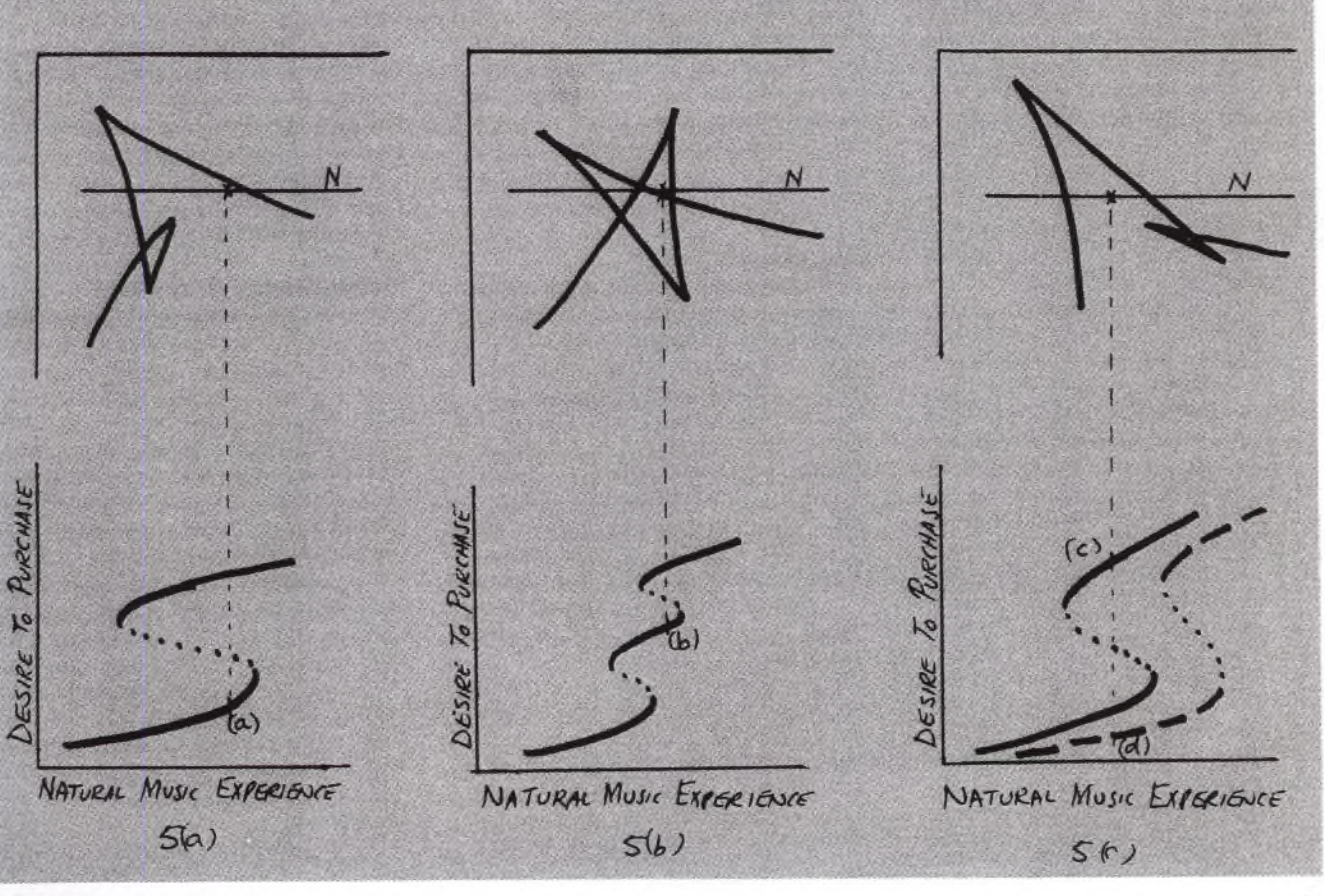
The role which advertising (and product reviews) plays in this process

Fig. 5—The butterfly factor introduces a trimodal condition of stability. This figure illustrates the influence which advertising has on desire to purchase an audio component, under conditions of constant cost but for the situation in which we suddenly begin to go to more live concerts. If we are passing along the trajectory N and are at the locations shown by the x, then advertising can be used to bump us up to a higher desire for purchase.

Fig. 5(a)—If there is minimal advertising, we are at the position shown as (a). We need to attend more live concerts before experiencing a substantial increase in desire to pay the cost of a new audio system for listening at home.

Fig. 5(b)—Advertising has the effect of rolling the manifold back toward lower levels of natural sound experience and of introducing an intermediate sheet. Even with no further increase in live music experience, advertising has bumped us up to position (b). We are in a pocket of compromise and have had the incipient catastrophe moved closer to our position. It now takes less of an exposure to pop us up all the way to a higher desire-to-purchase.

Fig. 5(c)—If the pressure of advertising is increased, it may be possible that as the pocket of compromise moves out, the wrinkle in the third sheet pops us up to position (c), a position we could never have achieved without the influence of advertising. Beyond this point, further pressures in advertising will continue to force the bifurcation set farther towards higher levels of live music listening. This means that there is an optimum level of advertising which can trigger us to a locally highest desire-to-purchase, but beyond which further sales pressures will tend to diminish desire. In extreme situations the additional warping caused by high levels of advertising may lose the sale by popping the point (c) back down to some place like (d). Supersaturation by advertising not only turns us off, but actually raises the amount of listening we must do in order to want to purchase that component.



can be helpful if we inject a delay between initial response and eventual commitment. (Within the framework of Catastrophe Theory, there is a process called delay which relates to the conditions under which a transitional response change must occur and refers to a "smoothing" effect on such changes — it is not this smoothing delay to which I refer.) While it is apparent that the initial impact of product awareness can be that of precipitating a response catastrophe, blatant claims and product puffery will have less of an effect during a "cooling-off" period and, in fact, can turn a sales away if it must compete with accurate product claims when the purchaser is allowed to exercise rational judgment.

A good salesperson instinctively knows that a customer who wants to "think about it" will probably not come back once he walks out the door. Thus, a salesperson who knows what he is doing will not present us, the customer, with an opportunity for a cooling-off period. We must take this step ourself.

During a cooling-off period between a desire catastrophe and commitment to purchase, increased product awareness can drive us back toward a disillusionment catastrophe if the added knowledge reveals things which *we ourselves would consider undesirable*. That is most important. The shape of the behavior manifold, and where we are on that manifold, is *different for each of us*. Two persons can react differently to the same product if they bought it during the passion of sales and then took it home for listening. At the moment of desire catastrophe, both persons might be equally convinced that this is the product for them. But the person with the higher gradient of response will become more quickly disillusioned with that product if additional negative features are revealed. Product awareness on audio components includes knowledge about sonic imperfections. If those imperfections are of a type which will detract from the enjoyment of sound, *in the frame of reference of a particular purchaser*, then it is helpful in the long range satisfaction of that purchaser that he know about such imperfections *before* buying the product, not after. The person who might become the most unhappy about the performance of a product is the one (with the higher gradient) who can best benefit from a cooling-off period during which he is allowed to compare product claims and do comparative listening.

If there is a moral which catastrophe theory teaches us, it is that the time to pay attention to product claims is *before* we are in the store and exposed to the heat of salesmanship; then we should allow ourselves a cooling-off period after we fall in love with a product.

If, during the cooling-off period, the customer becomes aware of enough negative factors, he may find himself dropped into a pocket of compromise. His trimodal condition lies between outright rejection of the product and wild acceptance. Because the drive of product awareness has been stalled by opposing trends, cost may now play a more dominant role in popping the desire to purchase upward to a higher sheet.

Since the bias factor (cost) tends to magnify response, a reduction in price (shopping around for a better deal) may set up a condition where an increase in splitting factor (time spent listening) triggers a jump upward in desire to purchase.

There is no surprise here; shopping around for a lower cost is, or should be, a requisite for any rational purchase. However, the combined effect of bias (cost) and splitting/normal (music listening) factors can now set up another trigger condition. A good salesperson can pull the butterfly trigger by letting a customer know that demand for the product has so thoroughly outpaced deliveries that this demonstrator model is the last one in stock. Desire catastrophe! Pull out the checkbook! What salesperson could be cruel enough to turn down the tearful request of a customer who wants to take home the prize of his desire. This particular ploy also sidetracks any further opportunity for a cooling-off period, so beware.

Approximate C. T.

The foregoing analysis applies when there are four distinct, identifiable factors and one response. One's own personal emotions, when considering purchase of an audio component, may involve more than four factors, and there may, indeed, be several responses which those factors elicit. In setting up an example which illustrates the application of Catastrophe Theory to audio, I have chosen what I consider to be the most significant factors and have tried to relate them to the mathematical terms to which they most nearly correspond.

It is my opinion that, at this time, the most important use we can make of Catastrophe Theory is to uncover trends in response under the influence of conflicting factors. When there are many factors, but four of them are dominant and relate to each other as does the normal, splitting, bias, and butterfly factors, and when there is one dominant response, we can use Catastrophe Theory to determine the *most likely* behavior. The influence of other, less dominant, factors will not change major aspects of the probable behavior, but will color the details of that response.

While not explicitly stated in most technical discussions of Catastrophe Theory, one can often simplify the analysis of a problem by reducing the dimensionality to that of the most dominant behavior space. For example, if there are 10 apparent factors, but only four of them stand out as dominant, then the situation can be reasonably approximated by the butterfly catastrophe. If, in turn, two of these four factors are considerably more important under a given set of circumstances, then we can resort to the cusp catastrophe.

The reason I suggest we use this simplification wherever possible (which I personally call "approximate catastrophe theory") is because of the enormous increase in detail complexity which occurs with rise in dimensionality. It simply gets out of hand and we may tend to lose the forest (general trends) for the trees (fine details).

When there are two responses, the type of catastrophe maps which are involved produce hypersurfaces which Thom calls "umbilics." Within Elementary Catastrophe Theory there are six umbilic catastrophes, ranked according to dimensionality of the behavior space. Time does not permit us to discuss the application of umbilic catastrophes to audio, although there are several of these and, perhaps, at a much later date we can discuss them.

Mathematics of Emotion

In this discussion we have essentially been considering a mathematics of human emotion. Improperly used, such a theory could cause considerable mischief; however, one of the quickest ways to defuse this potential weapon is to be aware of its existence. That, in part, is why I have not hesitated to discuss what might otherwise be considered a touchy subject.

Computer-programmed advertising campaigns, or, for that matter, political campaigns, may not be very far off. As a consequence we must be aware that any assault on our pocketbook must begin with a play on our emotions. That is the game, and we all play it. And as long as we all understand the rules, the match is balanced and the game is fair.

Up to now each of us has had to

learn the rules by experience, either our own or what we observe from others. Now, a weird, far-out, abstract mathematics has come into existence which, while not originally developed for that purpose, seems to model some of the primal rules of human emotion. As a service to the readers of *Audio*, I believe it is better to point out the existence of this new game plan and to let you know what might be coming, than to suppress the knowledge in hope that overzealous advertisers would never find out about it.

On a more pleasant note (that is, less sinister) we can now begin to appreciate the part that human emotion can play in our judgment of the subjective listening quality of audio components. All of us, I am sure, have experienced the situation where we "liked" the performance of a particular component one time, then something happened and we "didn't like" the same component when we heard it again. Perhaps we learned something which caused us to change our minds. It is only human; a new factor was introduced. But the component did not change — the measured technical performance did not change — *we changed*. Or, more properly, our response changed.

In light of what we have been discussing, this does not seem so mysterious. Yet think how capricious this might seem to a technically oriented "flatlander" who found that nothing had changed in the technical performance of that component.

The investigation of these delights and other properties of perception are yet to come. They can be the subject of a future discussion.

Subjective Impressions

The final point I would like to discuss in this three-part series involves the reason I am personally interested in Thom's geometric theory. The intangibles of audio include perception, cognition, and valuation. Perhaps with Catastrophe Theory we can, for the first time, begin to understand how one's personal subjective impressions of quality can be linked to conflicting circumstance.

We can begin to realize that it is not inconsistent for us to "like" something one time, then "not like" it another time, even under seemingly identical conditions. It is not where we are, that determines the intensity of our emotions, but *how we got there*.

The old complaint "why can't someone measure what I hear?" now takes on a different tone. We can "measure" under stable conditions, but that may not be sufficient for the determination of subjective valuation. Education, knowledge, experience, and training are factors involved in subjective valuation. It is often said that we can hear with different ears. It now appears that this is quite true.

We can be fooled into perceiving one thing while actually being subjected to something else. The stage artistry of a master illusionist can make us "see" things contrary to reality, such as flowers from thin air or a person sawed in half and then reassembled in front of us. Similar artistry can allow a ventriloquist to "project" a voice so that we "hear" it coming from an impossible location. And even as early as the 1920s a live vs. reproduced experiment was conducted in which an audience was substantially unable to distinguish whether the sound they heard coming from a stage was that of a live performer, whom they could see, or the playback of that performer from an Edison acoustic phonograph. The perceived sense of reality in all such cases is an illusion supported by factors other than those of the principal sensory organ which is involved.

So, too, it is in this present business of audio reproduction. It is not reality, but the illusion of reality which the present audio industry depends upon. We must perceive an acceptable illusion in order to have any success in so-called high fidelity reproduction.

The type of analysis embodied in Catastrophe Theory might begin to address this very important audio problem from a new direction. We can begin to look at the human process of perception, cognition, and evaluation as responses under sets of factors, some of which may be conflicting. We might also begin inquiry into a very important problem in present audio, namely, what are those factors necessary to support an acceptable illusion?

It is not a retreat either from reality or from technology to accept the existence of human emotion. Nor should technology ignore the role played by emotional response in establishing the quality of the listening experience.

Returning to the theme we discussed in the first part of this three-part series, those of us who pursue endless quests of measuring vanishingly smaller amounts of system distortion for sine-wave and square-wave signals are scrambling around under the wrong lamppost. The wrong lamppost, that is, if we want to correlate such measurements with subjective quality. It is the wrong place to look for two reasons: First, those distortion measurements are derived from linear theory and are, at best, on shaky ground; second, such measurements take no cognizance whatsoever of the intangibles of the listening experience. The light is not so bright and cheery when we walk into the bushes and search for audio truth with nonlinear tools.

Castastrophe Theory is one of those newly evolving mathematical tools which show promise of being able to provide us with a framework for handling those types of distortions, either real or imagined, which lead to instability of perceived form. It can score the most heavily in those situations poorest handled by linear methods. I really do not know whether Catastrophe Theory can be of any lasting value in our understanding of the dilemma of perception. But we will never find out until we give it a try. A

Reference

1. T. S. Kuhn, ***The Structure of Scientific Revolutions,*** The University of Chicago Press, Chicago, 1970.

Reprinted from *JAES*, vol. 32, no. 1/2, pp. 23–25, 1984 January/February.

Determining the Acoustic Position for Proper Phase Response of Transducers*

RICHARD C. HEYSER

Tujunga, CA 91042, USA

Sound-path delay has been accurately removed from the measured phase response of a transducer when the derivative of phase response with respect to frequency approaches zero as the frequency increases without limit.

The frequency transfer function of linear transducer systems, such as loudspeaker arrays and hydrophones, is often measured at some distance. This adds a term in the phase response due to the finite speed of sound in the measuring medium. Some concern has been expressed to me about the proper way to remove this unwanted phase term and determine not only the true phase response of the transducer under test, but also its effective acoustic position. There is a straightforward method of doing this, but it requires examination of the frequency response outside the normal passband of the device under test.

The delay plane expansion of a band-limited network coalesces to the proper time delay for very large values of pitch [1]. This means that the alternative frequency domain representation approaches a high-frequency limiting form such that the plot of phase angle versus frequency becomes a straight line whose slope is related to time delay offset in the measuring process. If the value of time delay in the measuring medium is to be subtracted in the measurement process, as can be done for impulse, cross spectrum, and time-delay spectrometry measurements, then the proper value of time delay has been subtracted, and the true acoustic position for phase measurement has been obtained, when the plot of phase angle versus frequency approaches a flat *horizontal* line for frequencies well outside the passband of the device under test. For most practical transducers this final value of phase will either be zero or 180° with respect to the polarity of the applied test stimulus, when the propagation time delay has been accurately removed.

The theoretical basis for this procedure is thoroughly developed in [1] and [2] which show that 1) all linear lumped constant systems, including those involving multipath, can be expanded as a linear *sum* of generalized all-pass terms, and 2) the all-pass is the *only* network for which true causal time delay is given by group delay.

The time delay that is determined by this procedure corresponds to the moment of arrival of the earliest possible energy component, caused by energetic stimulus to the transducer, which is capable of exerting causal influence at the point of measurement. It is the earliest signal that can do work, and indicates the nearest acoustic position of the transducer. This is the time delay to use for determining the proper phase response of a transducer. Subsequent arrivals, with higher relative energy content, will correspond to the frequency-dependent smearing of the effective acoustic position of the transducer behind this location, as predicted and measured in [2]. The mean average acoustic position for this subsequent bulk of energy arrival can be approximated by noting the time delay that causes the plot of phase angle to have zero slope (on average) over the largest bandwidth encompassing the highest amplitude response. The amount and phase of this arrival pattern of energy is more accurately displayed in the energy–time curve [3].

When multiple transducer subsystems are involved, such as woofer, midrange, and tweeter assemblies with their attendant crossover networks, it may be necessary to measure the acoustic position of each subsystem separately; otherwise a low-frequency driver that is positioned well in front of a higher frequency driver may have its advanced position masked by the larger response of the higher frequency driver at very high frequencies. Subsystems which are not involved in the test should be deactivated, with care taken to assure that this does not alter the acoustic properties of the subsystem under evaluation. This is where common

* Manuscript received 1983 July 5.

sense and a complete measurement of amplitude and phase, corrected for time delay, can be helpful.

The requirement for determining the asymptotic slope of phase places severe demands on the measurement process. To begin with, measurement at a single value of frequency is insufficient; it is necessary to know the shape of the phase response in order to estimate its asymptotic behavior. In many practical cases the true phase angle approaches its asymptotic value very slowly. Accurately scaled examples of this can be seen in [1, Fig. 7] and, coincidentally, in [2, Fig. 7]. Depending upn the desired accuracy and upon the nature of rolloff of transducer response, the measurement may need to be performed at signal levels as low as 60 dB below that which might otherwise be obtained in the transducer passband. Even then, the proper value of asymptotic phase may not be closely achieved if the amplitude cutoff is very rapid, and it may be necessary to estimate this limiting value by extending the measured curvature in phase to a final horizontal value.

As a practical engineering consideration, errors which are made in the estimate of this high-frequency limiting value will result in much smaller errors in knowledge of the phase angle within the passband of the transducer under test. For example, if, due to environmental noise, the time delay which is required to correct the arrival time of a 5-kHz low-pass driver can only produce an estimate valid to within 10° at 30 kHz, then the proper angle at 5 kHz has been established to within 1.7° and the arrival time has been corrected to slightly better than 1 μs.

On an additional technical point, examination of the all-pass expansion, developed in [1] and [2], shows that the locus of points on a polar (or, as it is commonly called, a Nyquist) plot of frequency response must always curl in a clockwise direction for increasing values of frequency if the system is causal. If it is not possible to isolate components of a larger system, or if it is suspected that some portion of a system has an earlier arrival masked by larger, later arrivals, then the Nyquist plot should be carefully examined. There may be many whorls on this plot, corresponding to individual all-pass arrival components, but if, after subtracting the presumed time delay from the measurement, a whorl is found which curls counterclockwise with increasing frequency, then such a whorl represents an earlier arrival and the presumed time delay is incorrect.

It should also be noted that since in-phase and quadrature parts of a causal frequency transfer function are always related by Hilbert transform, the limiting cutoff values of these two components should asymptotically approach zero for any band-limited system when the proper time delay is removed from the measurement. They will not oscillate above and below their final limiting value as they vanish. The tangent of the response phase angle is the quotient of quadrature and in-phase component, which approaches a constant value even as the components vanish. This also means that the final value of the corresponding Nyquist plot will be characterized by a "plunge" straight into the origin when the proper time delay is removed.

For simplicity, the delay plane expansion presented in [1] was developed around lumped-constant frequency transfer functions. Distributed systems can be similarly treated, and in [2, Fig. 7(b)] there is an example of this. As a result, the procedure presented in this report can be applied to any causal band-limited transducer, whether characterized as a lumped or a distributed system, and will work for nonminimum-phase as well as for minimum-phase systems. The possibility that a transducer may be nonminimum phase makes it inadvisable to use a Hilbert transform procedure to compute phase from amplitude in the passband. The residual nonminimum-phase all-pass component of the transducer response could only be separated from the all-pass component due to propagation time delay by investigating the asymptotic phase slope at frequencies outside the passband, which is the single procedure we are recommending in this correspondence.

REFERENCES

[1] R. C. Heyser, "The Delay Plane, Objective Analysis of Subjective Properties, Parts I and II," *J. Audio Eng. Soc.*, vol. 21, pp. 690–701 (1973 Nov.); pp. 786–791 (1973 Dec.).

[2] R. C. Heyser, "Loudspeaker Phase Characteristics and Time Delay Distortion, Parts I and II," *J. Audio Eng. Soc.*, vol. 17, p. 30 (1969); p. 130 (1969).

[3] R. C. Heyser, "Determination of Loudspeaker Signal Arrival Times, Parts I–III," *J. Audio Eng. Soc.*, vol. 19, pp. 734–743 (1971 Oct.); pp. 829–834 (1971 Nov.); pp. 902–905 (1971 Dec.).

THE AUTHOR

Richard C. Heyser received a B.S.E.E. degree from the University of Arizona in 1953. Awarded the AIEE Charles LeGeyt Fortescue Fellowship for advanced studies, he received an M.S.E.E. from the California Institute of Technology in 1954. The following two years were spent in postgraduate 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 Motorola Research. From 1956 until the present time he has been associated with the California Institute of Technology Jet Propulsion Laboratory in Pasadena, where he is presently a member of the technical staff in information systems research. His activities at JPL have involved communication and instrumentation design for all major space programs at JPL, commencing with conceptual design on America's first satellite, Explorer I. More recently he has been involved in the application of coherent spread spectrum techniques to improving underwater sound research and medical ultrasound imaging. Mr. Heyser also maintains a personal laboratory where he conducts research on audio and acoustic measurement techniques. This effort has resulted in a number of papers published in the *Journal of the Audio Engineering Society* and elsewhere, and he has been awarded nine patents in the field of audio and communications techniques, including time delay spectrometry.

Mr. Heyser serves as a reviewer for the *Journal* and is a member of the Publication Policy Committee of the AES. As a senior editor of *Audio* magazine, he has been responsible for the loudspeaker reviews of that publication for the past eight years. He is a member and fellow of the Audio Engineering Society and the Acoustical Society of America, and a member of the IEEE and the Hollywood Sapphire Club. In 1983 at the 74th Convention of the AES he was awarded the Silver Medal for the development of time delay spectrometry and its use in the study of loudspeakers and room acoustics. He is also listed in *Who's Who in the West*, *Who's Who in Technology*, and *Distinguished Americans of the West and Southwest*.

Reprinted from *JAES*, vol. 33, no. 6, pp. 463–465, 1985 June.

LETTERS TO THE EDITOR

COMMENTS ON "DETERMINING THE ACOUSTIC POSITION FOR PROPER PHASE RESPONSE OF TRANSDUCERS"*

We agree with the author of the above engineering report[1] that there is some concern about the proper way to remove the unwanted phase term caused by propagation delay from transducer measurements, and in particular from measurements on loudspeaker *systems*. Indeed, we personally raised this very point with Heyser some time ago. We find, however, that the procedure advocated by the author in this report is not well suited for general application. We would like to take this opportunity to explain why this is so and to suggest useful practical alternatives.

To begin with, we believe that a given transducer or loudspeaker system should be treated as a "black box" when submitted for measurement. It should not need to be dismantled or to have some of its subsections disabled in order to be properly measured. Knowledge of the internal design of a loudspeaker system may be helpful in the measurement process (for example, it may tell us what type of phase response to expect on the basis of the crossover network design), but it should not be necessary to determine these aspects by disabling first the woofer and then the tweeter in order to perform a proper phase measurement on the system.

We are in agreement with the author that if the transducer or system is essentially minimum phase over the bandwidth of the measurement, the correct procedure is clear and leads to an unambiguous and correct phase measurement. For, with minimum-phase systems, the phase response is the Hilbert transform of the log magnitude response. If one computes the Hilbert transform of the measured log magnitude (that is, frequency) response, then one has correctly removed the linear phase term due to propagation delay from the phase measurement when the measured phase response agrees with the computed Hilbert transform.

Many individual transducers are of minimum-phase type, but loudspeaker systems, by virtue of their crossover networks, generally have nonminimum-phase response. Some transducers (for example, "whizzer cone" drivers) may be nonminimum phase too. Clearly we cannot pull such a driver apart in order to perform a proper phase measurement on it.

Difficulties thus arise when one wants to measure the phase response of a nonminimum-phase device correctly. Such systems have transfer functions which can be represented as the product

$$H_{\text{min phase}}(s) \cdot H_{\text{all pass}}(s) \cdot H_{\text{time delay}}(s)$$

and the problem is one of "correctly" removing the third factor from the phase measurement on the system. If the all-pass term is known (for example, from the crossover design), one can compute the minimum-phase part (by Hilbert transformation of the measured frequency response), subtract it from the measured phase response, and then adjust the time delay until the remaining excess phase agrees with the expected all-pass part. If the system is of finite order, both the minimum-phase and the all-pass terms have a phase that tends asymptotically to a constant and hence have a phase slope which tends to zero as $\omega \rightarrow \infty$. Only the delay term contributes a nonzero asymptotic slope, if present. So removing delay until the high-frequency asymptotic slope is zero appears at first sight to be a reasonable procedure to adopt. This is what the author suggests, and this is what we wish to demonstrate to be inappropriate, in many cases, because of the complicated behavior of most drivers at high frequencies, and the unusual delay behavior of some common systems.

One practical problem is that most high-frequency transducers have transfer functions with poles and zeros beyond their useful band limit. This may imply measuring unreasonably far above the system's normal passband in order to characterize the phase asymptote in the required manner. This difficulty is pointed out by the author; in practice it may be so severe that the author's suggested procedure is not feasible.

We believe that it is preferable to characterize the transducer on the basis of an in-band measurement, rather than to depend upon indeterminate out-of-band poles and zeros. This approach is assisted by the fact that most individual loudspeaker drivers are minimum phase in their passbands, and the phase effects of loudspeaker crossover networks are largely localized to the vicinity of the crossover frequencies used. Where this is so, and *provided that one first subtracts the computed minimum-phase response*, the author's procedure is feasible and leads to "correct" time delay removal. The high-frequency asymptotic behavior of the excess phase response is then of the expected form.

To illustrate this point, we refer to Fig. 1, which shows the frequency and phase responses measured on a two-way loudspeaker system of no great distinction, which we happen to have available in our laboratory. The bottom phase curve is that measured by dual-channel fast Fourier transform techniques, while the middle phase curve shows the (cyclic) Hilbert transform com-

* Manuscript received 1984 May 7; revised 1985 January 10.

[1] R. C. Heyser, *J. Audio Eng. Soc. (Engineering Reports)*, vol. 32, pp. 23–25 (1984 Jan./Feb.).

puted from the measured log magnitude curve. We have adjusted the propagation delay time offset in the measurement so as to make the "excess phase" (the top phase curve) tend asymptotically to zero slope. The excess phase is the difference between the original data and the computed minimum-phase curve. It will be seen to display the second-order all-pass phase response expected on the basis of the crossover network design. As with most designs, the response of this loudspeaker system is nonminimum phase. The point we wish to make is that the excess phase slope tends to zero, whereas the loudspeaker's measured phase slope *does not*, at the top of the measured frequency band.

Indeed, even if we extend the measurement to higher frequencies, as shown in Fig. 2, the loudspeaker's phase slope does not show any clear asymptotic behavior, and the author's suggested procedure using the asymptotic high-frequency phase slope does not readily enable one to characterize this loudspeaker's phase response correctly. This tweeter's response is quite representative of current designs.

To substantiate our claim that the measurement of Fig. 1 is correct, we present in Fig. 3 the impulse response computed from the Fig. 1 measurement data. Note that the impulse response is maximally causal in that the first nonzero response occurs at time zero. This shows that the correct propagation delay term has been removed from our measurements. The author's phase slope criterion *applied to the excess phase* works well here and leads to a correct measurement because the tweeter itself is minimum phase. (Despite its severe null at 28 kHz, the tweeter's response by itself is minimum phase to 50 kHz.)

This procedure will not necessarily work if the tweeter is not minimum phase. In such a case, or even quite generally, the best procedure seems to be to remove from the measurement that maximum delay which still leaves the system causal. One way of doing this is to adjust the delay until the system's impulse response, computed from its measured complex electroacoustic transfer function, commences at time zero. This can easily be done on many dual-channel fast Fourier transform-type analyzers. This approach is consistent with the author's third to last paragraph, which insists that a correct result must be causal. However, it is important to note that the delay so obtained is not always the same as the delay obtained from a consideration of the high-frequency asymptotic phase behavior.

For example, let us consider the situation envisaged by the author in the last paragraph on p. 23, namely, a two-way system whose woofer is acoustically ahead of its tweeter. In such a case the author's suggested procedure (or indeed the procedure used above based upon the excess phase) results in a phase curve which implies acausal behavior, that is, the woofer signal begins to arrive before time zero. In such a case one should clearly set the time delay such that the correspondence between Hilbert transform and measured phase is obtained at low frequencies (below crossover) and not at high frequencies. The system is then causal: the woofer output still precedes the tweeter output, but now there is no output before $t = 0$.

This situation is illustrated by the measurement of Fig. 4, which is of a two-way system whose woofer is acoustically well ahead of its tweeter (by over 100 mm). The bottom phase curve is the causal one. This is clearly evident in the phase plot of Fig. 5, which has been made on a linear frequency axis. The tweeter's

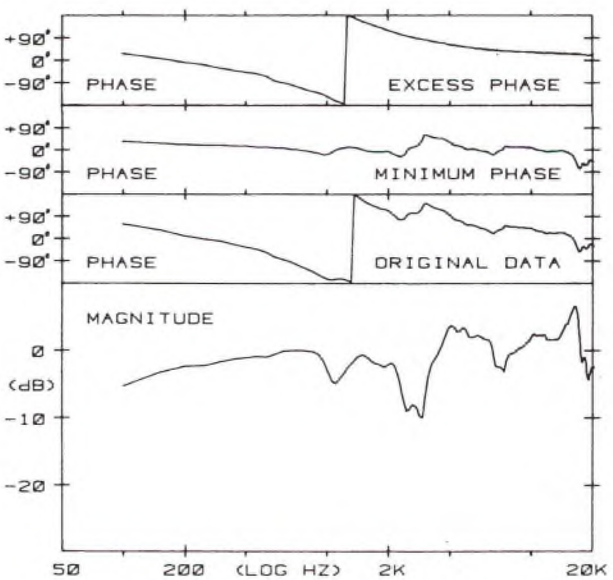


Fig. 1. Magnitude and phase responses of a two-way loudspeaker system.

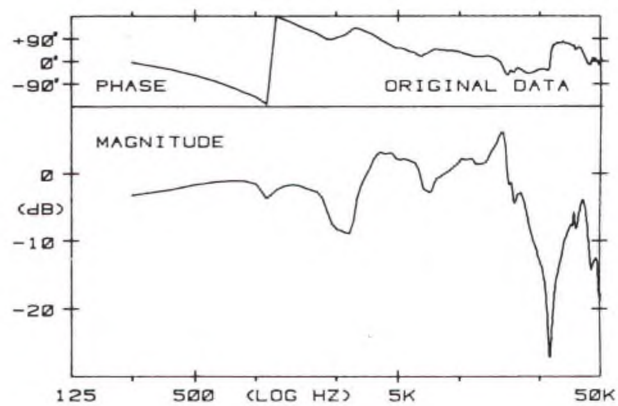


Fig. 2. Frequency response of the system of Fig. 1 extended to 50 kHz.

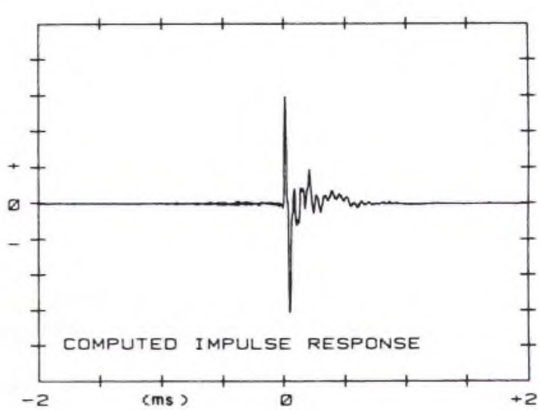


Fig. 3. Impulse response computed from the data of Fig. 1.

asymptotically linear phase lag (delay) passing through the point (0, 0) is well represented relative to the woofer's slope of zero in its passband. The middle phase curve of Fig. 4 is obtained by setting the high-frequency asymptotic phase slope to zero. It is acausal, and does not coincide with the computed Hilbert transform (top phase curve) for this nonminimum-phase system. We maintain that the bottom curve is the correct one.

The author's suggested procedure making use of the limit of the group delay

$$\tau_{g\infty} = \lim_{\omega \to \infty} \left[- \frac{d\phi(\omega)}{d\omega} \right] \tag{1}$$

where ω is the radian frequency and $\phi(\omega)$ the phase response, is reminiscent of Papoulis's concept of the "signal front delay" [1], [2], which is based upon the limit of the phase delay

$$\tau_{p\infty} = \lim_{\omega \to \infty} \left[- \frac{\phi(\omega)}{\omega} \right] . \tag{2}$$

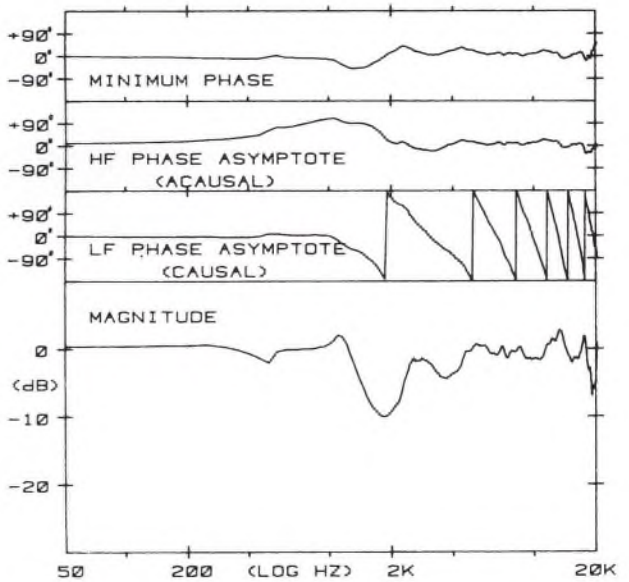


Fig. 4. Magnitude and phase responses of a different two-way loudspeaker system whose woofer is well ahead of its tweeter.

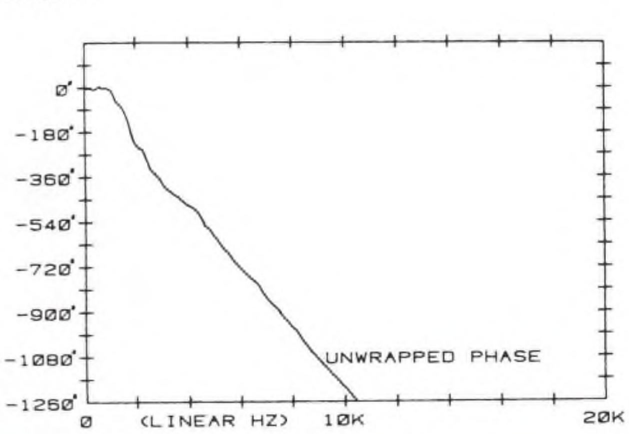


Fig. 5. Unwrapped phase of the causal measurement from Fig. 4 on a linear frequency axis.

In fact, since the existence of the former limit implies, under any reasonable conditions on $\phi(\omega)$, the existence of the latter limit too and their equality (i.e., if $\tau_{g\infty}$ exists, so does $\tau_{p\infty}$ and $\tau_{p\infty} = \tau_{g\infty}$), it follows that the author's asymptotic phase slope $\tau_{g\infty}$, if it exists, is equal to Papoulis's signal front delay $\tau_{p\infty}$. The trouble is that Papoulis's theorem [1, p. 136], purporting to show that the signal front delay $\tau_{p\infty}$ is the time at which the system's impulse response begins, is false—otherwise the middle phase curve of Fig. 4 would have corresponded to an impulse response starting at time zero. The problem arises when Papoulis takes a true result concerning the complex Laplace transform [1, sec. 9.5, p. 187] or [2, sec. 7.1, p. 225] and misapplies it without proof to the signal front delay $\tau_{p\infty}$. The signal front delay $\tau_{p\infty}$ as defined by Papoulis, or the asymptotic value of the group delay $\tau_{g\infty}$ as used by the author, does not always coincide with the first arrival signal. This is the reason that a criterion based solely upon $\tau_{g\infty}$ cannot always yield the correct result.

To summarize, if the system under measurement is minimum phase, the author's procedure will work if applied to the *excess phase*. In other cases, which represent the vast majority of loudspeaker system measurements, even this procedure has been shown not always to yield correct results, especially in the absence of further knowledge of the system's makeup. The procedure we recommend is to remove the maximum delay which leaves the system's impulse response (computed from the necessarily band-limited transfer function measurement) causal. If some system information is available, one can trim this delay adjustment on the basis of the in-band high- or low-frequency correspondence between the measured phase response and the computed minimum-phase response.

ACKNOWLEDGMENT

The authors would like to thank the reviewer for helpful comments.

STANLEY P. LIPSHITZ
JOHN VANDERKOOY
University of Waterloo
Waterloo, Ontario N2L 3G1, Canada

REFERENCES

[1] A. Papoulis, *The Fourier Integral and Its Applications* (McGraw-Hill, New York, 1962), pp. 134–136, 187.

[2] A. Papoulis, *Signal Analysis* (McGraw-Hill, New York, 1977), pp. 224–225.

Author's Reply

I feel that it is necessary to pass along a sage bit of advice which was directed at me and my activities many years ago: "When I tell you not to do something because to do so, might get you into trouble and you go ahead and do it anyway—and get into trouble—don't come back and blame me." I made it clear, or thought I did, that I am describing a transducer, not a transducer system. To quote my report, "When multiple transducer systems are involved, such as woofer, midrange, and tweeter assemblies with their attendant crossover networks, it may be necessary to measure the acoustic position of each subsystem separately. . . ." I also warned, to continue the quote, "otherwise, a low-frequency driver that is positioned well in front of a higher frequency driver may have its advanced position masked by the larger response of the higher frequency driver at very high frequencies. . . ." You measured such a combined system in Fig. 4 and found that the high-frequency phase in no way told where the lower frequency driver was. I am not surprised; I predicted it.

I also cautioned that ". . . the measurement may need to be performed at signal levels as low as 60 decibels below that which might otherwise be obtained in the transducer passband. Even then, the proper value of asymptotic phase may not be closely achieved if the amplitude cutoff is very rapid. . . ." Based on this warning, I am not at all surprised that the phase slope of Fig. 2 does not show any clear asymptotic behavior at 50 kHz, where the amplitude response is only 20 dB down and falling rapidly.

I suggest that when computer facilities are not available, an engineer can determine the impulse response, and from it the moment of earliest signal arrival, by driving the transducer with a sufficiently narrow pulse and measuring the microphone response with a calibrated delay trigger oscilloscope.

With the foregoing exceptions, I believe that your comments on this matter are valuable and should be carefully considered by engineers who measure multiple transducer systems. I would only add the caution that I also presented in my report: "This is where common sense and a complete measurement of amplitude and phase, corrected for time delay, can be helpful."

RICHARD C. HEYSER
Tujunga, CA 91042, USA

INSTANTANEOUS INTENSITY

2399 (E-9)

Richard C. Heyser
California Institute of Technology
Jet Propulsion Laboratory
Pasadena, California

**Presented at
the 81st Convention
1986 November 12-16
Los Angeles, California**

AES

AN AUDIO ENGINEERING SOCIETY PREPRINT

INSTANTANEOUS INTENSITY

Richard C. Heyser
California Institute of Technology Jet Propulsion Laboratory

ABSTRACT

Using an energy theorem first published in the Journal in 1971, and reconsidering sound pressure and particle velocity as components of a four-dimensional space-time vector, results in the prediction of a complex intensity whose components are related through the Hilbert operator. This intensity vector has an instantaneous time dependence. The conventional time average of the product of sound pressure and particle velocity gives incorrect answers for anything but a steady tone and is not much use in dynamic sound measurements. The energy-time curve (ETC) is a valid measure of instantaneous intensity.

INTRODUCTION

Although the subject matter of this paper is sound intensity, there is actually a much deeper aspect, an aspect that changes much of our modern approach to analysis. This paper is extracted from one small part of a much larger paper which I am preparing for presentation in the Journal of the Audio Engineering Society. The subject of that larger paper is the fundamental theory behind time delay spectrometry. Time delay spectrometry is not, as some persons believe, merely the use of a sweeping oscillator and a delay tracking filter. Time delay spectrometry is the implementation of a concept of mapping among a great many domains by the use of a specific class of integral transform, one special case of which is the Fourier transform. It is the theory behind TDS, the theory that nature does not have a preferred frame of reference, which forces us to look more carefully at our modern approach to analysis.

There is no nice way to say this, so I might as well be blunt: our personal human prejudice has caused us to do some pretty silly things, some of these things leading to error. I won't go into details, that is covered in the larger paper, but one of the most visible manifestations of this prejudice is our avoidance of complex numbers for considering the "real" world. This prejudice is not a recent phenomenon, but goes back to the beginnings of what might euphemistically be called "modern science". It may come as a surprise to some persons to realize that it is only within the past few centuries that mathematicians finally accepted irrational numbers. The acceptance of zero as a number and the acceptance of negative numbers as anything but absurdities did not come easy, or all that long ago. And the new type of number which was spawned from negative numbers, complex numbers, were not fully accepted until WELL AFTER DEVELOPMENT OF THE MATHEMATICS WHICH WE NOW USE TO ANALYZE SOUND!

We now know that all the previous concepts of number, such as integer, rational, irrational, zero and negative numbers, together with the full set of fundamental rules for combining them, are contained within the realm of complex numbers. They are all special cases, including the so-called "real" and "imaginary" components. The very name, "imaginary number", is a fossil remain of an early prejudice that hangs on even to this day.

- 1 -

It was Rene Descartes, one of the most liberal free-thinkers of his day, who made fun of persons who used such numbers. He claimed that these were images of the mind and hence imaginary. He was wrong, but the name, the stigma, remains. In this manifestation of a prejudice of modern science, I regard the term "imaginary" as the technological equivalent of a racial slur which must be struck from our language.

SOUND INTENSITY

Sound is what happens when air gets pushed. Pushing on air, whether by a loudspeaker cone or any other means, does work on the air because air pushes back as it is forced to move. The energy to move the air comes from the source of sound. Once in motion, the air nearest the source of sound must itself move air which is farther away from that source. And so it proceeds, as ever expanding pushing motions being transmitted from certain regions of the air to more distant regions. Sound.

The work that was done by the source of sound is passed outward and away from that source by the motion of the air. In its passage, the pushing-shoving motions of the air contain the energy which was initially imparted from the source, less any amount which may have been converted to heat due to viscosity in the air through which it passes. In most of those cases of interest to audio engineers, the amount of energy lost to heat up the air is much less than the total amount contained in the sound.

So sound contains energy. If nothing is there to convert that energy to heat or work, then the energy simply passes and, once past, the local air returns to its normal equilibrium state with its neighbors.

The sound energy transmitted per unit of time through a unit area of space is called the SOUND INTENSITY of the sound wave. Sound intensity is power flow expressed in Watts per square meter. Sound intensity is a "now" thing, having an instant by instant meaning at each place in the sound field.

Sound intensity, sound power flow in Watts per square meter, is a "real world" concept of tremendous practical benefit. By measuring three-dimensional sound intensity we can tell where the sound is coming from, where it is going, and how much of it is present at each location in between. Dr. Harry F. Olson introduced the concept to audio engineering, although the apparatus available to him in those early days was less than adequate for the task [1]. The more recent introduction of new signal processing methods has led to renewed interest in sound intensity and several excellent instruments are presently available to audio engineers.

There is one small problem. It is clear from reading Dr. Olson's 1932 patent that he was contemplating measurement of the moment by moment intensity of the sound, its instantaneous intensity. In the main, contemporary instruments do NOT measure instantaneous intensity; they measure the intensity in bands of frequency. Furthermore, if we really pull the mathematics apart, piece by piece, we find the prejudices of old intruding upon considerations of the new.

- 2 -

What I will now do is present two mathematical derivations of sound intensity. The first derivation is that of contemporary analysis. The second derivation is based upon considerations that are forced upon us by the more general transformation theory that underlies TDS. In the frequency domain there will be no difference because the prejudice against complex numbers does not exist within the frequency domain. The time domain, however, suffers a substantial difference because it is there where the full wrath of a prejudice against "imaginary" numbers is felt.

I have two additional comments to make prior to the analysis. First, obtain a copy of S. Gade's excellent article in the Bruel and Kjaer Technical Review and read it. This is the most lucid and straightforward discussion of contemporary sound intensity theory which I have seen. Second, those hundreds of engineers who now utilize ETC measurements in the analysis of sound should take heart. For free field sound, the ETC is a direct measure of instantaneous intensity. You have been measuring total energy density, and hence instantaneous intensity (minus direction of the sound), all along.

CONTEMPORARY SMALL SIGNAL ACOUSTICS [2],[3],[4],[5]

Small signal acoustics operates under the assumption that air is a fluid which is vortex free and whose viscosity can be ignored. This is a necessary and sufficient condition for the existence of a velocity potential whose gradient gives the point by point value of the small particle velocity of that fluid. Since the motion is assumed to be exceedingly small, we can ignore nonlinearities in the fluid motion and the local change of fluid density, its condensation, can be related to a more familiar parameter, sound pressure. If Ψ is the velocity potential, $\vec{u}$ is the particle velocity, p is pressure and ρ_3 is the equilibrium density of air, then,

$$\vec{u} = -\text{grad}\,\Psi = -\sum_{j=1}^{3} \vec{n}_j \frac{\partial \Psi}{\partial x_j}\ , \quad p = \rho \frac{\partial \Psi}{\partial t} \qquad (1)$$

where $\vec{n}_j$ is a unit vector in the j[th] direction.

For a freely advancing sound wave at great distance from its source, pressure and particle velocity are precisely in phase (scalar multiples of each other) under the relation,

$$p = \rho c \vec{u} \qquad (2)$$

where c is the speed of sound.

The classic energy density of the sound wave is expressed as,

$$E = \frac{1}{2}\rho(\vec{u})^2 + \frac{1}{2}\frac{p^2}{\rho c^2} \qquad (3)$$

The first term is identified with kinetic energy density while the second term is identified with potential energy density.

When condition (2) prevails, this energy density can be simplified to,

$$E = \rho(\vec{u})^2 = \rho\frac{p^2}{(\rho c)^2} = \frac{1}{c}(\rho c)(\vec{u})^2 = \frac{1}{c}\frac{p^2}{(\rho c)} \qquad (4)$$

Energy density is expressed in Watt-seconds per cubic meter. Sound speed is expressed in meters per second. Intensity I is power density expressed in Watts per square meter under the relation,

$$E = I/c \qquad (5)$$

Classic sound intensity is thus,

$$I = (\rho c)(\vec{u})^2 = \frac{p^2}{(\rho c)} = p\vec{u} \qquad (6)$$

It is traditional to evaluate the active work producing component of intensity as the TIME AVERAGE of the product of pressure and particle velocity,

$$\bar{I} = \overline{p\vec{u}} = \frac{1}{2T}\int_{-T}^{T} p(t)\,\vec{u}(t)\,dt \qquad (7)$$

where the time interval T is chosen to be of the order of the period of the lowest usable frequency component. For a sinusoidal pressure time dependence, particle velocity will be in phase with the sound pressure for a free-field plane wave. In that special case the time average sound intensity becomes,

$$\bar{I} = \frac{1}{2N\pi}\int_0^{2N\pi}(p_0 \sin \omega_0 t)(u_0 \sin \omega_0 t)\,dt$$

$$= \frac{1}{2N\pi}\int_0^{2N\pi} p_0 u_0 \left(\frac{1-\cos 2\omega_0 t}{2}\right)dt = \frac{p_0 u_0}{2} \qquad (8)$$

Where reflections exist, particle velocity will no longer be in phase with sound pressure and the time average value of intensity will diminish. When particle velocity is ninety degrees out of phase with sound pressure, as it will be for sound directly in front of a perfectly reflecting boundary, the time average intensity is zero, under,

$$\bar{I} = \frac{1}{2N\pi}\int_0^{2N\pi}(p_0 \sin \omega_0 t)(u_0 \cos \omega_0 t)\,dt$$

$$= \frac{1}{2N\pi}\int_0^{2N\pi}\frac{\sin 2\omega_0 t}{2}\,dt = 0 \qquad (9)$$

The interpretation is that the intensity of (6) consists of an active component (inphase pressure and particle velocity) and reactive component (quadrature pressure and particle velocity). FOR THE SPECIAL CASE of sinusoidal time dependence of pressure, it is found convenient to resort to complex notation in which,

$$p = p_0 e^{i\omega_0 t}$$

$$u = u_0 e^{i(\omega_0 t + \phi)} \qquad (10)$$

Time average active intensity is then,

$$\bar{I} = I_A = \frac{1}{2}\mathcal{Re}(pu^*) \tag{11}$$

and time average reactive intensity is

$$I_R = \frac{1}{2}\mathcal{Im}(pu^*) \tag{12}$$

where $\mathcal{Re}$ and $\mathcal{Im}$ stand for "real part of" and "imaginary part of", respectively, and * represents complex conjugation.

I want to point out two very important things. First, A TIME AVERAGE REMOVES ALL TIME DEPENDENCE OVER THE INTERVAL OF INTEGRATION. We lose all moment to moment information about what happens. What is normally done at this place in the analysis is to show what happens when we transform over to the frequency domain. That is fine, because we must lose all time dependence when we transform to frequency. The frequency domain nicely has all the necessary real and imaginary numbers to present active and reactive intensity. The problem is the time domain. The closest thing we have to instantaneous intensity is relation (6). I claim that relation (6) is incomplete because it uses "real" numbers only.

Second, the complex representations of (10), (11) and (12) are correct, but only apply to TIME AVERAGE POWER FLOW FOR A PERIODIC MONOCHROMATIC SOUND FIELD. They do not apply to an arbitrary non periodic sound field.

INSTANTANEOUS INTENSITY

What I will now do is rederive the acoustic intensity expression from a first principle which DEMANDS the use of complex numbers. I am not patching up an existing theory. I am developing a new approach which is based upon the principles from which TDS was developed.

As before, there is a velocity potential. But now we must consider a space-time representation of an acoustic wave in terms of a four-dimensional set of coordinates x_1, x_2, x_3 and x_4. The first three: x_1, x_2 and x_3, are the usual orthogonal coordinates of "space". The fourth coordinate is,

$$x_4 = ict \tag{13}$$

where $i = \sqrt{-1}$ is a unit normal to all three space directions and c is a constant to be determined. It will, of course, be the speed of sound under free field conditions.

This gives us a four-dimensional observable

$$\phi(x) = -\mathrm{grad}\,\psi = -\sum_{j=1}^{4} \vec{\eta}_j \frac{\partial \psi}{\partial x_j}$$

$$= -\sum_{j=1}^{3} \vec{\eta}_j \frac{\partial \psi}{\partial x_j} + \frac{i}{c}\frac{\partial \psi}{\partial t} = \vec{u} + \frac{ip}{\rho c} \tag{14}$$

Under conditions where the denominator does not vanish, the factor ρc can be identified from the ratio of time to space gradients as,

$$\frac{p}{u} = \frac{\rho\left(\frac{\partial\psi}{\partial t}\right)}{-\left(\frac{\partial\psi}{\partial x}\right)} = \rho\frac{dx}{dt} = \rho c \tag{15}$$

from which c is the speed dx/dt.

It is a requirement (not obvious here but developed from the theory as presented in part in [6]) that $\phi(x)$ be the real part of a complex $\Theta(x)$, under,

$$\Theta(x) = \phi(x) + i\hat{\phi}(x) \tag{16}$$

where $\hat{\phi}(x)$ is the Hilbert transform of $\phi(x)$. Thus,

$$\Theta = (u + i\hat{u}) + \frac{i(p + i\hat{p})}{\rho c} \triangleq [u] + \frac{i[p]}{\rho c} \tag{17}$$

The total energy density as a function of time at a particular position of space becomes the positive real entity,

$$E(x) = K\cdot\Theta(x)\,\Theta^*(x) \tag{18}$$

where $K = \frac{1}{4}\rho$. Expanding this,

$$E(x) = \frac{1}{2}\rho\frac{(u^2+\hat{u}^2)}{2} + \frac{1}{2}\left(\frac{1}{\rho c^2}\right)\frac{(p^2+\hat{p}^2)}{2} - \frac{(\hat{p}u - p\hat{u})}{2c} \tag{19}$$

Comparing this with relation (3) we see that the first two terms are the analytic signal equivalent of classic kinetic energy density and potential energy density, respectively. But now a new term appears, a cross product term. This cross product term can be identified as the "imaginary" component of a complex intensity,

$$I = \frac{1}{2}[p][u]^* = \frac{1}{2}(p+i\hat{p})(u-i\hat{u})$$

$$= \frac{(pu+\hat{p}\hat{u})}{2} + i\frac{(\hat{p}u-p\hat{u})}{2} = I_A + iI_R$$

$$= |I|(\cos\alpha + i\sin\alpha)\quad,\ \alpha = \tan^{-1}\frac{[p]}{[u]} \tag{20}$$

The intensity expression of (20), which is DERIVED from the energy terms, is what I shall call INSTANTANEOUS INTENSITY. The expression for total energy density can be expressed as,

$$E(x) = (1/2)\,\rho\frac{[u][u]^*}{2} + (1/2)\frac{1}{\rho c^2}\frac{[p][p]^*}{2} - \frac{I_R}{c} \tag{21}$$

where the [ ]'s indicate the complete complex term, impulse plus Hilbert transform doublet component; not just the real-only term as is used in the contemporary expression (3).

It is a property of the Hilbert transform that dual application produces negation such that,

$$\hat{\hat{p}} = -p \tag{22}$$

It is also a property of the Hilbert transform that if f and g are of Class L^2, then [7],

$$\int_{-\infty}^{\infty} f\,g\,dx = \int_{-\infty}^{\infty} \hat{f}\,\hat{g}\,dx \tag{23}$$

From these we can derive the relationship,

$$\int_{-\infty}^{\infty} \hat{f}\, g\, dx \;=\; -\int_{-\infty}^{\infty} f\, \hat{g}\, dx \tag{24}$$

Instantaneous active intensity is,

$$I_A \;=\; (1/2)(p\,u) \;+\; (1/2)(\hat{p}\,\hat{u}) \tag{25}$$

The time average of instantaneous active intensity is, from (23),

$$\overline{I_A} \;=\; \lim_{T\to\infty} \frac{1}{2T}\int_{-T}^{T} I_A\, dt \;=\; (1/2)\overline{(p\,u)} + (1/2)\overline{(\hat{p}\,\hat{u})} = \overline{p\,u} \tag{26}$$

Instantaneous reactive intensity is,

$$I_R \;=\; (1/2)(\hat{p}\,u) - (1/2)(p\,\hat{u}) \tag{27}$$

The time average of instantaneous reactive intensity is, from (24),

$$\overline{I_R} \;=\; (1/2)\overline{(\hat{p}\,u)} - (1/2)\overline{(p\,\hat{u})} \;=\; \overline{\hat{p}\,u} \tag{28}$$

What this means, quite simply, is that the expression,

$$I \;=\; p\,u \tag{29}$$

is not correct; and the expression for time average of intensity,

$$\overline{I} \;=\; \overline{p\,u} \tag{30}$$

is correct only so long as reactive intensity is zero and all time dependence is removed.

If complex pressure [p] is a scalar multiple of the complex particle velocity [u],

$$(p + i\hat{p}) \;=\; \rho c(u + i\hat{u}) \tag{31}$$

then the instantaneous reactive intensity vanishes under,

$$I_R = \frac{\hat{p}\,u - p\,\hat{u}}{2} = \rho c\frac{(\hat{u}\,u)}{2} - \rho c\frac{(u\,\hat{u})}{2} \;=\; 0 \tag{32}$$

This is a general signal relationship.[2] For the special kind of signal called a "sine wave", it corresponds precisely to the classic case of sound pressure being in phase with particle velocity.

When instantaneous reactive intensity vanishes in this manner the total energy density expression of (21) reduces to,

$$E \;=\; \frac{I_A}{c} \tag{33}$$

In the most general case, sound pressure will not be a scalar multiple of particle velocity, but will have a complex ratio,

$$(p + i\hat{p}) \;=\; \rho\, c_o(r_a + i\,x_a)(u + i\hat{u}) \tag{34}$$

The entity $(r_a + i\,x_a)$ is what I shall call specific acoustic wave impedance, where r_a is specific acoustic wave resistance and x_a is specific acoustic wave reactance. The entity c_o is the speed of sound in free space and ρ is the density.

For many practical audio measurements, we can use the complex inverse of this, the specific acoustic wave admittance, expressed as,

$$(r_a + i\,x_a)^{-1} \;=\; (g_a + i\,b_a) \tag{35}$$

where g_a is specific acoustic wave conductance and b_a is specific acoustic wave susceptance.

Going back to (14), we see that there are NOT "two kinds" of acoustic properties - particle velocity and sound pressure - there is "one kind" which expresses the space-time gradients of a four-dimensional potential. Particle velocity is the space gradient while pressure (divided by ρ c) is the time gradient. This means that we can obtain particle velocity from sound pressure by noting the fundamental forms,

$$u = -\frac{\partial\psi}{\partial x}\quad,\quad p = \rho\frac{\partial\psi}{\partial t} \tag{36}$$

or

$$\psi = \int \frac{p}{\rho}\, dt = -\int u\, dx \tag{37}$$

so that,

$$u = -\frac{\partial}{\partial x}\left(\int \frac{p(t)}{\rho} dt\right) \tag{38}$$

and

$$p = -\rho\frac{\partial}{\partial t}\left(\int u(x)\, dx\right) \tag{39}$$

Note that relation (39) comes directly from our four-dimensional expression (14). We do not have to know Euler's relation, the acoustic equivalent of Newton's Second Law that force equals mass times acceleration. In fact, if we look at relation (39) we see that the theory predicts that pressure gradient equals density times the time rate of change of particle velocity.

The equivalent of relation (6) for instantaneous intensity under general conditions is,

$$\begin{aligned} I &= \rho c_o(r_a + i x_a)\frac{(u^2+\hat{u}^2)}{2} = \frac{(p^2+\hat{p}^2)}{2}\frac{(g_a - i b_a)}{\rho c_o} \\ &= \frac{(p^2+\hat{p}^2)(r_a + i x_a)}{2\rho c_o(r_a^2+x_a^2)} = I_A + i I_R \end{aligned} \tag{40}$$

For free field conditions these reduce to,

$$I = \rho c_o\frac{(u^2+\hat{u}^2)}{2} = \frac{(p^2+\hat{p}^2)}{2\rho c_o} = \frac{pu + \hat{p}\hat{u}}{2} = I_A \tag{41}$$

Useful additional relationships are,

$$\frac{(p^2+\hat{p}^2)}{(u^2+\hat{u}^2)} = (\rho c)^2(r_a^2 + x_a^2) = \frac{(\rho c)^2}{(g_a^2+b_a^2)} \tag{42}$$

and

$$|I| = \sqrt{I_A^2 + I_R^2} = \frac{(p^2+\hat{p}^2)}{2\rho c}\sqrt{g_a^2+b_a^2} = \frac{(p^2+\hat{p}^2)\cdot 1}{2\rho c\sqrt{r_a^2+x_a^2}} \tag{43}$$

ETC

When the reactive field vanishes, total energy density, active intensity and sound pressure are related from (19) and (33) by,

$$\frac{p^2+\hat{p}^2}{2} = \rho c I = \rho c^2 E \qquad (44)$$

Thus the sum of the squares of the sound pressure (i.e. impulse response squared) and of the Hilbert transform of that sound pressure (i.e. doublet response squared) is proportional to total energy density and also to active intensity. The conventional ETC now used by acoustical engineers, which expresses the logarithm of the magnitude of the full sound pressure, as above, is DIRECTLY PROPORTIONAL TO TRUE TOTAL ENERGY DENSITY OF THE SOUND, as has been stated all along, and, in addition, is a DIRECT EXPRESSION OF INSTANTANEOUS INTENSITY FOR A FREE FIELD SOUND WAVE.

Instantaneous sound intensity has an instantaneous space direction as a function of time. If we use a single pressure microphone in a single unchanging position, as is conventional practice with the ETC, then we do not measure the direction of power flow, but we do have a measure of the total value of energy density passing that microphone. We can get spatial direction by taking three orthogonal space gradients of sound pressure.

The ETC is measured in decibels versus time. Decibels of what? Decibels of sound intensity in watts per square meter. If the decibel value of the ETC is equal to the associated decibel value of the sound pressure level, as now measured with sound level meters, and if we measure in air at standard conditions, then zero decibels on the ETC corresponds to 10^{-16} Watts per square centimeter, or one PicoWatt per square meter. This is because the basic reference unit of 0 dB for SPL is NOT a unit of pressure, but is a unit of INTENSITY. That is why 0 dB SPL is 20.4 MicroPascals, instead of some pressure-only basis such as one MicroPascal which is commonly used in underwater sound.

A RECOMMENDATION ON UNITS OF SOUND INTENSITY

The real hero in this saga of sound intensity, the person who started it all, is the late Dr. Harry F. Olson. After more than fifty years, we do not yet have a unit of sound intensity. Instead, sound power flow is expressed as the ratio of units, namely Watts per square meter.

In recognition of his pioneering work in this field, I recommend that we adopt a unit of sound intensity and that this unit be called the Olson, signified by the single capital letter O. We already have a fundamental basis for this proposed unit, the intensity level from which our present sound pressure levels are defined. If one Olson is defined as an intensity level of one PicoWatt per square meter, then for air, at standard conditions, ALL NUMBERS STAY THE SAME. That is, 74 dB SPL is also 74 dB for intensity in Olsons. I further suggest that we might adopt the designator dBO for decibels of sound intensity relative to one PicoWatt per square meter.

This has the further advantage that the same numeric value of dBO represents the same sound intensity, regardless of temperature, material density or speed of sound. As sound passes from one medium to another, the dBO stays constant if the power flow is constant, whereas dB SPL may change drastically.

Engineers long ago adopted dB notation because of the ease it imparts to signal analysis. We rarely try to figure network gains and losses in terms of amps or volts, taking into account impedance transformations. So why persist in the same thing with acoustic units? We can add and subtract dBO's, which will automatically take acoustic impedance into account, and enormously simplify our acoustic calculations.

The Olson is a unit which acoustic and audio engineering needs.

CONCLUSION

Instantaneous intensity is NOT

$$p\,u \qquad (45a)$$

but IS

$$\frac{p\,u}{2} + \frac{\hat{p}\,\hat{u}}{2} + i\left(\frac{\hat{p}\,u}{2} - \frac{p\,\hat{u}}{2}\right) \qquad (45b)$$

The error in using (45a) is both quantitative and philosphical. Quantitative in that (45a) gives a result which passes through zero when either pressure or particle velocity is passing through zero. Philosophical in that, in an attempt to avoid complex numbers, the resultant cannot explain the existence of reactive power.

It is the Hilbert transform (actually, it is an <u>operator</u>, not a <u>transform</u>; but that is left to the larger paper) which fills in the oscillatory gaps and which accounts for the reactive power term.

In the very commonplace circumstance in which particle velocity is essentially a scalar multiple of sound pressure,

$$p = \rho c\, u \qquad (46)$$

then instantaneous sound intensity is NOT

$$\rho c\, u^2 = \frac{p^2}{\rho c} \qquad (47a)$$

but IS

$$\rho c \frac{(u^2 + \hat{u}^2)}{2} = \frac{(p^2 + \hat{p}^2)}{2\rho c} \qquad (47b)$$

In this commonplace case, the dB magnitude of instantaneous sound intensity, relative to a reference level of

$$0 \text{ dB} = 10^{-12} \text{ Watt/(meter)}^2$$

is IDENTICAL to the ETC. The distinction between the ETC and instantaneous sound intensity is that the energy-time curve is an expression of TOTAL ENERGY DENSITY, whereas instantaneous sound intensity is an expression of POWER DENSITY FLOW and thus has a space direction as well as a magnitude.

REFERENCES

[1] Olson, H.F., "System Responsive to the Energy Flow of Sound Waves", U.S. Patent No. 1,892,644 (1932)

[2] Gade, S.,"Sound Intensity (Part I Theory)", Bruel & Kjaer Technical Review, No. 3 (1982)

[3] Gade, S., "Sound Intensity and its Application in Noise Control", Sound and Vibration, pg 14-26, March 1985

[4] Pascal, J.C., "Structure and Patterns of Acoustic Intensity Fields", Recueil de Conferences, 2nd International Congress on Acoustic Intensity, CETIM, pg 97-104 (1985)

[5] Elko, G.E., "Frequency Domain Estimation of the Complex Acoustic Intensity and Acoustic Energy Density", Ph.D. Thesis, The Pennsylvania State University (1984)

[6] Heyser, R.C., "Determination of Loudspeaker Signal Arrival Times", J. Audio Eng. Soc., pg 902 (1971)

[7] Titchmarsh, E.C., Fourier Analysis, (Oxford University Press, London, 1959) pg 138

SPECIFIC ACOUSTIC WAVE ADMITTANCE

2400 (E-10)

Richard C. Heyser
California Institute of Technology
Jet Propulsion Laboratory
Pasadena, California

**Presented at
the 81st Convention
1986 November 12-16
Los Angeles, California**

AN AUDIO ENGINEERING SOCIETY PREPRINT

SPECIFIC ACOUSTIC WAVE ADMITTANCE

Richard C. Heyser
California Institute of Technology Jet Propulsion Laboratory

ABSTRACT

The relationship of active to reactive instantaneous intensity of a sound wave is expressed by its specific acoustic admittance. This reveals important features of a sound wave under measurement. For example, the direction to a source, the distance from that source, whether the source is a monopole or dipole radiator, and whether the sound under measurement is free field or reverberant, and by how much, is contained in the specific acoustic admittance. Preliminary measurements utilizing time delay spectrometry (TDS) are presented and discussed.

INTRODUCTION

This is a companion paper to "Instantaneous Intensity". I will draw upon some of the concepts which are presented in that paper. But, whereas "Instantaneous Intensity" is written to present the concept of instantaneous time domain power flow, this paper will consider the concept of power flow in both the frequency domain and the time domain. As in "Instantaneous Intensity", this paper will draw from another larger paper in progress on the theory of time delay spectrometry (TDS).

FOUR DIMENSIONAL SOUND

In "Instantaneous Intensity" I present a four-dimensional space-time representation of an acoustic signal, using the coordinates x_1, x_2, x_3 and x_4, where x_1, x_2 and x_3 are the usual coordinates of "space" and x_4 is a fourth orthogonal space coordinate,

$$x_4 = ict. \tag{1}$$

This gives a four-dimensional observable which is obtained from the four-dimensional gradient of a velocity potential. Particle velocity along the x_1 direction is obtained as the partial derivative of this potential with respect to axis x_1. Similarly for particle velocity along x_2 and x_3. Sound pressure, divided by a factor ρc, is obtained as the partial derivative of this potential with respect to axis x_4. In other words, particle velocity is the "space gradient" while sound pressure (divided by ρ c) is the "time gradient" of the SAME POTENTIAL.

The strange factor of ρ c is the link between "space-type" and "time-type" behavior of the sound which is characterized from particle velocity and sound pressure. We can identify this factor by taking the ratio of "time-type" to "space-type" as follows,

$$\frac{P}{u_i} = \frac{\rho\left(\frac{\partial\psi}{\partial t}\right)}{-\left(\frac{\partial\psi}{\partial x_i}\right)} = \rho\left(\frac{dx_i}{dt}\right) = \rho c_i \tag{2}$$

where ψ is the potential. We see that ρ is the equilibrium density of the air and that c is some kind of speed. The factor ρc is a characteristic property of the sound. It is this factor which I will investigate in this paper.

The seemingly weird four-dimensional representation, which smacks of the metric of special relativity, is something which makes sense when we stop to think about it. Both the sound pressure and particle velocity are departures from equilibrium conditions in the air. If we impart a momentary impulse of energy at a position of space and moment of time, then that energy cannot simply stay there. We can poke our finger in sand and then remove our finger, and a hole stays in the sand. But air cannot do that. The hole of air must be filled in order to restore equilibrium, and in being filled, the hole must essentially move outward. The relation between how far it goes in space and how far it goes in time is governed by that strange four-dimensional relationship. In order to follow the energy which we put in, we must follow the sound in the four dimensions of space and time.

If we go (and wait) far enough, then the relationship (2) will approach a fixed number. Right at the beginning, at the shorter distances in space and time from our "finger poke", the entity ρc will not be simple. But far enough out it will approach a constant which I shall designate as,

$$\rho c_o \tag{3}$$

It is a result of the theory presented in [1] that sound pressure at a point in space and moment in time MUST have the form,

$$p(x) + i\hat{p}(x) \triangleq [p(x)] \tag{4}$$

where $\hat{p}(x)$ is the Hilbert transform of p(x) and $i = \sqrt{-1}$. Particle velocity MUST have the form,

$$\vec{u}(x) + i\hat{\vec{u}}(x) = [\vec{u}(x)] \tag{5}$$

where I place an arrow over each component to signify that it is a space vector having three space coordinates in the x_1, x_2 and x_3 directions.

In general, from relation (2), we will have a complex ratio of pressure and particle velocity as defined by,

$$\frac{p + i\hat{p}}{\vec{u} + i\hat{\vec{u}}} = \rho c_o(\vec{r}_a + i\vec{x}_a) = \rho c_o(\vec{r}_a + i\hat{\vec{r}}_a) \tag{6}$$

The inverse ratio is defined as,

$$(\vec{r}_a + i\vec{x}_a)^{-1} = \vec{g}_a + i\vec{b}_a = \vec{g}_a + i\hat{\vec{g}}_a \tag{7}$$

I haven't named these things yet. Let us do it now.

sawa and sawi

Equation (6) defines what I shall call ACOUSTIC WAVE IMPEDANCE. This is my notation. The ratio of sound pressure to particle velocity is often called specific acoustic impedance. I do not like that notation, since the word "specific" is normally used to mean that something is referenced relative to an agreed standard, either of the same kind, or per unit measure, such as per square centimeter. The specific gravity of water, for example, is unity since water is the standard for liquid and solids. I want to define a specific wave impedance such that it is unity if some agreed upon condition exists. I therefore incorporate the word "wave" in my designation in order to separate my unitless ratio from the conventional designation of the entire ρc.

We are discussing waves of sound. If the wave of sound is passing unhindered through space, that is, it is free-field, then I want this condition to be a reference standard against which all other wave motion is to be compared. At very great distance and time from a source of sound, the acoustic wave impedance approaches ρc_o. The ratio of actual acoustic wave impedance to ρc_o is what I shall call SPECIFIC ACOUSTIC WAVE IMPEDANCE (sawi). I use lower case letters for the acronym because we are considering the time domain. Sawi is unity if we have distant free field conditions. Sawi is thus $(\vec{r}_a + i\vec{x}_a)$. Its inverse, $(\vec{g}_a + i\vec{b}_a)$, is SPECIFIC ACOUSTIC WAVE ADMITTANCE (sawa). Both sawi and sawa can be functions of space and time.

Thus,

$$\text{sawi} = \vec{r}_a + i\vec{x}_a$$
$$\text{sawa} = \vec{g}_s + i\vec{b}_a \qquad (8)$$

Please note that since I am defining these terms as <u>specific</u> impedance and admittance, respectively, they are <u>unitless</u>. They are pure complex numbers whose value depends upon a position in four-dimensional space and time. I use the $\vec{r}_a$, $\vec{x}_a$, $\vec{g}_a$ and $\vec{b}_a$ notation in order to identify what their respective roles are, with regard to net acoustic impedance and admittance.

Sawi and sawa represent the geometry-dependent properties of the sound which is characterized by (4) and (5). Sawi and sawa do not tell us what the sound is, but they tell us what is happening to that sound.

If we now look at the equations of instantaneous intensity, as represented solely in terms of sound pressure, we find,

$$\vec{I} = \frac{(p^2+\hat{p}^2)}{2}\frac{(\vec{g}_a - i\vec{b}_a)}{\rho c_o} = \frac{(p^2+\hat{p}^2)}{2}\frac{(\vec{r}_a + i\vec{x}_a)}{\rho c_o(\vec{r}_a^{\,2} + \vec{x}_a^{\,2})}$$
$$= \vec{I}_A + i\vec{I}_R \qquad (9)$$

We thus find that <u>the division of power flow into active and reactive components is governed solely from sawi and sawa.</u> As with all good things, this makes stunning good sense. It is the geometry which determines the division into active and reactive power flow. Up close, near the source of sound, with no foreign object to influence the sound, sawi will be reactive because of the sharp curvature of the expanding sound wave. Sawi rapidly loses its reactive component and approaches unity as we proceed away from the otherwise unhindered source of sound.

If something gets in the way, and upsets the nice four-dimensional space-time relation, such as a reflecting boundary, then sawi changes.

Put into engineering terms: sawi and sawa tell us whether our measurements are "free field" and "distant" or not, and if not, then how much they depart from free field and distant conditions.

<u>TIME AND FREQUENCY</u>

As engineers, we have absolutely no qualms at all about taking the Fourier transform of a microphone signal and stating that we have the frequency domain representation of the sound which the microphone picks up. It is only our ignorance of the details that lets us get away with doing this. In other words, for our purposes it is correct, but not for the reasons we may have thought.

The microphone is somewhere. When we transform, we transform the "when", but not the "where" of the microphone signal. Instantaneous intensity is space-time power flow. It is possible to map this to space-frequency power flow through the use of a one-dimensional Fourier transform which operates only on the time component. We get away with this when there are no space-time cross coupling terms which can tangle the time part up with the space part, such as can be caused by some kinds of nonlinear interactions. Under conditions (1) we do not have such interaction. In nonlinear acoustics we do, and for some audio situations we are perilously close to running into such things.

Our prejudice against complex numbers, which I mentioned in the introduction to the companion paper "Instantaneous Intensity", has led us into one of those silly situations which only a prejudice can induce: frequency can be negative as well as positive, while time is only positive. How deep the mystery of negative frequencies! No mystery. We did it to ourselves by insisting (incorrectly) that a time representation is real-only while a frequency representation can be complex (in fact, it's got to be complex, whether we like it or not).

Although not explicitly stated in any book on the Fourier transform of which I am aware, we can readily show that:

> Under the Fourier transform, single-sided complex maps into single-sided complex, where the inphase and quadrature of each complex are related by Hilbert transform.

My demonstration goes like this:

1. Real a maps into complex A + iB

$$a \longleftrightarrow (1/2)(A + iB)$$

2. The Hilbert transform of a maps into,

$$b = \hat{a} \longleftrightarrow -i(1/2)\text{sgnx}\cdot(A + iB)$$

where sgnx means a scalar multiplier which is +1 if the coordinates of A and B are positive, and -1 if the coordinates of A and B are negative.

3. The "imaginary" operator is unchanged under mapping, so that

$$ib \longleftrightarrow (1/2)\text{sgnx}\cdot(A + iB)$$

4. Thus,

$$a + ib \longleftrightarrow \frac{(A + iB)}{2} + \frac{\text{sgnx}(A + iB)}{2}$$

or

$$a + ib \longleftrightarrow \begin{cases}(A + iB) & \text{if } x > 0\\ 0 & \text{if } x < 0\end{cases}$$

5. A second forward Fourier map restores the original coordinates of a and b, but in a reverse direction. Thus if ξ are the coordinates of a and b and B is the Hilbert transform of A, then this produces a single-sided complex representation (c + id)

$$A + iB \longleftrightarrow \begin{cases} c(-\xi) + i\,d(-\xi) & \text{if } -\xi > 0\\ 0 & \text{if } -\xi < 0\end{cases}$$

6. Two additional forward Fourier maps restore a, b and the direction of their coordinates. It must follow that c=a and d=b such that,

$$A(x) + iB(x) \longleftrightarrow a(\xi) + ib(\xi) \quad \text{if } x, \xi > 0$$

and

$$A(x) + iB(x) = 0 \quad \text{if } x < 0$$
$$a(\xi) + ib(\xi) = 0 \quad \text{if } \xi < 0$$

The frequency domain alternative of instantaneous intensity is sound intensity as a function of space and frequency. And, since they are functions of time and space, sawi and sawa transform into frequency dependent SAWI and SAWA.

Sawi and sawa are complex in the time domain, where the two components are related by Hilbert transform (see relations (6) and (7)) and therefore are complex in the frequency domain, where the corresponding parts are also related by Hilbert transform. It MUST be this way. No other. The frequency domain SAWI has the form,

$$\text{SAWI} = \vec{R}_a + i\vec{X}_a = \vec{R}_a + i\hat{\vec{R}}_a \qquad (10)$$

and the frequency domain SAWA has the form,

$$\text{SAWA} = \vec{G}_a + i\vec{B}_a = \vec{G}_a + i\hat{\vec{G}}_a \qquad (11)$$

The coordinates of (6) and (7) are space-time, while the coordinates of (10) and (11) are space-frequency. We could also have taken the four-dimensional Fourier transform of (6) and (7) to give a wavenumber-frequency set of coordinates.

LOGARITHMIC FORM

The logarithm of a complex representation is itself a complex representation whose "real" part is the logarithm of the magnitude of the original representation and whose "imaginary" part is the angle of that representation plus integer values of 2π. We usually discount the additional 2π terms and use the angle itself for the principal value of the logarithm.

If

$$Z = X + iY = r(\cos\theta + i\sin\theta) \qquad (12)$$

then

$$\log Z = \log r + i\theta \qquad (13)$$

If we wish to express acoustic intensity in decibels, then we must take the logarithm of the intensity. From relation (9),

$$10 \log I = 10 \log \frac{(p^2 + \hat{p}^2)}{2} - 10 \log \rho c_o + 10 \log (\text{sawa})^* \qquad (14)$$

where $(\text{sawa})^*$ is the complex conjugate of sawa. Sawa is the only complex term with which we must be concerned.

This has a very significant engineering interpretation:

(1) Instantaneous acoustic intensity, expressed in decibels, is equal to the decibel value of the ETC plus the decibel value of the time domain $(\text{sawa})^*$.

(2) Frequency domain acoustic intensity, expressed in decibels, is equal to the decibel value of the energy-frequency curve (EFC) plus the decibel value of the frequency domain $(\text{SAWA})^*$.

This imparts complete symmetry to the time domain and frequency domain representations of acoustic intensity. Using the recommended notation which was presented in the paper "Instantaneous Intensity":

For the time domain:

$$\text{dBO} = \text{ETC} + \text{dB}(\text{sawa})^* \qquad (15)$$

For the frequency domain

$$\text{dBO} = \text{EFC} + \text{dB}(\text{SAWA})^* \qquad (16)$$

where dBO is decibels of intensity relative to one Olson (One PicoWatt per square meter), $\text{dB}(\text{sawa})^*$ is the dB value of the complex conjugate of the time domain specific acoustic wave admittance, and $\text{dB}(\text{SAWA})^*$ is the dB value of the complex conjugate of the frequency domain specific acoustic wave admittance. It is understood that both the EFC and the ETC are expressed in decibel values which are precisely equal to the associated sound pressure levels. That is what is now done (or should be done) for EFC and ETC when using TDS measurements.

$(\text{SAWA})^*$, which may be called SAWA-STAR, is unity for free field conditions, so that for free field,

$$\text{dB}(\text{SAWA})^* = 0 \text{ dB} + i\,0 \qquad (17)$$

The same is true for sawa-star.

Equations (15) and (16) give the TOTAL ACOUSTIC INTENSITY MAGNITUDE and its ANGLE. The inphase component is the ACTIVE ACOUSTIC INTENSITY and the quadrature component is the REACTIVE ACOUSTIC INTENSITY.

Equations (15) and (16) are broken into two major components. The first component, ETC in the case of the time domain and EFC in the case of the frequency domain, represents the acoustic signal and its dynamics. The second component represents the geometric properties of the supporting medium.

Thus if we measure sawa or SAWA at a position in space, due to any arbitrary signal which is caused by a source of sound, then, simply by measuring the sound pressure, we know the instantaneous power flow which is due to that same source for any subsequent signal.

We gain another advantage by expressing the decibel value of SAWA: the decibel value of SAWI is the negative of the decibel value of SAWA. This is because,

$$\log(1/Z) = -\log Z = -\log r - i\theta \qquad (18)$$

If the decibel value of SAWA is (0 + i0), then the dB value of SAWI is (-0 - i0). If the dB value of SAWA is (3 + i37) then the dB value of SAWI is (-3 - i37). And so on. The convenience of being able to do this is rapidly grasped once one begins to use this technique to measure sound intensity.

MEASURING sawi

The four-dimensional potential which I have introduced is continuous in its derivatives. This allows me to obtain the complex ratio of complex sound pressure to complex particle velocity as follows. Since,

$$[p] \hat{=} (p + i\hat{p}) = \rho\frac{\partial}{\partial t}(\psi + i\hat{\psi}) \hat{=} \rho\frac{\partial}{\partial t}[\psi] \qquad (19)$$

and

$$[\dot{u}] \hat{=} (\dot{u} + i\hat{\dot{u}}) = -\frac{\partial}{\partial x}(\psi + i\hat{\psi}) \hat{=} -\frac{\partial}{\partial x}[\psi] \qquad (20)$$

then it must follow that,

$$\frac{(p + i\hat{p})}{(\dot{u} + i\hat{\dot{u}})} = \frac{\rho(\frac{\partial}{\partial t}[\psi])}{(\frac{\partial}{\partial x}[\psi])} = \frac{\rho\frac{\partial}{\partial t}(\frac{\partial}{\partial t}[\psi])}{\frac{\partial}{\partial t}(\frac{\partial}{\partial x}[\psi])} \qquad (21)$$

And because of the differential properties of ψ we can interchange the order of differentiation to get,

$$\frac{[p]}{[\dot{u}]} = \frac{\rho\frac{\partial}{\partial t}(\frac{\partial}{\partial t}[\psi])}{-\frac{\partial}{\partial x}(\frac{\partial}{\partial t}[\psi])} = \rho\frac{(\frac{\partial}{\partial t}[p])}{(\frac{\partial}{\partial x}[p])} = \rho\frac{dx}{dt} = \rho(c_o \, sawi) \qquad (22)$$

Thus,

$$sawi = \frac{-1}{c_o}\frac{(\frac{\partial}{\partial t}[p])}{(\frac{\partial}{\partial x}[p])} \qquad (23)$$

This is a remarkable relationship: sawi is proportional to the ratio of the time gradient of sound pressure to the space gradient of sound pressure, where the constant of proportionality is the inverse of the free field speed of sound. Furthermore, we can perform the same operation on particle velocity, as follows,

$$\frac{[p]}{[\dot{u}]} = \frac{\rho(\frac{\partial}{\partial t}[\psi])}{-(\frac{\partial}{\partial x}[\psi])} = \frac{\rho(\frac{\partial}{\partial x}(\frac{\partial}{\partial t}[\psi]))}{-\frac{\partial}{\partial x}(\frac{\partial}{\partial x}[\psi])} = \frac{\rho\frac{\partial}{\partial t}(\frac{\partial}{\partial x}[\psi])}{-\frac{\partial}{\partial x}(\frac{\partial}{\partial x}[\psi])}$$

$$= \frac{\rho(\frac{\partial}{\partial t}[u])}{-(\frac{\partial}{\partial x}[u])} = \rho c_o(sawi) \qquad (24)$$

It is also true that,

$$\frac{[p]}{[u]} = \frac{[p]}{-\frac{\partial}{\partial x}(\frac{1}{\rho}\int[p]dt)} = \frac{\rho[p]}{-\frac{\partial}{\partial x}\int[p]dt} \qquad (25)$$

These are not different relationships, they only look different. Relations (22), (24) and (25) are one and the same. The reason for this high strangeness is tied back to my four-dimensional Euclidean frame of reference, of which relation (1) is the fourth orthogonal "space" component. The things which we call particle velocity and sound pressure (divided by ρ c) are different parts of the same four-dimensional observable. Particle velocity is the x_1, x_2 and x_3 gradient, while sound pressure (divided by ρc) is the x_4 gradient.

To measure sawa and sawi we only need one type of microphone, such as a pressure microphone. Everything is computed from that microphone signal. From relation (4) we need to compute the Hilbert transform of the pressure response and add it as a complex component. If we can use the commercially available TDS instruments for the measurements, then this is done automatically in the process. From relation (23) we need to make three microphone measurements for each space direction of x_1, x_2 and x_3. The gradient in time, the numerator of (23), is the vector difference between two measurements made at the same physical location but separated by a very short interval (short, that is, compared to a period of the highest frequency component in the signal being measured). The gradient in space, the denominator of (23), is the vector difference between two measurements made at precisely the same moment in time, but at two places close to each other and along the direction of the sound axis of interest. If one can use TDS for the measurement, then not only can the "clock" be adjusted for whatever time we wish, but also the same microphone can be used for three separate measurements. This has the technical advantage that the amplitude and phase characteristics, including all microphone polar patterns and structure diffraction effects, can be precisely identical for all three measurements. Even a "cheapie" microphone can make good sawa measurements with TDS.

The quantity obtained from relation (23) is a complex number. Three of these are required to characterize three-dimensional (spatial) sound: east-west, north-south and up-down. It makes no difference WHAT the sound is that we measure: if that sound is passing the measuring microphone under free field conditions and the microphone location is (for normal audio frequencies) at least several meters away from the sound source, then the three-dimensional sawi vector will be unity, with no phase angle, and be pointed in the direction of the sound power flow.

If the sound has reactive, as well as active, power flow associated with it, then the sawi will have a reactive component as well as a resistive component in each of the three space directions. Either sawi or its inverse, sawa, is a valuable acoustic property to know in itself, since this indicates the manner in which the sound wave interacts with its physical environment (including, incidentally, the measuring microphone itself). When added to the ETC, it yields the net instantaneous sound intensity: type and direction.

Relation (23) contains another surprise: it is valid for both the time domain and the frequency domain. The gradient in time delay of frequency dependent pressure divided by the gradient in direction of frequency dependent pressure yields SAWI.

PLOTS

Sawa represents four-dimensional space-time flow of sound. SAWA represents four-dimensional space-frequency flow of sound. Time gated and apodized sawa Fourier transforms to an equivalent SAWA, precisely as is now done for the one-dimensional conversion from time response to frequency response. Relation (7), (9) and (23) then give us a most useful acoustic tool:

> From appropriate space and time gradient measurements, using only pressure responsive microphones, we can determine when a sound arrives, the direction from which that sound comes, and the spectral/temporal content of the intensity of that sound, both active and reactive.

That, to my way of thinking, is useful information in a broad range of sonic disciplines, from underwater sound to room acoustics.

My personal experience using these relationships convinces me that there are so many possible variations on these measurements that I cannot possibly go into details in the brief space allotted to this paper. I shall therefore give a few examples.

Figure 1. is the measured SAWA, amplitude and phase, for a deliberately misterminated analog audio delay line. Designed for audio analysis, this delay line has a maximum delay of 600 microseconds and has selectable taps every 5 microseconds. This electric network is a close analog of acoustic standing wave tubes of the type used to measure absorption coefficient and complex specific impedance. The amplitude of SAWA is in decibels relative to a value of unity, and the phase is in degrees. As seen from relation (18), the advantage of using logarithmic notation is that, on the dB scale, a mirror-rotation of SAWA around 0dB yields SAWI. Thus, +1.5 dB SAWA, relative to the reference impedance/admittance, is -1.5 dB SAWI, relative to the same reference. The nominal delay line impedance is 500 ohms and the measurement is made with a resistive termination of 122 ohms. Peak magnitude SAWA at dc should be,

$$10 \log(500/122) = 6.1 \text{ dB}$$

which it is. Slight progressive delay line attenuation with frequency (it is, after all, a lumped constant physical network) reduces the magnitude of SAWA with frequency. The line isn't as good at 20kHz as it is at dc.

Figure 2. shows a superposition of measured pressure, computed particle velocity and computed active intensity flow, all on the same dB scale. If the delay line were properly terminated, there would be no standing waves and both pressure and particle velocity would coincide at a value representing reactionless power flow. This condition is shown in Figure 3. The upper curve in Figure 3. is the active intensity with the delay line terminated in its characteristic impedance of 500 ohms, while the lower curve is the computed active intensity (using relation (9)) with the line terminated in 122 ohms. The difference in intensity between these two cases should be directly related to the reflection coefficient as follows:

$$K = \frac{R_1 - R_2}{R_1 + R_2} = \frac{500 - 122}{500 + 122} = 0.6077$$

$$10 \log(1 - K^2) = 2.0 \text{ dB}.$$

- 9 -

The measured separation is indeed 2 dB at zero frequency where line loss is smallest. The deviations above about 5 kHz are due to the fact that the least increment delay taps on this line are 5 microseconds, which gives a coarse estimate of space gradient at the higher frequencies whose wavelength periods are no longer large compared to 5 microseconds.

Figure 4. is a superposition of active intensity and reactive intensity on the same scale. Note that reactive intensity is minimum where either pressure or particle velocity is a local null.

All of these measurements are obtained from SAWA. Particle velocity is pressure times SAWA. Intensity is EFC times SAWA-STAR. A more interesting application of SAWA is shown in Figures 5 and 6. Figure 5. is the pressure response (EFC) of a loudspeaker system composed of two identical wide-range loudspeakers which are physically displaced and electrically driven in parallel. From first principles we get a comb filter interference response. This one is deliberately made as bad as I can make it. A practical question is: not knowing what the actual loudspeaker may be, is this measured response due to loudspeaker misalignment or due to acoustic reflections picked up in the microphone, or both? The measured single-axis SAWA, along the line between the microphone and the center of the loudspeaker cluster, is shown in Figure 6, amplitude lower curve, phase upper curve. SAWA tells us that the comb filter, large dips and peaky low frequency response is indeed due to the sound source, and not to flanking signals getting into the measurement. But look at the phase details from 6 kHz to about 14 kHz. This tells us that there is a very small reflection signal coming in at these frequencies. In this particular case, the measurement is made with TDS in a small room at a distance of one meter from the loudspeaker system. A combined ceiling and floor reflection is indeed sneaking through the TDS filter and SAWA tells us <u>how much the direct path signal is in error from true free field response</u>. For example, at 8 kHz, the direct response is notched by 24 dB and the flanking signal, which does not suffer such a notch due to its off-axis location, introduces a 4 dB peak to peak error in amplitude and a 45° peak to peak error in phase.

HOW TO DO IT

These measurements are made utilizing time delay spectrometry as implemented in the TECRON System 10 TEF machine. Relation (9) and either (23) or (25) may be used in real world acoustic measurements of SAWA using TDS and <u>one pressure microphone</u>. The procedure is as follows: (1) the microphone is firmly placed in a position where SAWA is to be determined. (2) The time gradient of sound pressure at this location is obtained by vector differencing two TDS measurements, whose time delay differs by a small amount (6 microseconds, by way of example). It makes no difference whether you add or subtract microseconds, but a negative time gradient must accompany a positive space gradient to yield a SAWA phase which is zero degrees when you are lined up precisely with the direction of sound. (3) The space gradient of sound pressure at this location is obtained by vector differencing two TDS measurements which

- 10 -

are taken at the same time delay, but physically offset by a small amount in the direction of the SAWA component of interest (2 millimeters, again by way of example). If the space gradient is called X and the time gradient is called T, then the dB magnitude of SAWA is the scalar difference of X with respect to T, obtained using the TEF machine and the amplitude mode. The phase of SAWA is obtained in precisely the same manner, using the TEF machine and the phase mode. Regrettably, the dB amplitude of SAWA will be twice as large as it should be. That is, an admittance ratio of 2:1 will show as 6 dB, rather than 3 dB. This is because the unmodified TEF programs were not designed to measure impedance and admittance in dB. Appropriate software will probably become available for SAWA as well as for the things which may be obtained from SAWA, such as sound intensity.

SAWI is the inverse of SAWA and is obtained by the scalar difference of T with respect to X. Three-dimensional SAWA is readily obtained by keeping the vector difference in T and scalar differencing the appropriate space gradients with respect to T.

As with all measurements, the best instrument you have is common sense. Most microphones are not only larger than 2 millimeters, but have diaphragms which lie on a defined surface. I use a B & K condenser mic and "point" it along the direction of sound which I want to measure. I have thus successfully used spatial increments smaller than a tenth millimeter with no difficulty. Were I to move the mics laterally, then most of the diaphragm would be overlapped in the two measurements. It works, but kills even the large signal to noise ratio gained by TDS.

If the space gradient is lined up precisely with the direction of arriving sound, then the phase of SAWA will be correct if the delta-x of the space gradient spacing, divided by the delta-t of the time delay offset is equal to the speed of sound. If it is not, then the phase will be tilted with respect to linear frequency. It is not necessary to repeat the measurement, simply unwrap the phase. Since the amplitude of SAWA is expressed in dB, the same effect may show as a vertical offset in the dB value. Pending available software, this is a small penalty to pay for a practical measurement, since it is normally the deviations from constancy in dB which are significant in most measurement situations.

The larger the space offset used for space gradient, the larger the signal to noise ratio for low frequencies, but the delta-x must be less than about a tenth wavelength at the highest frequency which you want to measure. A useful "frequency ratio" of SAWA obtained in this manner is about 100:1 before you begin to run into practical noise limitations in a quiet environment; 20:1 in a noisy environment.

SUMMATION

Sound moves in both space and time. It is the physical geometry of the space that determines what will happen to the sound, whether it moves outward in unhindered fashion, or rebounds upon itself in multiple cascades.

We can capture a piece of sound at a particular place in space and determine whether this piece represents sound that is passing in unhindered fashion or not, as well as the direction of its stream. This time domain property is called sawa, while the frequency domain counterpart is SAWA.

Sawa, specific acoustic wave admittance, expresses the division of sound power flow into active and reactive components. Sawa is numerically obtained as the ratio of the way a sound changes in space, at a fixed moment in time (the space gradient of sound), to the way that sound changes in time at a fixed position in space (the time gradient of sound). The inverse of sawa is sawi, specific acoustic wave impedance. For sound which is passing in free field fashion at a great distance from its source, sawa, and its inverse, sawi, is a directed vector whose value is unity, with no reactive component.

Audio engineers now measure the amplitude of sound pressure level as a function of frequency (the EFC dB amplitude) as well as the amplitude of sound pressure level as a function of time (the ETC dB amplitude). Both the EFC and the ETC measure total energy density. That is what the "E" stands for in their respective acronyms. The EFC and the ETC characterize the signal, not the space. Because of this, neither the EFC nor the ETC indicates direction of sound. SAWA and sawa characterize the space, not the signal. When we put the two together we have a powerful expression of what the sound is and what is happening to that sound due to objects which it encounters in space. Combining the EFC with SAWA yields sound intensity as a function of frequency. Combining the ETC with sawa yields sound intensity as a function of time. For free field sound, the amplitudes of EFC and ETC are numerically equal to the sound intensity in their respective domains. This numeric relationship is facilitated by using notation in which the dB value of EFC and ETC is identical to the dB value of sound pressure level from which they are measured. Thus expressed, the sound intensity is then given in decibels of power flow in Watts per square meter relative to a reference level of one PicoWatt per square meter, or one Olson.

The simplest way to measure EFC, ETC, SAWA and sawa is by the use of pressure responsive microphones. It is not necessary to use velocity microphones, nor is the space gradient of sound pressure a poor approximation to particle velocity. Rather, both sound pressure and particle velocity can be considered different aspects of the same entity, a potential whose four-dimensional gradient yields the sound pressure and particle velocity components with which we are familiar.

REFERENCE

[1] Heyser, R.C., "Determination of Loudspeaker Signal Arrival Times", J. Audio Eng. Soc., pg 902 (1971)

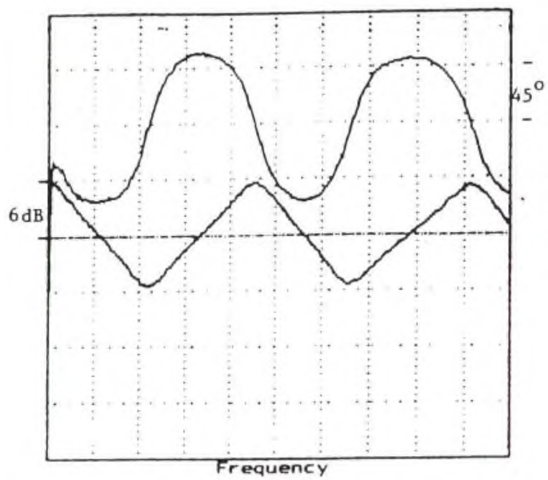


Figure 1. SAWA, amplitude (middle curve at 6dB/div) and phase (upper curve at 45^{o}/div) for a misterminated delay line.

Vertical: 6dB/div with base of display at 11.0dB
0dB is located at .001 volt

Horizontal: 0.00Hz to 19998.10Hz
scale: 5467.68Hz/inch or 2152.63Hz/cm.

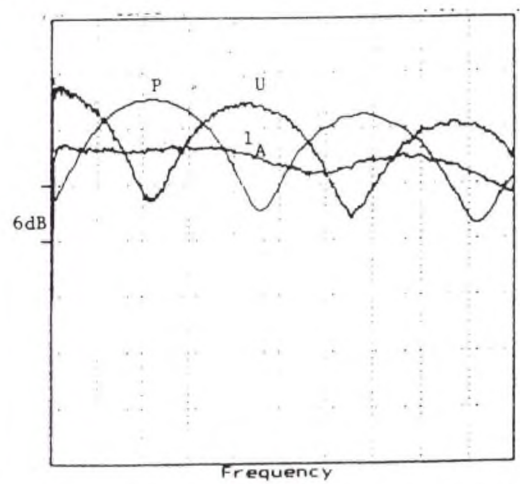


Figure 2. Sound pressure P, particle velocity U, and active intensity I_A for the delay line of Figure 1.

Vertical: 6dB/div with base of display at 11.0dB
0dB is located at .001 volt

Horizontal: 0.00Hz to 19998.10Hz
scale: 5467.68Hz/inch or 2152.63Hz/cm.

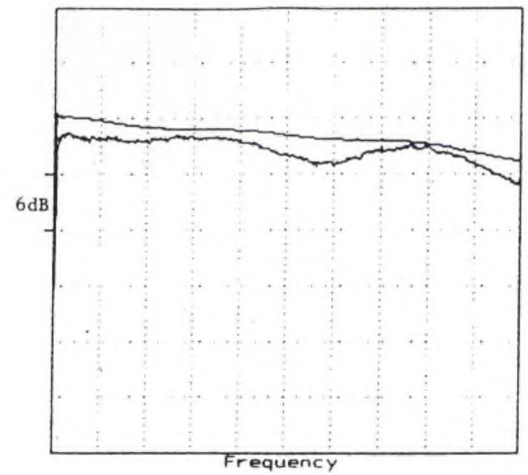


Figure 3. Active intensity for terminated (upper curve) and misterminated (lower curve) delay line of Figure 1.

Vertical: 6dB/div with base of display at 11.0dB
0dB is located at .001 volt

Horizontal: 0.00Hz to 19998.10Hz
scale: 5467.68Hz/inch or 2152.63Hz/cm.

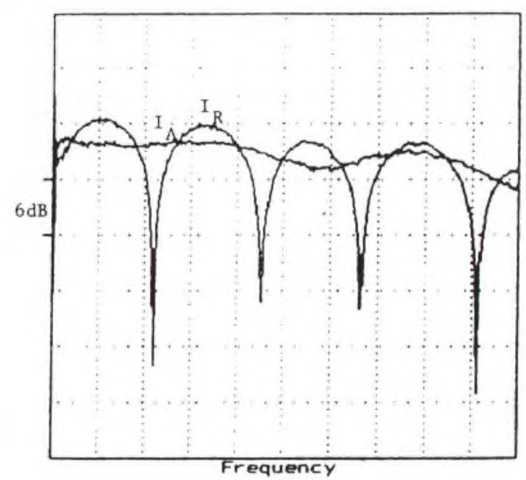


Figure 4. Active intensity I_A and reactive intensity I_R for delay line.

Vertical: 6dB/div with base of display at 11.0dB
0dB is located at .001 volt

Horizontal: 0.00Hz to 19998.10Hz
scale: 5467.68Hz/inch or 2152.63Hz/cm.

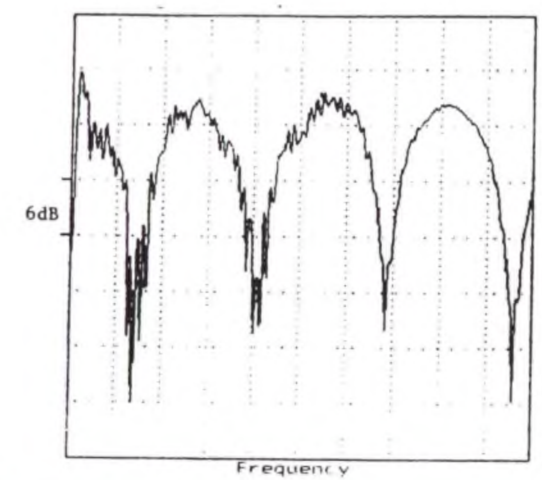


Figure 5. Sound pressure response of misaligned loudspeaker system.

```
Vertical:   6dB/div with base of display at    29.0dB
            0dB is located at  .001  volt

Horizontal: Auto        0.00Hz to  19998.10Hz
     scale:    5467.68Hz/inch or    2152.63Hz/cm.
```

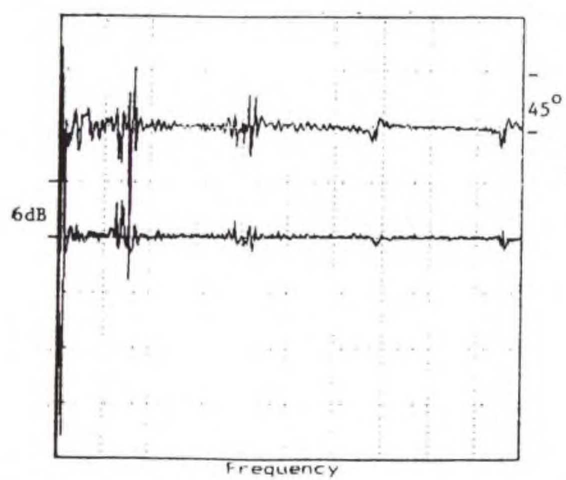


Figure 6. SAWA, amplitude and phase for loudspeaker system of Figure 5.

```
Vertical:   6dB/div with base of display at    29.0dB
            0dB is located at  .001  volt

Horizontal: Auto        0.00Hz to  19998.10Hz
     scale:    5467.68Hz/inch or    2152.63Hz/cm.
```

Fundamental Principles and Some Applications of Time Delay Spectrometry

Richard C. Heyser
Tujunga, California

INTRODUCTION

The paper that follows represents Richard C. Heyser's final contribution to the literature on time delay spectrometry (TDS). He was in the process of writing it when he died. Although we cannot say whether he would have wanted the paper published in this form, I felt that it was important to include it in this anthology.

Dick had completed a number of experiments and was conducting others to verify the theories that he advanced. He may have wished those experiments to be included here, but they were "lost with Dick."

Of the three drafts found, one appeared substantially more complete. That manuscript, which has been prepared for publication by Dr. D. H. Le Croissette, required only minor editing.

This paper provides "food for thought," and I hope it will encourage further progress in the creative realms of science. The information presented here is primarily oriented to audio, yet the scope of Dick's work extends far beyond, with significant implications in the fields of mathematics and physics. Some may view these concepts as controversial; that is another good reason for presenting them. There are concepts which challenge the mind and get the creative processes flowing—this paper contains just such intellectual inspiration.

John R. Prohs

Fundamental Principles and Some Applications of Time Delay Spectrometry

by Richard C. Heyser
Tujunga, California

BACKGROUND

The original conception of time delay spectrometry was motivated by the need to find some method for measuring the free field response of loudspeakers, while those loudspeakers were in reverberant environment. The problem was to find some method of isolating the direct sound from later arrivals and measuring the frequency and time behavior of this direct sound. The solution involved the use of a specific class of excitation signal which, in effect, placed a time tag on each frequency component. This allowed a special filtering procedure to extract signals taking a path of known time delay and to display the frequency and time behavior of those signals, to the selective exclusion of signals taking paths of longer or shorter delay. Since the complex spectrum was obtained as a function of this time delay, the procedure of measurement was called time delay spectrometry (TDS).[1]

It soon became apparent that this conception not only provided one possible solution to the original problem, but led to a much deeper insight into the more general problem of characterizing signals propagating through media which can be characterized as dispersive and absorptive. In order to understand how this method worked, and to provide some firm mathematical basis for its operation, I was forced into deeper levels of analyses than I had expected to penetrate.

As can happen in such things, this more thorough mathematical analysis of the elemental procedure revealed a much deeper structure. Whereas the original intent was to measure traditional time domain and frequency domain properties, analysis showed that the procedure was, in effect, a key that could be used to unlock other domains. Stated simply, time delay spectrometry is the implementation of a specific class of integral transform that maps among domains of differing dimensionality.

This opens the door to an enormously rich field of analysis. Its effect may be conceptualized by imagining what contemporary signal theory would be without the tool of Fourier analysis.

Measurement could proceed, but at a much more difficult pace, and certain elementary facts (which we assume to be elementary only because we know about Fourier spectrum analysis) would not be known. Imagine the increment in understanding brought about by introducing Fourier analysis to such a barren theory; for example, by adding the frequency domain to the traditional time domain. With this in mind, imagine what might result when we learn how to add new domains.

Although the methodology of TDS soon outstripped the simpler concepts of the frequency and time domain for which the procedure had originally been named, no need has been felt to change the name, or acronym, even when applied to the more general mapping procedures.

My original interest in measuring the acoustic performance of loudspeakers stemmed not from concern with loudspeakers, but from my personal interest in the mysteries of human subjective perception. There is often an apparent discrepancy between subjective perception and the objective measures of the ingredients of that perception. I was interested in trying to understand why, to use a popular phrase, things do not always measure the way they sound. Although my original concern was with the listening qualities of then-new types of power amplifiers[2] it was necessary to measure the performance of loudspeaker, since these were the source of sound uses in acoustic perception.

It had occurred to me that perhaps the reason why conventional signal analysis and subjective perception did not necessarily agree was due to a difference in their respective frames of reference and, in particular, due to differences in dimensionality between these frames of reference. They were, after all, both descriptions of the same "event". It was this decision to investigate differences in frames of reference that precipitated the concept of alternatives which is described in Section I.

Although the original problem was acoustic in nature, I soon found that I was involved in relationships that, at first, seemed far afield of acoustics. From the beginning, it was apparent that the mathematical structure which I was developing paralleled those which could be found in other branches of sciences, including quantum mechanics. This is not an unusual occurrence. The famous indeterminacy relation, for example, was quickly applied to acoustics,[3] and Gabor[4] utilized the mathematics of quantum mechanics in order to align the apparent "frequency" of a tone burst with its mathematical description. But because I had abandoned dimensionality as an invariant of a description, I was forced to tread what was, to me, new ground. Whenever possible, I would map my results back to a discipline and dimensionality with which I was familiar and test my hypotheses. The results were sometimes surprising. Higher dimensional analysis not only confirmed the existence of relationships that were well known in lower dimension, but gave new insight into their cause. Several examples of this are presented in Section I.

The first open literature descriptions of a practical time delay spectrometer were based on conventional engineering practice.[1,5] Since the time domain and frequency domain are so much a part of our present analysis methodology, the first descriptions employed the model of a swept frequency excitation and a tracking filter. This proved quite successful in obtaining time domain and frequency domain properties of linear time invariant systems and was an easy model to grasp. But this simple model could not begin to exploit the broader concepts which a thorough analysis began to reveal. Subsequent publications[6-9] began to outline some of these concepts, but, quite understandably, remained somewhat mysterious to those accustomed to contemporary techniques, and the first TDS implementation model remained a model of choice for experimenters in acoustics and audio engineering.

My first published papers on TDS did not come at a fortuitous time. The FFT had just been introduced and, other than some pioneering spirits, few persons were interested in utilizing the early implementation of TDS which, like Fourier analysis of a prior decade, could only be assembled with analog instrumentation.

There was the further misconception of the time that since the first TDS instrumentation utilized a so-called "linear frequency sweep", it must be a variant of conventional chirp technology. To some extent this misconception still exists among those who have given present TDS instruments only a cursory glance. The linear frequency sweep is indeed utilized in most of the present instruments, but only because this provided ready access to the alternative time domain (equation 71) and frequency domain (equation 66) with which signal analysts have the most familiarity. The general TDS kernel (equation 54) has been successfully used in other phase programs, including pseudorandom, to pull out selected properties of a signal process, although this is not generally available with present commercial TDS instruments. It is my hope that by presenting the more general mathematical development of Section II, other experimenters will be prompted to break away from traditional analyses which depend so heavily on what, as I have shown, are only two of the infinite number of alternative representations, the time and frequency domains.

I believe that the energy relations which are developed in Section I are important, not only in the understanding of TDS, but in analysis in general. These are truly general relations and do not depend upon any particular frame of reference, or dimensionality of representation. Any observation of a linear energetic exchange process must consist of two parts in each of the dimensions involved, and these two parts must be related in the specific way developed in Section I. Acoustic impedance, for example, must conform to the rules. Not only must the "real" and "imaginary" components of acoustic impedance be related by Hilbert transform, but there must be a preferred relationship between them such that a complex plot made on the fg-plane (relation 7) must curl in a constant direction with increasing values of the coordinate of expression. For example, the complex plot must curl clockwise with increasing values of frequency for frequency-dependent descriptions of impedance, and counterclockwise for increasing values of time for a time-dependent description of impedance. It may never deviate from this form if the proper coordinate basis is chosen. Thus, not only can the "imaginary" component be computed from the "real" component, but if one is presented with a complete plot of only one of them, it is possible to identify by visual inspection whether it is the "real" or "imaginary" component.

Any experiment which results in a single "real" component, for example, seismic detection and evoked potential response to stimulus, to mention only two, will benefit from utilizing, through measurement or computation, the Hilbert transform related counterpart described in Section I. The fg-plot, a complex representation in which projections on quadrature axes are related by Hilbert transform, is a true energy representation and can be of great value in analysis. The detection of a deterministic process within an otherwise stochastic process can often be facilitated through proper use of this plot. And the general geometric properties of "radius" and "angle" on this plot can tell much about a process which is so represented. One example of this occurs when the "radius" and "angle" are themselves related by Hilbert transform, a condition known as minimum phase in the frequency representation of network response. The energy properties of this very special geometric relationship reveal that there will be an optimum clustering of total energy content in the particular alternative representation which is reached by the Fourier transform.

With the increasing use of TDS in performing acoustic measurements, now is an appropriate time to present a more complete description of the theoretical and mathematical basis from which TDS was developed. I hope that this present tutorial discussion may spark others to examine the possibilities which the underlying concepts present for advanced acoustic measurements.

OUTLINE OF PRESENTATION

TDS is now becoming an accepted method of acoustic measurement, taking its place alongside conventional, long-standing techniques. The ranks of users of this technique have continuously expanded during the past decade, and TDS has been applied in such diverse applications as architectural acoustics, loudspeaker measurement, underwater sound, and medical ultrasound imaging.[10-23] There are now two commercially available instruments for TDS acoustic testing.

A number of papers have been written on this subject and an excellent analysis of the technique has recently appeared.[24] What, then, could this present paper offer that is not known about the technique? The answer is that the concepts underlying TDS are considerably more extensive than might be apparent to those who use its present implementation. Future versions of the technique will draw upon concepts which are presented for the first time in this paper. There is, as I intend to show, the equivalent of a new paradigm for modelling and problem solving purposes which TDS uniquely implements.

In order to present the broader picture, this paper is divided into four sections. The first section deals with the conceptual basis that underlies the proposed paradigm and presents some energy relations which derive from that paradigm. The assertion that this is a new paradigm demands justification and several examples are presented in order to provide that justification.

The second section extracts those analytical aspects of the paradigm which underlie TDS. This leads to an analytical expression for mapping between alternative frames of reference which is a special type of integral transform.

The third section describes the manner in which TDS implements the integral transform. The presentation progresses from the abstract, through a simple mathematical model, to the block diagram of presently available TDS instrumentation.

The fourth section discusses a number of practical applications of TDS which have already been made in the field of acoustics.

SECTION I

THEORETICAL BACKGROUND

Although contemporary TDS instrumentation presents its results of measurement either as time domain or frequency domain properties, I intend to show that eventual use will not necessarily be limited to these specific domains. In addition, contemporary TDS instrumentation makes use of a particular complex pair of entities in both the time and frequency domains, which I consider to be related to energy density within the frame of measurement used for the particular TDS measurement. In order to understand the basis for both of these assertions, it is necessary to present a more penetrating analysis of the fundamental considerations from which this technique was derived. Accordingly, this section briefly summarizes the necessary theoretical and conceptual background to understand TDS.

A Paradigm of Alternatives

The manner in which each of us approaches a problem-solving situation depends upon the particular conceptual model which we use to characterize that situation. This conceptual model is the very foundation of what may be called our "understanding" of the problem. Thomas Kuhn[25] formalized this idea of basic conceptual models used within technical disciplines, and referred to them as paradigms. It has been Kuhn's assertion that scientific revolutions within a discipline do not so much depend upon a continual expansion of knowledge within that discipline, as if present experts were merely standing on the shoulders of their forebears in order to see more clearly, but is often a result of paradigm changes that produce an entirely new perspective.

Although Kuhn's concept of paradigm has been applied (or misapplied) by overzealous followers far beyond his original intent, the basic premise of his work is still worth considering.

A proper understanding of the theoretical basis of time delay spectrometry requires a recognition of what I believe to be an implied paradigm in ongoing analysis and the introduction of a different paradigm, which I shall refer to as the paradigm of alternatives.

The underlying assumptions that link these paradigms relate to the concept of frame of reference. This can be illustrated through the example of generalized Fourier series. The set of all functions defined and (Lebesgue) measurable on a set **T** in N-dimensional space, and with the property that the square of the norm is finite and induced by an inner product,

$$\| \mathbf{x} \|^2 = (\mathbf{x},\mathbf{x}) = \int \mathbf{x}(t)\,\overline{\mathbf{x}(t)}\,dt < \infty \tag{1}$$

is a separable Hilbert space, which is designated $L^2(T)$ when the inner product is taken to be,

$$(\mathbf{x},\mathbf{y}) = \int \mathbf{x}(t)\,\overline{\mathbf{y}(t)}\,dt \tag{2}$$

where the overbar denotes complex conjugation.

When this is established, it follows that there is for every **x** in the space a unique expression,

$$\mathbf{x} = \sum^{\infty} C_N e_N \tag{3}$$

where $C_N e_N$ is a countable set of a complete orthonormal set and $C_N=(x,e_N)$. The series representation is called the generalized Fourier Series of **x** with respect to e_N. If, for example, **T** is the interval $[0, 2\pi]$ of the real line, then the expansion is the usual Fourier sine and cosine series.

Let us take a more careful look at a hidden assumption. The orthonormal set is expressed in terms of the same frame of reference as **x**. That is what we mean when we state that **x** can be expanded in terms of elemental expressions taken from an orthonormal set over **T.** In the case of traditional modal analysis, we presume that a particular response in terms of coordinate **t** can be decomposed into terms that are also expressed in terms of coordinate **t**.

There is nothing wrong in such an approach and its mathematical pedigree is beyond reproach; but a paradigm lies at its very heart, a problem-solving model which presumes that one needs to stay in the same frame of reference. The point I would make is that we do not think of this as a paradigm until it is called to our attention.

Consider another approach. Let us elevate the significance of the frame of reference which is to be used in any measurement (or observation). In terminology quite familiar to students of quantum mechanics, assume that we must first establish a frame of reference; then we can express our theory of how things work in this frame of reference and conduct experiments in consonance with this theory.

When confronted with a problem-solving situation, rather than persisting in a search for elemental components expressible in terms of the same frame of reference, suppose we search for a different frame of reference in which the part of the problem of interest to us is now of more manageable form, even though abstract.

It is quite apparent that this does *not* alter any part of the existing mathematical structure of our analysis. But it does introduce the new assumption that there is no "preferred" frame of reference. The situation can be conceptualized in terms of the symbolic diagram of Figure 1. An observer, utilizing frame of reference A, can interact with and measure an event. The presumption is that not only can another observer, utilizing some other frame of reference B, observe the same event from his special perspective, but that there are an indefinitely large number of alternative frames of reference which can be used to observe the same event.

As far as each observer is concerned, the event occurs solely within his specific frame of reference. He sees everything that can be seen and there is no part of the event inaccessible to him. It may, in fact, be totally inconceivable to such an observer that there could be some other way of describing the event. In a previous publication this concept was presented as a principle of alternatives.[10] In a particular problem-solving situation, we may be likened to an observer in some frame of reference A. Properties which are of interest to us may be considerably enmeshed with other undesired properties, so long as we stay in system A. There are an infinite number of alternative frames of reference in which we might view the problem of interest. Perhaps one of these reference systems, for example C, will be such that the property of interest, as expressed in that system, is separated from the undesired property. Rather than try to express the abstract frame of reference C in terms of A, TDS is a methodology for mapping our entire problem to the alternative frame of reference C; where we can separate the desired property, and then choose to map the results either back to frame of reference A for observation of the results in that system or to some other alternative frame of reference for further analysis and consideration.

The foregoing example does not address deep philosophical considerations, such as the meaning to be assigned to the terms "observe" and "event", but it does outline the concept of alternatives. Space does not permit a more general approach, so from this point forward we shall limit the present discussion to a specific type of frame of reference which encompasses the linear

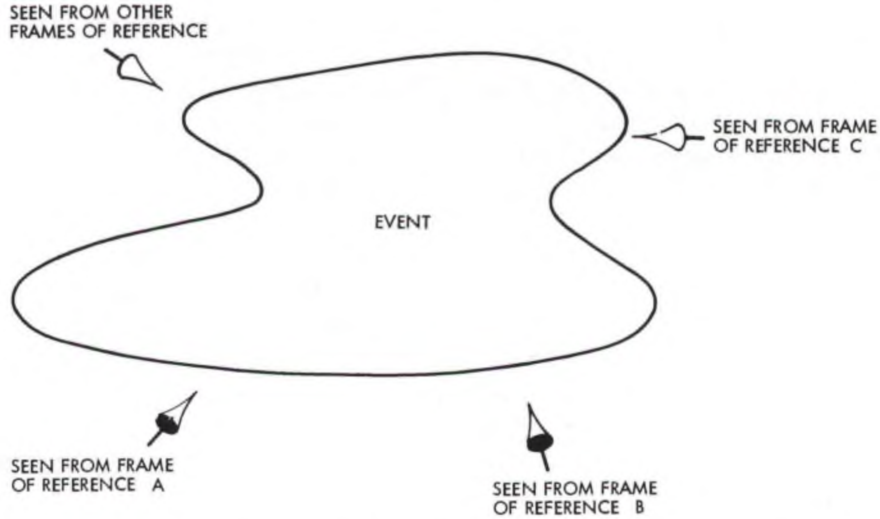


Figure 1 Symbolic diagram showing views of the same event from various frames of reference.

time and frequency domain of conventional analysis, and which leads to the development of linear TDS.

C-Alternatives

The principle of alternatives states that, if there is a valid functional representation **f**, which is expressed in some frame of reference **x**, then there will exist maps $\mathcal{M}$, which can transform this **f**(x) into equally valid representations in terms of alternative frames of reference. Those representations, which are equally valid under a defined set of conditions **C**, constitute a set of alternatives under conditions **C**. In the case of linear signal analysis, the condition of Lebesgue square summability, in which the representations are of class L^2, defines a particularly useful set of alternatives.

If representations **f** and **g** are **C**-alternatives under $\mathcal{M}$, then in symbolic notation,

$$\mathcal{M}: \mathbf{f} \overset{\mathrm{C}}{\longmapsto} \mathbf{g} \tag{4}$$

In words, $\mathcal{M}$ is the procedure that converts an **f** into a **g** under conditions **C**.

For the remainder of this article, I will assume that **C** is the condition of equal Lebesgue square measure. The term "observe" can then be related to transfer of Lebesgue measure between the observer and that which is observed, for only things of non-zero Lebesgue measure can be carried between alternative observables.

Let us assume that **f** is a valid functional representation expressed in an N-dimensional frame of reference, which, for simplicity, we will assume is Euclidean in its geometry. On the other hand, **g** is an equally valid representation expressed in a different frame of reference. The frame of reference for **g** may be M-dimensional and also is assumed to be Euclidean; this makes condition **C** the requirement that each space be a Hilbert space on which an inner product can be defined. The mapping procedure $\mathcal{M}$ then has a most interesting task; it must transform expressions from one frame of reference to another without altering their geometry under conditions **C**, even though the frames of reference may be of different dimensionality.

If all possible alternatives are equally valid for characterizing an observable event and are nothing more than different ways of representing that event to an observer, then we can pose a very interesting question: What observable property does not change when we alter our frame of reference? Since we can alter dimensionality of a representation by mapping from one alternative to another, we can no longer depend upon the familiar concepts of topology. Dimension is a topological invariant,[29] and we can no longer carry such concepts as continuity of functions as self-evident truths to be applied to alternative representations.

A truly invariant property under such maps must be one which has no dependence whatsoever upon frame of reference. It must be scalar. Net scalar Lebesgue measure is such a property. It is a property of **C**-alternatives that net Lebesgue measure is conserved. If we presume that the process of "observe" equates to transfer of Lebesgue measure—that is, nothing can be observed which does not have Lebesgue measure—then the conservation of Lebesgue measure corresponds to conservation of net energy under conditions of complete alternative representation. This seemingly abstract consideration can now be developed into some useful energy relations.

Energy Relations

If there is a finite scalar entity E, identified as total energy, which is to remain invariant when all aspects of a process are considered over some frame of reference s, then we can postulate the existence of a total energy density E(s) such that the net sum of this density over all values of this frame of reference is equal to the total energy. This can be expressed as,

$$\int_S E(s)\, ds = E \tag{5}$$

The set of all frames of reference s for which this invariance holds will be defined as generalized coordinates of energy measure. When energy representations in generalized coordinates s_i are mapped to energy representations in generalized coordinates s_j under conditions of conservation of total energy E, then

$$\int_{S_i} E(s_i)\, ds_i = \int_{S_j} E(s_j)\, ds_j \tag{6}$$

As it stands, this expression is of little utility, but if we further define a new complex entity h(s), the square of whose magnitude is equal to E(s), then we can cast relation (5) into a form that is known to be of use in descriptions of energy related events. Defining

$$h(s) = f(s) + ig(s) \tag{7}$$

we have

$$\int_S h(s)\,\overline{h(s)}\, ds = \int_S |h(s)|^2\, ds = E \tag{8}$$

Since E is finite, h(s) is square integrable, and consequently will be of class L^2 if ds is the Lebesgue measure of the frame of reference s. The one assumption we have to make, in order to use h(s), is that energy density is positive definite. If energy density is positive definite, then we can define an h(s) whose square is equal to that energy density. The function h(s) needs to be complex for completeness, since the set of the square of real numbers is a subset of the set of the square of complex numbers.

The strategy of defining a complex number of class L^2 allows us to draw from the extensive mathematical literature concerning linear spaces of class L^2. Of particular benefit is a theorem due to Titchmarsh[26] which identifies the necessary and sufficient condition under which a complex h(s) will be analytic and of L^2.[(27)] The condition is that f(s) and g(s) be conjugate functions related to each other through the Hilbert transform.

The Hilbert transform is of strong L^2 type,[27, 28] which means that not only is

$$| h(s) |^2 = [f(s)]^2 + [g(s)]^2 \tag{9}$$

everywhere except over sets of Lebesgue measure zero, but that

$$\int_S (f(s))^2\, ds = \int_S (g(s))^2\, ds \tag{10}$$

The result of this is that when we impose L^2 structure on our analysis the expression which is identified as total energy density is partitioned into two terms,

$$E(s)=V(s) + T(s) \tag{11}$$

where

$$E(s) = | h(s) |^2 \tag{12a}$$

$$V(s) = [f(s)]^2 \tag{12b}$$

$$T(s) = [g(s)]^2 \tag{12c}$$

A further result is that if we integrate over the whole of s in order to evaluate the scalar partition of E which is imposed by this additional structuring, then,

$$E = V + T \tag{13}$$

and

$$V = T = (1/2)\, E \tag{14}$$

where

$$V = \int_S V(s)\, ds \tag{15a}$$

$$T = \int_S T(s)\, ds \tag{15b}$$

Some discussion of terminology is in order. E(s), by definition, embodies the total energy density that can be measured in system s; there is no additional measurable term which relates to total energy density. In order for the net scalar energy to be conserved, there must be two components of E(s), and they must take the forms which I have identified as V(s) and T(s). One of these terms clearly corresponds to that expression which we have traditionally called potential energy density, as expressed in system s. Accordingly, I have shown it as V(s). The other term, the other "half" of total energy density, is identified as T(s), and henceforth I shall refer to it as kinetic energy density, When this relationship was first discovered, I anguished over the meaning of the term which I called T(s), but could see no other identification than that of a kinetic component; and the theorem was published with this nomenclature.[7] Relationships (11) - (15) have survived many subsequent challenges. The issue is discussed later when we consider the significance of the energy time curve, but a simple interpretation of these relationships is offered when we consider the place-by-place measurable value of f(s); namely, g(s) is that global distribution which is necessary to support the local departure from equilibrium identified as f(s) and which is necessary for the conservation of net scalar measure for all alternative representations of f(s).

With these considerations in mind, we can identify a fundamental relationship:

The necessary and sufficient condition for the conservation of total energy in a linear system is the partitioning of any observation of that energy into two components related through the Hilbert transform.

A second relationship is that,

A "real" observable f(s) whose square is proportional to energy density yields half the total energy that can be attributed to a linear process.

In Defense of the Paradigm

Let me return to the assertion that we are dealing with a new paradigm. Although not explicitly stated by Kuhn, it would seem that there are two conditions that must be satisfied by a new paradigm. First, the new paradigm must completely embrace everything that is known in the ongoing paradigm; there can be nothing modelled in the ongoing paradigm that cannot be modelled in the new paradigm. Second, the new paradigm must lead to predicted results which are not explainable in the ongoing paradigm but which can be readily tested and demonstrated without violating any of the accepted constraints of the ongoing paradigm. Clearly, the concept of alternatives satisfies the first condition, since it does not alter any of the analytical structure now used in problem-solving situations. The concept of alternatives also satisfies the second condition, as will now be shown.

Three examples will be presented. The first example relates to energy density and its partitioning. The second example is that of the noncommutation of the procedures of the operator and transform, and the third example is the predicted noncommutation of certain important classes of operator used in quantum mechanics as well as acoustics.

Example 1

Consider the simple electrical circuit of Figure 2. A battery of potential V is applied through a switch and unspecified network N to a capacitor C. Assume that network N may be anything whatsoever that does not have measurable steady-state potential energy in terms of the voltage frame of reference used to characterize the network, and which allows the capacitor to eventually equilibrate in voltage with the battery. The network may consist of a short wire, a resistor, or combinations of coils, transformers, lamps, motors, and electromagnetic antennas. Regardless of the initial stored potential energy of the capacitor, what is the relationship, when the switch is closed and conditions have equilibrated, between the net energy transferred to or from the battery and the net potential energy remaining on the capacitor?

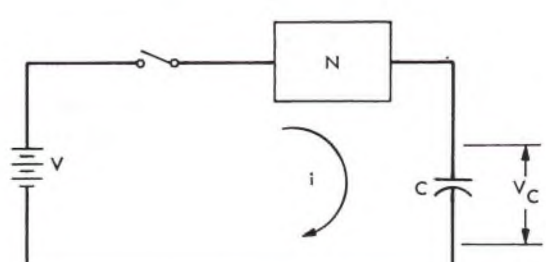


Figure 2 Simple electrical circuit.

The previously derived energy relations predict that it takes energy to store energy. If a net energy E were transferred away from the battery, then the prediction is that a net potential energy of (l/2)E will be observed in the storage medium. The other half will be attributed to a quadrature component which is not observable in a potential energy frame of reference.

This assertion can readily be demonstrated using conventional circuit relations, The net energy transferred to (or from) the battery is related to its voltage and the circuital current i, by

$$\int_0^\infty V \cdot i dt = V \int_0^\infty i dt \tag{16}$$

But since the voltage across the capacitor, V_c, is related to the current by,

$$i = C \frac{dV_c}{dt} \tag{17}$$

it follows that,

$$V \int_0^\infty C \frac{dV_c}{dt} dt = CV \int_0^V dV_c = CV^2 \tag{18}$$

Under steady-state conditions we know that the potential energy stored in the capacitor is,

$$\frac{1}{2} CV^2 \tag{19}$$

which is precisely half that which was transferred. Although presented as an electrical circuit, it can be readily demonstrated that the same situation prevails for hydraulic and mechanical situations. There is nothing in our present paradigm that predicts why such a precise energy partitioning should occur, although the physical and mathematical demonstration is quite simple to perform.

Example 2

The second demonstration draws from conventional functional analysis. Under contemporary analysis, the terms map, operator, function and transform are considered synonymous.[29] There is no intrinsic distinction made among them, except for convention of use. Thus, there is nothing to prevent us from considering the Fourier transform as a Fourier operator, for example.

The concept of alternative representations, on the other hand, brings a new perspective to these mathematical procedures. The paradigm recognizes the independence of alternative frames of reference and thus requires that a distinction be made between those procedures whose initial and final states reside within the same frame of reference, and those procedures whose initial and final states exist in alternative frames of reference.

The procedure that takes a representation from one frame of reference to an alternative frame of reference will henceforth be called either a map or a transform and will be represented by an upper case script notation (e.g., $\mathcal{M}, \mathcal{N}, \ldots$), The procedure that alters a representational form within an alternative frame of reference will be called an operator and will be designated by an upper case letter (e.g., R, S, ...). Functional representations within an alternative will be designed by lower case (e.g., **f**, **g**, ...).

The concept of alternatives is brought into functional analysis through the requirement that if **g** is to be a valid alternative to **f**, then there can be nothing contained in **f** that does not appear in g, and conversely. This means that **f** can be mapped into **g** and then recovered from **g** everywhere (except over sets of Lebesgue measure zero) for the L^2 condition. Formally, if

$$\mathcal{M}\mathbf{f} = \mathbf{g} \tag{20}$$

then there will exist an inverse map $\mathcal{N}$ such that

$$\mathcal{N}\mathbf{g} = \mathcal{N}\mathcal{M}\mathbf{f} = \mathbf{f} \tag{21}$$

This requires that

$$\mathcal{N}\mathcal{M} = \mathcal{M}\mathcal{N} = \mathbf{I} \tag{22}$$

where **I** is the identity operator.

If there is an operation R in the space of **f** which corresponds to an operation S in the $\mathcal{M}$-transform space of **g**, then

$$\mathcal{M}\mathrm{R}\,\mathbf{f} = \mathrm{S}\,\mathbf{g} = \mathrm{S}\mathcal{M}\mathbf{f} \tag{23}$$

This leads directly to the following equivalence relations on any measurable **f**,

$$\mathcal{M}\mathrm{R} - \mathrm{S}\,\mathcal{M} = 0 \tag{24a}$$

$$\mathcal{M}\mathrm{R} - \mathrm{R}\mathcal{M} = \mathrm{S}\mathcal{M} - \mathrm{R}\mathcal{M} \tag{24b}$$

It is clear that, unless S has the identical form to R, then map M cannot commute with R; that is

$$\mathcal{M}\mathrm{R} - \mathrm{R}\mathcal{M} = [\mathcal{M},\mathrm{R}] \neq 0 \tag{25}$$

Quite simply, this means that the transform of an operator is not the same as the operator of a transform, a conclusion quite unexpected if one considers map, transform, operator and function to be synonymous terms. Like the example of the electrical circuit, this noncommuting of the procedures of transform and operator can readily be demonstrated by means of examples drawn from contemporary analysis.

Example 3

While this is an interesting result, a far more significant consequence of this concept of alternatives is the effect which is predicted when sequential operational procedures are applied to functional forms within a specified frame of reference. There is a long-standing observation of science that certain operational procedures do not commute. Thus, for example, the procedures of differentiation and multiplication by a monomial do not commute in the order of their application. This noncommutation is still considered a mystery and, to the best of my knowledge, no general explanation has been offered which would allow one to predict whether two arbitrarily chosen operational procedures will commute. I will now show the derivation of such a general relation based on this mapping concept.

Relation (23) can be cast into the forms,

$$\mathrm{R} = \mathcal{N}\mathrm{S}\mathcal{M} \qquad \mathrm{S} = \mathcal{M}\mathrm{R}\mathcal{N} \tag{26}$$

which signify that S is the $\mathcal{M}$-transform alternative of R, while R is the $\mathcal{N}$-transform alternative of S.

A symbolic diagram that is helpful in understanding the mapping relations for alternative frames of reference is shown in Figure 3. Two alternative systems are symbolized as A and B. Placing alternatives A and B on the same diagram is only a symbolic convenience which is intended to portray correspondence between procedures in these systems. An **f**(x) in system A is transformed to a **g**(y) in system B by map $\mathcal{M}$. Similarly, map $\mathcal{N}$ converts **g**(y) to **f**(x). Operator R, acting solely within A, converts **f**(x) to R**f**(x). The corresponding alternative form of this operator in B is S. If, for example $\mathcal{M}$ is the Fourier transform and R is the derivative operator on system A, then multiplication by iy is the alternative operator on system B. This symbolic diagram is quite helpful in demonstrating the distinction demanded by the principle of alternatives between the procedure of map and operator.

Figure 4 demonstrates the symbolic basis underlying the noncommuting of the procedures of map $\mathcal{M}$ and operator R.

The case of commutation or noncommutation of operational procedures is also symbolized in Figure 4. The question to be addressed is, under what conditions will operator R, followed by operator S, produce the same functional form on A as operation S followed by operation R? It should be apparent that there is absolutely no answer to this question which is available to an observer who resides solely in system A.

The geometric symbolism of the diagrams of Figures 3 and 4 can now be replaced by the more formal mapping diagram of Figure 5. This diagram shows the relationship between operations (upper case letters and lateral displacement), which act on representational forms to create a new form within the same frame of reference; and alternative maps (upper case script and vertical displacement), which transform to a new frame of reference. Each node is a representational form, with the direction of arrows indicating forward procedures; inverse procedures act contrary to the direction of the arrows. Nodes at same vertical location correspond to representational forms within the same frame of reference, while nodes at the same horizontal location correspond to alternative representations of those above and below it.

Whenever two procedures demonstrate a dependence upon each other, such as noncommu-

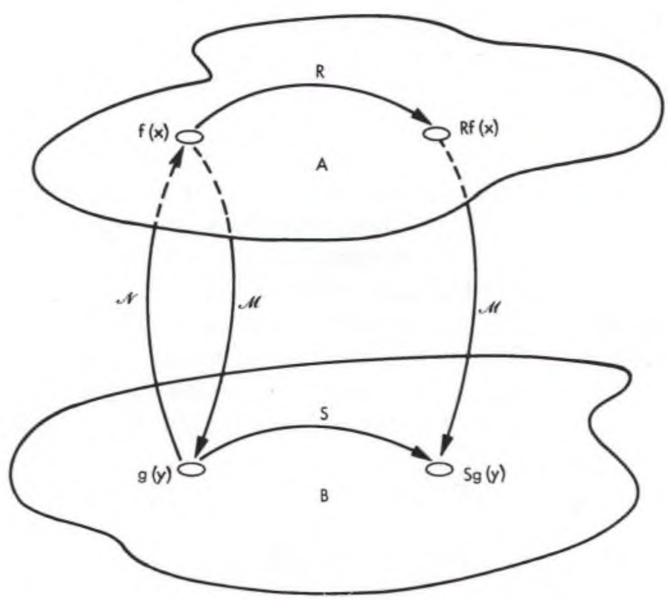


Figure 3 Symbolic diagram showing mapping relationships between alternative frames of reference.

tation in their application, then these procedures must be related to each other in some manner. I intend to show that this is indeed the case for noncommuting operators and that the underlying relationship is that they are different versions of each other under a particular linear transform.

Figure 5 shows the hierarchical relationships that exist among operators when these operators are expressed in the same frame of reference. If these are to be expressed in the same frame of reference, say System A, then operator S can be interpreted as the $\mathcal{M}$-transform of operator

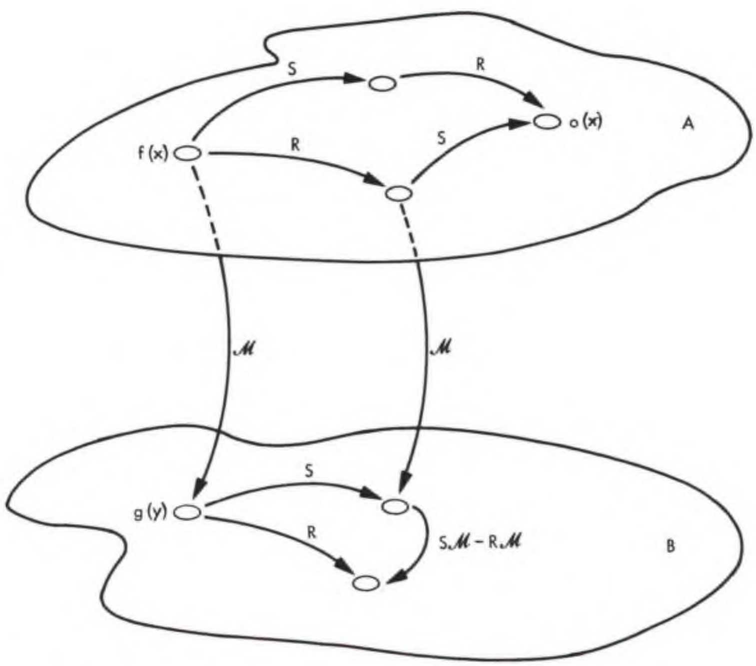


Figure 4 Symbolic diagram illustrating commutation for noncommutation of operational procedures.

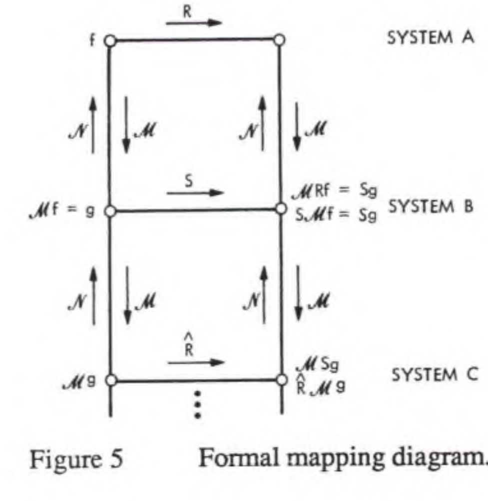


Figure 5 Formal mapping diagram.

R, while operator $\widehat{R}$ is the dual $\mathcal{M}$-transform of operator R. This can be represented as,

$$(\mathcal{M}R) = S \tag{27a}$$

$$(\mathcal{M}^2 R = \widehat{R}) \tag{27b}$$

where the parentheses indicate that the procedures are to be expressed in the same frame of reference as that of operator R.

As an example, if $\mathcal{M}$ is the Fourier transform and R is the derivative operator, then, since dual Fourier transformation reverses coordinate direction,

$$R = d \,/\, dx \tag{28a}$$

$$S = ix \tag{28b}$$

$$\widehat{R} = d \,/\, d\,(-x) \tag{28c}$$

We know from relation (25) that the procedures of transform and operator cannot commute. This means that,

$$\mathcal{M}R \neq R\mathcal{M} \tag{29}$$

Similarly,

$$\mathcal{M}RR \neq R\mathcal{M}R \tag{30}$$

The dual procedure of R followed by $\mathcal{M}$ maps a function from System A to System B. Since this involves only procedures $\mathcal{M}$ and R, there must be a unique correspondence between $\mathcal{M}$R and the single domain procedure ($\mathcal{M}$R). That is, for one $\mathcal{M}$ and one R there can only be one ($\mathcal{M}$R). This correspondence allows us to rewrite (30) as,

$$(\mathcal{M}R)R \neq R(\mathcal{M}R). \tag{31}$$

But, from (27a), this means that

$$SR \neq RS\,, \tag{32}$$

and hence R and S cannot commute. The commutator of R and S can thus be expressed,

$$RS - SR = [\,R,S\,] \neq 0. \tag{33}$$

This, I regard, as one of the stronger arguments that we are dealing with a new paradigm, a new problem solving model and set of relations. Although the noncommutation of certain operational procedures has been known for more than 60 years, and this noncommutating lies at the heart of contemporary quantum mechanics, no general reason for this noncommutation has been previously presented.[30] With the introduction of the concept of alternatives, it can now be appreciated why this failure to commute was not obvious in the ongoing paradigm: it involves mapping equivalence relationships in alternative frames of reference. We could only demonstrate the noncommutation of R and S by considering the forms they take in alternative frames of reference: S is the form taken by R under map $\mathcal{M}$ and $\hat{R}$ is the form taken by S under map $\mathcal{M}$.

Rather than two independent procedures, R and S, the mapping diagrams of Figure 6 illustrate how one can envision the dual application of R and S as equivalent to the action of only one procedure; in this case, procedure S. Figure 6a is the equivalent of procedure R followed by procedure S; while Figure 6b is the equivalent of procedure S followed by procedure R.

The result of the foregoing analysis can be readily explained in the following terms: when a first operational procedure in one frame of reference is mapped into the form of a second operational procedure in an alternative frame of reference, then the two forms of operational procedure cannot commute if applied in the same frame of reference.

Examples can readily be produced which verify the predictions of this new paradigm. Fourier related procedures cannot commute; thus the second derivative operator and multiplication by the quadratic coordinate cannot commute. Other procedures which are related to these by simple maps or operations will likewise fail to commute, thus the second derivative operator will not commute with the derivative operator acting on multiplication by the quadratic coordinate. And while convolution cannot commute with multiplication, convolution can commute with differentiation of any degree.

As a comment on the notation which I use for relation (27b), there is often a special relationship that exists between an operator and its dual transform alternative. I therefore identify the dual M-transform of R as $\hat{R}$ rather than some other form, such as T. This comes about in the following manner. As seen in the source frame of reference, a transform between the source and a second alternative frame of reference presents the coordinates of that second frame of reference as parameters which are expressed in terms of the source frame of reference. There can be transforms, of which the TDS transform is one example, for which an identical transformation on the second alternative results in a third alternative form whose coordinates are scalarly related to those of the first frame of reference.

My purpose in presenting the foregoing analysis in defense of the proposed paradigm is not only to alert the acoustic analyst to its viability in practical situations with which he may have gained much familiarity by other methods, but to the realization that some new and unexpected results can derive from its use. In particular, it is exceedingly important to release oneself from any artificial limitation on the dimensionality of a signal to be analyzed. In the next section we will consider a special class of integral transform that can be interpreted in terms of this paradigm of alternatives.

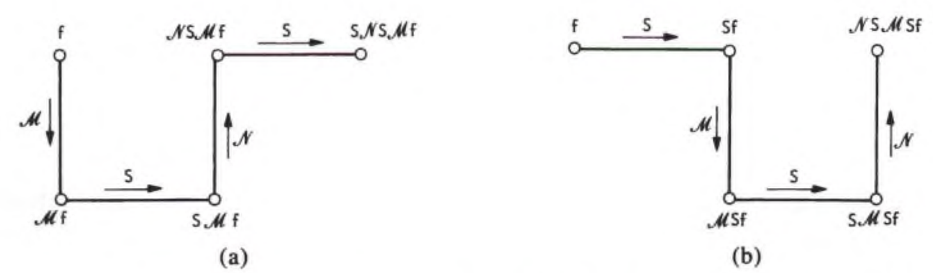


Figure 6 Two mapping diagrams. (a) Procedure R followed by procedure S, (b) Procedure S followed by procedure R.

SECTION II

MATHEMATICAL DEVELOPMENT

When we try to reduce the conceptual model to useful mathematical relationships, we are faced with a most interesting problem. What happens if we are allowed to look at something in terms of a completely different frame of reference? And in particular, what results if we convert a familiar representation, expressed in terms of a dimensionality which has meaning to us, into a frame of reference whose dimensionality is altogether different. As a heuristic example, how might a chair, which we recognize in the four-dimensional geometry of space-time, appear if we could view it in terms of an alternative six-dimensional geometry? It is apparent that in such a world, time, as we know it has no local description but must appear as a geometric relationship throughout the six dimensions. The same will be true for each of those familiar lower dimensional attributes which we call spatial location.

Our task is made even more formidable once we recognize that we cannot depend upon familiar concepts drawn from the branch of mathematics called topology. Topology considers almost every conceivable deformation, except that of changing dimension; dimension is a topological invariant.

Once we begin to appreciate that any particular view we might take is never the "only" view, we can begin to understand how to convert a representational form from one view to an alternative form in an alternative view. The touchstone, as pointed out earlier, is that no view be preferred (that is, no view can see something that cannot be accounted for in all other views), and the only property that does not change as we progress from one alternative view to another is that of total scalar measure. Starting with these fundamental considerations, this section develops some simple mathematical relationships for converting representational form among alternative frames of reference.

An Integral Transform

Any generalized procedure for mapping a representation **f** (expressed in terms of frame of reference x) into an alternative representation **g** (expressed in terms of frame of reference y) must satisfy the condition that every part of x must be uniquely accounted for in y, and conversely. This means that for every value of coordinate x there must be a unique distribution of coordinates y into which that x is to be mapped. This gives rise to the identification of a mapping kernel $\mathbf{m}(y,x)$ which is a geometric manifold involving parameters y in terms of coordinates x. If we have an $\mathbf{f}(x)$ and wish to convert it to a $\mathbf{g}(y)$ under conditions of equal square summability (class L^2), then we can use the formula,

$$\mathbf{g}(y) = \int_x \mathbf{m}(y,x)\,\mathbf{f}(x)\,dx \qquad (34)$$

$$y = y_1, y_2, \ldots y_M \qquad x = x_1, x_2, \ldots x_N .$$

$\mathbf{m}(y,x)$ is a mapping kernel that expresses certain parameters y in terms of frame of reference x. The product of this mapping kernel with any (Lebesgue) measurable $\mathbf{f}(x)$ results in a distribution over the whole of space x whose net Lebesgue sum over x leaves a representation in terms of parameters y.

If **f** is of class L^2, then by definition

$$\int_x |\mathbf{f}(x)|^2 \, dx < \infty \tag{35}$$

and if mapping kernel **m**(y,x) preserves total Lebesgue measure, then **g** is also of class L^2, and

$$\int_y |\mathbf{g}(y)|^2 \, dy = \int_x |\mathbf{f}(x)|^2 \, dx \tag{36}$$

In that case, the transform(34) is an isomorphism of L^2 onto L^2 which preserves total measure and maps space x of dimension N onto space y of dimension M, where N may be the same or different than M.

The mapping kernel identifies the way in which coordinates y will appear in space x. The parameters y become the coordinates of the transformed expression. Representations **g**(y) and **f**(x) are L^2 alternatives under map **m**.

Let us now consider one special type of mapping kernel. The mapping procedure,

$$\mathbf{g}(y) = \frac{1}{K} \int_x e^{i\phi(y,x)} \mathbf{f}(x) \, dx \tag{37}$$

where K is a constant to be determined under condition C, will be called the TDS map, and analysis performed under this mapping procedure will be called time delay spectrometry.

If **f** is complex and of class L^2, then the real and imaginary parts of **f** are also L^2. The representation **g** may be complex. The complex conjugate of **g** will be,

$$\overline{\mathbf{g}(y)} = \frac{1}{K} \int_x e^{-i\phi(y,x)} \, \overline{\mathbf{f}(x)} \, dx \tag{38}$$

If **g** and **f** are both L^2 and have the same sum, then (36) can be written,

$$\int_y \mathbf{g}(y) \, \overline{\mathbf{g}(y)} \, dy = \int_x \overline{\mathbf{f}(x)} \, \mathbf{f}(x) \, dx \tag{39}$$

This can be expressed, under conditions of uniform convergence, as,

$$\int \mathbf{g}(y) \left[\int e^{-i\phi(y,x)} \overline{\mathbf{f}(x)} \, dx \right] dy = \tag{40}$$

$$= \int \overline{\mathbf{f}(x)} \left[\int e^{-i\phi(y,x)} \mathbf{g}(y) \, dy \right] dx$$

which means that an inverse map exists and is given by,

$$\mathbf{f}(x) = \frac{1}{K} \int_y e^{-i\phi(y,x)} \mathbf{g}(y) \, dy. \tag{41}$$

Or, a forward TDS map followed by an inverse TDS map restores the original form under the relation,

$$\mathbf{f}(x) = \frac{1}{K^2} \int_y e^{-i\phi(y,x)} \int_x \mathbf{f}(x)\, e^{-i\phi(y,x)}\, dx dy \tag{42}$$

The factor K^2 is that which is required to restore **f** under a dual procedure of a forward map followed by an inverse map. Relations (37) and (41) are forward and inverse TDS maps, respectively.

This shows that the TDS map is invertible and hence is a legitimate transformation between coordinates. An **f** can be changed into a **g**, and then recovered as an **f** (everywhere, except over sets of Lebesgue measure zero). What we have done is to break the (artificial) bond of dimensionality from analysis. We are no longer restricted to working at a fixed level of dimensionality, but can transform up or down in dimensionality of representation. This is the key that lets us into other domains.

TDS Mapping Relationships

Iterating the TDS map, that is, applying two forward maps, will not restore the original function. In particular, if $\mathbf{h}(u)$ is the result of two forward maps, then, neglecting the scalar constant K,

$$\begin{aligned} \mathbf{h}(u) &= \int_y e^{i\phi(y,u)} \left[\int_x \mathbf{f}(x)\, e^{i\phi(y,x)}\, dx \right] dy \\ &= \int_x \mathbf{f}(x) \left[\int_y e^{i\{\phi(y,x) + \phi(y,u)\}}\, dy \right] dx \,. \end{aligned} \tag{43}$$

Another relationship which can be demonstrated is the TDS transform of products. If,

$$s(y) = \int e^{i\phi(y,x)}\, \mathbf{r}(x)\, dx, \tag{44}$$

then the transform of the product of two functions in y is

$$\begin{aligned} \mathbf{p}(x) &= \int g(y)\, s(y)\, e^{-i\phi(y,x)}\, dy \\ &= \int g(y)\, e^{-i\phi(y,x)} \left[\int e^{i\phi(y,u)}\, \mathbf{r}(u)\, du \right] dy \\ &= \int \mathbf{r}(u) \left[\int g(y)\, e^{-i\{\phi(y,x) - \phi(y,u)\}}\, dy \right] du \\ &= \int \mathbf{r}(u)\, \mathbf{f}^{\dagger}(x,u)\, du \end{aligned} \tag{45}$$

where

$$\mathbf{f}^{\dagger}(x,u) = \int \mathbf{g}(y)\, e^{-i\{\phi(y,x) - \phi(y,u)\}}\, dy \tag{46}$$

The TDS transform of the product of two functions is a special integral of the separate transforms of those functions.

Not surprisingly, the TDS transform of the derivative of a function has a special. form. This can be seen from the expansion—

$$\frac{d}{dx}\left[e^{i\phi(y,x)}\, \mathbf{f}(x)\right] = e^{i\phi(y,x)} \frac{d\,\mathbf{f}(x)}{dx} + i\, \frac{d\phi(y,x)}{dx}\, e^{i\phi(y,x)}\, \mathbf{f}(x)\,. \tag{47}$$

Taking the bounded integral with respect to x, and allowing the integration limits to grow without limit,

$$\lim_{X\to\infty} \int_{-X}^{+X} d\left[e^{i\phi(y,x)}\, \mathbf{f}(x)\right] = \lim_{X\to\infty} \int_{-X}^{+X} e^{i\phi(y,x)} \left[\frac{d\mathbf{f}(x)}{dx}\right] dx \tag{48}$$

$$+ \lim_{X\to\infty} \int_{-X}^{+X} i \left[\frac{d\,\phi(y,x)}{dx}\right] \mathbf{f}(x)\, e^{i\phi(y,x)}\, dx$$

which becomes the relation,

$$\int e^{i\phi(y,x)} \frac{d\mathbf{f}(x)}{dx}\, dx = -i \int e^{i\phi(y,x)} \left[\frac{d\phi(y,x)}{dx}\right] \mathbf{f}(x)\, dx\,. \tag{49}$$

The TDS transform of the derivative of a function is minus the imaginary operator times the TDS transform of the product of that function and the derivative of the phase of the TDS kernel.

If we define the transform of $\mathbf{f}(x)$ as $\Phi\,[\mathbf{f}(x)]$, under the relation,

$$\Phi[\,f(x)] \triangleq \int e^{i\phi(\xi,x)}\, f(x)\, dx = F(\xi), \tag{50}$$

then

$$\Phi\left[\frac{d\mathbf{f}(x)}{dx}\right] = -i\,\Phi\left[\,\mathbf{f}(x)\cdot\frac{d\phi(\xi,x)}{dx}\right]. \tag{51}$$

If, in particular,

$$\phi(\xi,x) = \xi x^{r}, \tag{52}$$

then the following transform relations result

$$\Phi\,[\mathbf{f}'(x)] = -i\,\xi\, r\, \Phi\left[x^{r-1}\,\mathbf{f}(x)\right] \tag{53a}$$

$$\Phi\left[x^{1-r}\,\mathbf{f}'(x)\right] = -i\,\xi\, r\,\Phi[\mathbf{f}(x)] \tag{53b}$$

$$\Phi\left[g^{\cdot}(\xi)\,\mathbf{f}(\xi)\right] = \int \mathbf{G}\,(u)\,\mathbf{F}\,(x^r - u^r)\,du \tag{53c}$$

$$\Phi\left\{\Phi\left[f(x)\right]\right\} = \Phi^2\left[f(x)\right] = \mathbf{f}\left(\sqrt[r]{-1}\,x\right) \tag{53d}$$

$$\Phi^{2r}\left[\mathbf{f}(x)\right] = \mathbf{f}(-x) \tag{53e}$$

$$\Phi^{4r}\left[\mathbf{f}(x)\right] = \mathbf{f}(x) \tag{53f}$$

Relations (53d), (53e) and (53f) show that simple polynomial TDS transforms are 4r-fold cyclic. Forward application of the map 4r times brings us back to our starting frame of reference, where r is the exponent of the TDS kernel.

The Special Case Called Fourier Transform

Let us now consider the significance of the TDS map. The mapping kernel has the form,

$$\mathbf{m}\,(y,x) = e^{i\,\phi\,(y,x)} \tag{54}$$

Geometrically, $\phi\,(y,x)$ = constant, is a hypersurface in x, defined in parameters y. The nature of this hypersurface is established by the parameters y. As a special case, if the source frame of reference is N-dimensional in x and the destination frame of reference is also N-dimensional, then the number of parameters y will equal the number of source coordinates x. The simplest hypersurface having one parameter for every x is the *hyperplane,* defined by the inner product,

$$\langle y,x\rangle = y_1x_1 + y_2x_2 + \ldots + y_Nx_N \tag{55}$$

The hyperplane is "flat" in Euclidean space; it has no curvature. A hyperplane in one-dimension is a point; in two-dimensional space it is a straight line; and in three-dimensional space is a plane. One could identify the parameters y as the inverse of "direction cosines", whose values determine the "angle" of the hyperplane in space x. As the y's pass through all their values, the hyperplane is N-dimensionally rotated through all its possible "angle". This procedure is formalized by expressing the hyperplane in the role of an angle by placing it in a complex exponent. When the N-dimensional function **f**(x) is multiplied by the exponential hyperplane, the product is a distribution over the whole of x in terms of the parameters y. Even if **f**(x) exists only at a "point" in x, the distribution is i over the whole of x. Summing, with Lebesgue measure, all x will completely remove any x-dependence, leaving the N-dimensional representation in terms of y.

It should be clear that what we have just developed is the N-dimensional Fourier transform:

$$\mathbf{g}\,(y) = \left(\frac{1}{2\pi}\right)^{N/2} \int e^{i\langle y,x\rangle}\,\mathbf{f}\,(x)\,dx \tag{56}$$

It also follows that the inverse Fourier transform is,

$$\mathbf{f}\,(x) = \left(\frac{1}{2\pi}\right)^{N/2} \int e^{-i\langle y,x\rangle}\,\mathbf{g}\,(y)\,dy\,. \tag{57}$$

And dual iteration produces,

$$h(u) = \int f(x)\left[\left(\frac{1}{2\pi}\right)^N \int e^{i<y,\,(x+u)>}dy\right]dx$$
$$= \int \mathbf{f}(x)\,\delta(x+u)dx = \mathbf{f}(-u) \tag{58}$$

Using our previously derived expression (46) we can see that the Fourier transform of products equates to the convolution of the Fourier transforms, under,

$$f^{\dagger}(x,u) = \left(\frac{1}{2\pi}\right)^{N/2} \int g(y)\, e^{-i\,\{<y,x>-<y,u>\}}dy\,.$$
$$= \mathbf{f}(x-u)\,, \tag{59}$$

and the Fourier transform of the derivative of a function is the Fourier transform of the function times the transform coordinate multiplied by the negative imaginary unit.

Relations (56), (57), (58) and (59) are well known properties of the Fourier transform in L^2. My purpose in deriving them is to show how the concept which we are presenting in this TDS approach can: (1) lead to an entirely new interpretation of an existing procedure, and (2) lead to new concepts.

The analysis tells us the following about the Fourier transform in L^2 that is, when used to describe situations under the law of conservation of total measure (energy or probability):

[1] The representations **f**(x) and **g**(y) are different descriptions of the same thing, and hence are different descriptions of each other. It is not possible to form a legitimate 2N-dimensional representation in L^2 by adding y onto x. When applied to the simplest expressions in signal theory, this means that it is absolute nonsense to talk about the moment in time when a frequency component appears. The mutual spreading often referred to as the uncertainty relation is due to nothing more complicated than the definition of the properties involved.

[2] The frame of reference x not only has the same dimensionality as y, but there is a pairwise coordinate association such that everything expressible along the i^{th} coordinate in x is found only along the corresponding i^{th} coordinate in y. (Remember, this derivation from the concept of alternatives directly produced the general N-dimensional Fourier transform. We did not have to start, as is conventional practice, from a definition in one-dimension, then try to expand upward in dimensionality.)

[3] Each "point" in x is mapped to a special geometric figure spread uniformly over all of y, and conversely. This figure has the form of the mapping kernel,

$$e^{\,i<y,\,x>}$$

The geometric interpretation of this figure is such that, henceforth, we shall call it a *zero curvature wave*. There is thus not only a point-wave duality between Fourier transform alternatives, but the wave is of a special type from hyperplanes of zero curvature.

[4] From considerations of energy, both **f**(x) and **g**(y) are complex in N-dimensions, with a special combination of real and imaginary components for each coordinate and with each complex pair related by Hilbert transform. The zero curvature wave, in conjunction with the forward and inverse transform relations, gives a special geometric property to expressions in terms of those paired complex representations in x and y. Along the i^{th} coordinate, representations in **f**(x)

will curl clockwise with increasing values of x, while representations in g(y) will curl counterclockwise with increasing values of y. In the special dimensional form used for signal analysis, where,

$$h\,(t) = \left(\frac{1}{2\pi}\right)^{\frac{1}{2}} \int_{-\infty}^{\infty} H\,(\omega)\, e^{i\,\omega\, t}\, d\omega, \tag{60}$$

the Nyquist plot of $H(\omega)$ will curl clockwise with increasing frequency, while the Nyquist plot of h(t) will curl counterclockwise with increasing time. Furthermore, we must now recognize that the time response is also complex, having a real part and imaginary part related by Hilbert transform.

A most interesting technical point can be seen by investigating the mapping relation (53f) as it relates to the Fourier transform. The Fourier transform must be four-fold cyclic by virtue of its zero curvature hyperplane kernel. There are thus four forms into which a representation can be mapped by the Fourier transform,

$$f\,(x) \rightarrow F\,(y) \rightarrow f\,(-x) \rightarrow F\,(-y) \rightarrow f\,(x) \tag{61}$$

A quadratic kernel, where r is two, corresponds to the simplest quadratic phase "chirp" used in contemporary TDS. The transform using this quadratic kernel is eight-fold cyclic, and yields the forms,

$$\begin{aligned} &f\,(x) \rightarrow g\,(y) \rightarrow f\,(ix) \rightarrow g\,(iy) \rightarrow f\,(-x) \rightarrow g\,(-y) \\ &\rightarrow f\,(-ix) \rightarrow g\,(-iy) \rightarrow f\,(x) \end{aligned} \tag{62}$$

As one continues this process with higher curvature kernels, the number of alternative spaces rapidly increases, but the interesting point is that all of them can reach the simple alternatives reached by the Fourier transform. A by-product is that we must not only recognize that functional forms within an alternative are complex, but that some alternative frames of reference can be recognized as complex versions of others. Quite simply, this means that with quadratic and higher curvature kernels we can cast a simple expression h(t) into h(it). We can, thus not only have complex functions of time but complex functions of complex time.

Since the purpose of this paper is the introduction of the general TDS map, we will not dwell further on this newer geometric interpretation of the Fourier transform. Suffice it to say that this gives us a much more powerful insight.

SECTION III

IMPLEMENTATION OF THE TRANSFORM

The basic concept underlying TDS is symbolized in Figure 7, Three alternative frames of reference are symbolized, referred to as the x-, y-, and z-domains. Each is a valid and complete system for the representation of an observable process. For the purpose of this paper, a process will be said to be observable if it has finite Lebesgue measure as expressed in terms of the coordinates of its particular frame of reference. As we have seen, this condition leads to a particular form of analytic representation for the observable energy density of such a process. An observer determines which particular alternative system he will utilize by means of his choice of frame of reference with its coordinates of representation. He may exit to another alternative (in effect, take an alternative view) only by means of a map (or transform) which preserves total Lebesgue measure. This assures that the total energy of the process, as determined by an observer in the x-domain, is the same for every alternative domain which he may choose.

Referring to Figure 7, there may be some property P which is of interest when expressed in terms of the x-domain. Unfortunately, there may be some other property Q which overlaps P and interferes with an accurate determination of P. If the map $\mathcal{M}$ is a dimension-preserving map, such as the Fourier transform, it may happen that P', the alternative form of P as expressed in the y-domain, is also overlapped by Q', the y-domain alternative to Q. In that event, the best that one may be able to do is map from the x-domain to the y-domain, separate (filter) as much of Q' as overlaps P', and map the results back to the x-domain to observe the consequent filtered form of P.

It may happen that there is a higher dimensional alternative representation, symbolized here as the z-domain, in which the transformed version of P (shown as P") has a much more complete separation from the corresponding transformed version of Q. This (presumed) higher dimensional alternative representation may be abstract and unimaginable to a person accustomed to either the x-domain or y-domain alternatives. Be that as it may, a filtering process in the z-domain can perform the surgical equivalent of resecting the undesired Q" from P". This filtered P" may still be unintelligible while expressed in terms of the z-domain alternative frame of reference. It is then necessary to map P" back to either of its (lower dimensional) alternative forms, P or P', for a

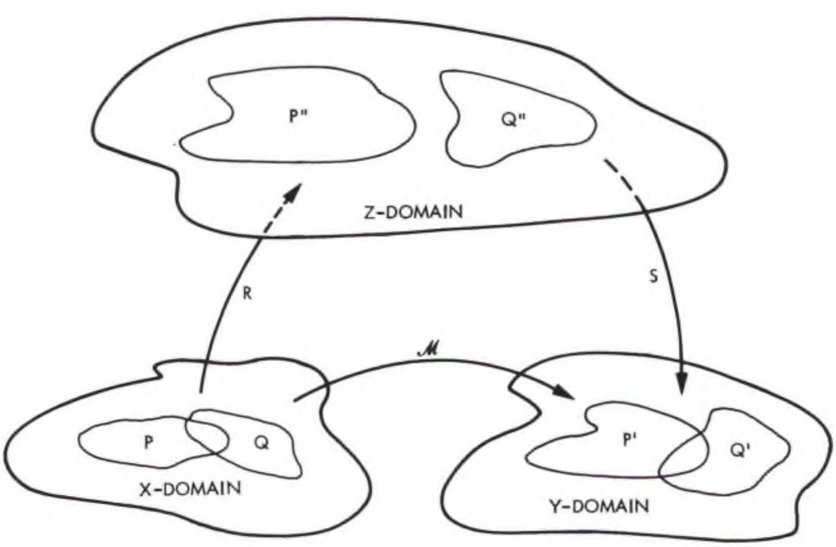


Figure 7 Symbolic diagram showing three alternative frames of reference.

presentation that is understandable to those whose frame of reference is x or y. As we shall see, time delay spectrometry is a methodology for accomplishing this process of mapping and filtering.

Time delay spectrometry is the name applied to a process consisting of four basic steps. First, an excitation signal is applied to a system which is to be analyzed; this signal has the specific form of a complex representation whose amplitude is constant and whose phase varies in a predetermined manner in accordance with a set of phase parameters. The term "phase" refers to the relationship between the two components f(s) and g(s) of the energy descriptor (7) as expressed in exponential form. Second, the response of the system (to which the excitation is applied) is interpreted in a frame of reference which is uniquely related to the phase parameters; that is, the phase parameters of the excitation signal determine the coordinates of this new frame of reference. Third, a filtering process is performed in which a desired component of response is separated in this new frame of reference. Fourth, the filtered result is mapped to an appropriate frame of reference for interpretation; this may, for example, be the original frame of reference which was utilized for system excitation.

The process of TDS is symbolized in Figure 8. An excitation signal exp $i\phi(\xi,x)$ is applied to a system whose impulse response (in terms of coordinates x) is h(x). This elicits a response $H(\xi,t)$, which is mathematically expressed in terms of the convolution integral involving h(x) and exp $i\phi(\xi,x)$. This is a key step to understanding how the concept of TDS is implemented. Traditional analysis attempts to maintain the same frame of reference for $H(\xi,t)$ as for h(x). This means that in traditional analysis the convolution integral is summed over all x and presented as a representation in terms of coordinate t with parameter ξ. Since t and x are of like units, the response is often interpreted in terms of an offset t in coordinates x.

It is a special property of the constant amplitude excitation signal that the normal convolution integral can be directly interpreted as the special transform used in TDS. Thus the system response would be observed as a signal expressed in terms of x to a person whose frame of reference was x, and as a signal expressed in terms of ξ and t to a person whose frame of reference was the higher dimensional ξ and t. In TDS analysis the response $H(\xi,t)$ is interpreted in terms of a new set of coordinates ξ and t.

Suppose there were some component of response P(x) which was of special interest in the frame of reference x. This P(x) will take the form $P''(\xi,t)$ in the higher dimensional response frame of reference. (Note that for illustrative purposes the example assumes x to be of lower

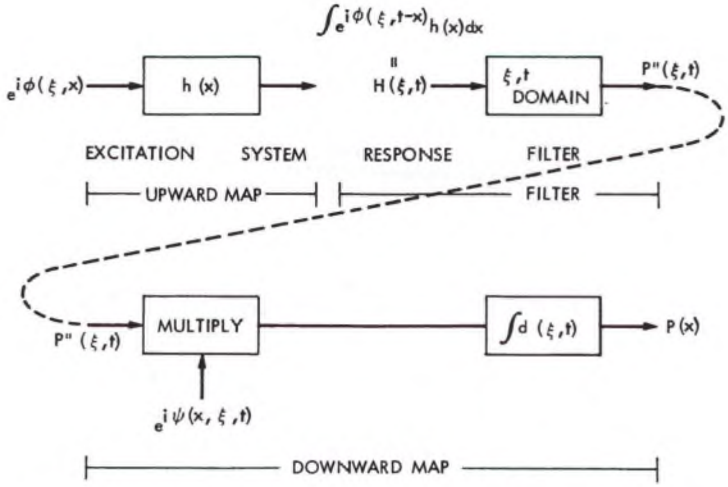


Figure 8 Symbolic representation of the TDS process.

dimensionality than (ξ,t) but this is not necessary; x may be the same or higher dimensionality than (ξ,t)). There are an infinite number of excitation signals of different functional dependence φ that one may employ for TDS. The particular excitation signal to use would be one where there is maximum separation of P"(ξ,t) from some identifiable interference Q"(ξ,t) which is mapped upward from the interference Q(x). Once filtered, the P"(ξ,t) may then be mapped down to the frame of reference x, as shown in Figure 7. Or it may be mapped to some alternative frame of reference if desired. The downward mapping is precisely the converse of the upward mapping process.

Block Diagram Analysis

Figure 7 outlined the abstract process for an overall discussion of TDS, whereas Figure 8 diagrammed an elementary mathematical interpretation of the process for purposes of spatial filtering. Figure 9 is a block diagram of a physical embodiment of the process which is capable of performing several simple acoustic measurements in the time and frequency domains. Although there are differences in detail of implementation, Figure 9 characterizes the presently available TDS instruments now being used in architectural acoustics, loudspeaker testing and medical ultrasound.

Referring to Figure 7, the implementation of Figure 9 equates time with x, frequency with y and a two-dimensional delay plane with z. In terms of Figure 7, $\mathcal{M}$ is the Fourier transform while R and S are upward and downward TDS transforms, respectively.

A constant amplitude linear phase rate sweep,

$$w(t) = e^{i(a/2)t^2} \tag{63}$$

is applied to the network whose impulse response is s(t). This elicits the response,

$$r(\tau) = \int_{-\infty}^{\infty} w(\tau - t)\, s(t)\, dt \triangleq w(\tau) \otimes s(\tau), \tag{64}$$

where the symbol ⊗ represents convolution. When we expand this convolution integral, it takes the form of the TDS integral.

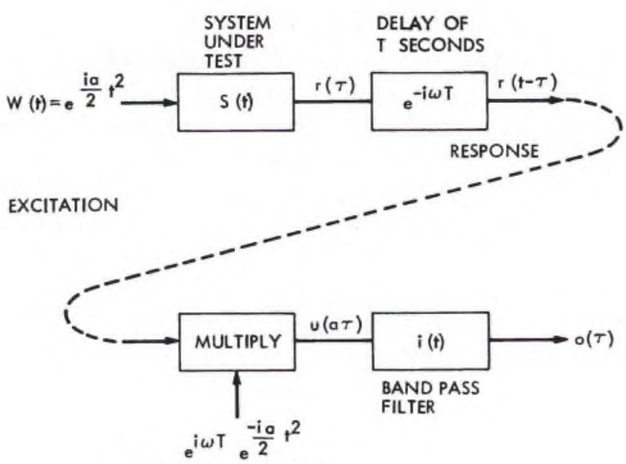


Figure 9 Implementation of TDS in currently available commercial instrumentation.

$$r(a,\tau) = \int_{-\infty}^{\infty} e^{i(a/2)(\tau-t)^2} s(t)\, dt\,. \tag{65}$$

This integral can also be expressed, in this special case, in the form of a Fourier transform,

$$\begin{aligned} r(a,\tau) &= e^{i(a/2)\tau^2} \int_{-\infty}^{\infty} s(t)\, e^{i(a/2)t^2} e^{-ia\tau t}\, dt \\ &= e^{i(a/2)\tau^2} \left[S(a\tau) \otimes W(a\tau) \right] \end{aligned} \tag{66}$$

where $S(a\tau)$ is the Fourier transform of $s(t)$ and $W(a\tau)$ is the Fourier transform of $w(t)$. The response $r(a,\tau)$ is a two-parameter function of the sweep parameter a and a delay τ. If our interest is in obtaining the frequency response $S(\omega)$ of the network whose impulse response is $s(t)$, it is available in this dual-parameter form by mapping from $a\tau$ to ω.

The interpretation to be placed on the parameter τ can more clearly be seen by considering the case in which a fixed delay of T seconds is experienced by the signal which emerges from the network. In this case the response becomes,

$$r'(a,\tau) = w(t-T) \otimes s(t-T) = r(a,(\tau+T)) \tag{67}$$

Thus the parameter τ is directly related to time delay in the response of the network. This means that the two-parameter response has one parameter corresponding to time delay and the other parameter corresponding to frequency divided by delay under the relation,

$$a\tau = \omega \tag{68}$$

Extracting frequency information about this network in the presence of time delay will thus consist of passing the information through a process in which components on the (a,τ) domain are accepted when their trajectory corresponds to the appropriate frequency relationship. The physical embodiment of this process is that of a delay tracking filter whose frequency of acceptance is that of relation (68) and whose frequencies of rejection are those of delay components occurring at specified periods prior to and subsequent to the response to be accepted. This delay tracking filter can be implemented as shown in Figure 9. The network response is passed through a process which first multiplies this response by a complex conjugate sweeping signal $w(t)$ which is delayed by the time T. In effect, this produces a frequency translation of the sweeping response signal into a steady frequency response. The complex amplitude and phase of this steady response contains the network frequency response as a special modulation. All components of the network responses of all delays are present as modulation on this multiplier output. The process of extracting the desired delay component consists of passing the multiplier signal through a filter which has its own impulse response $i(t)$. The output of this filter can be interpreted as the convolution of three terms. Formally, the time delayed response coming out of the network is,

$$r(a,\tau-T) = e^{-i\omega T} e^{i(a/2)\tau^2} \left[S(a\tau) \otimes W(a\tau) \right]. \tag{69}$$

When multiplied by the delayed and conjugated sweeping excitation, this produces

$$u(a,\tau) = \left[S(a\tau) \otimes W(a\tau) \right], \tag{70}$$

which, when passed through the filter with impulse response i(t), results in

$$o(\tau) = \left[i(a,\tau) \otimes W(a\tau) \otimes S(a\tau) \right] \tag{71}$$

Under the associated properties of convolution, this results in,

$$\left[i(a,\tau) \otimes W(a\tau) \right] \otimes S(a\tau) = A(a\tau) \otimes S(a\tau). \tag{72}$$

What this means, quite simply, is that the desired frequency response $S(a\tau)$ appears as a term which is convolved with a filter of known response. Since both the excitation signal w(t) and the filter response i(t) can be accurately determined, there is no distortion of $S(a\tau)$ which is not under the control of the experimenter who wishes to extract $S(a\tau)$ from interfering signals which arrive earlier or later than the component from which $S(a\tau)$ can be determined.

This is the desired response of this simple time delay spectrometer. A signal w(t) which is of constant energy with time dependent phase rate is fed to a network to be analyzed. The signal in passing through this network will undergo a spectrum change by $S(a\ \tau)$ and suffer a time delay T. The signal thus emerging from the network is multiplied by the complex conjugate of the original signal delayed by a time T. This product is passed through a narrow bandwidth filter. The result is that the output signal is the Fourier transform of the impulse response of that network with the delay T. The time domain selectivity producing the apodization of the desired signal, $A(a\ \tau)$, is performed by the joint action of filter bandwidth and excitation. Signals with an earlier or later arrival time than T will be suppressed by $A(a\ \tau)$, while signals arriving at time T will be shaped (apodized) in a manner producing the best representation of $S(a\ \tau)$ consistent with the suppression of improper arrival times. In other words, we can choose the network with response i(t) such that the desired response $S(a\ \tau)$ is apodized in whatever manner we wish.

The foregoing simplified analysis shows how the frequency response of a network of known time delay can be extracted from the two-parameter domain $(a\,\tau)$. It may also be of interest to extract the time response of all signals which come out of the network, and do so for a selected band of frequencies, even though the excitation signal elicits a response for all frequencies. This can be accomplished as shown in Figure 10.

In this case, the filter, whose impulse response is i(t), may now be a very broad bandwidth filter, or even no filter at all. What we wish to do is sum all terms along the delay parameter τ which are contained within a defined range of delays. The action of the sweeping excitation signal is to produce a response in which time delay appears as a frequency offset, relation (68). This means that we can select a particular time delay epoch by selecting the appropriate frequency components in the signal of relation (72). Referring to Figure 10, the response $A(a\tau) \otimes S(a\tau)$ is multiplied by a cisoid representing the desired delay epoch Ω, and integrated from the beginning of the sweep w(t) to its end. The output of the integrator at the end of this period, $0(\Omega)$, corresponds to the Fourier transform of the signal $A(a\tau) \otimes S(a\tau)$.

What is particularly interesting about this time response is that not only is the full analytic signal s(t) expressed, with its inphase and quadrature time components in terms of parameters (Ω/a), but it appears multiplied by an apodization term.

Practical Instruments

Figure 11 shows the block diagrams of two practical methods of instrumenting a time delay spectrometer of the special type outlined in the previous section Figure lla is the implementation

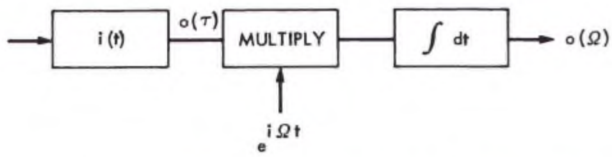
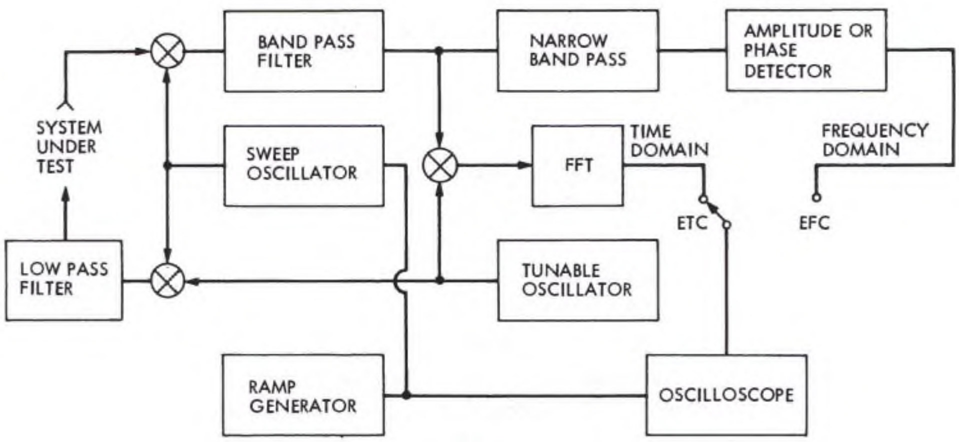
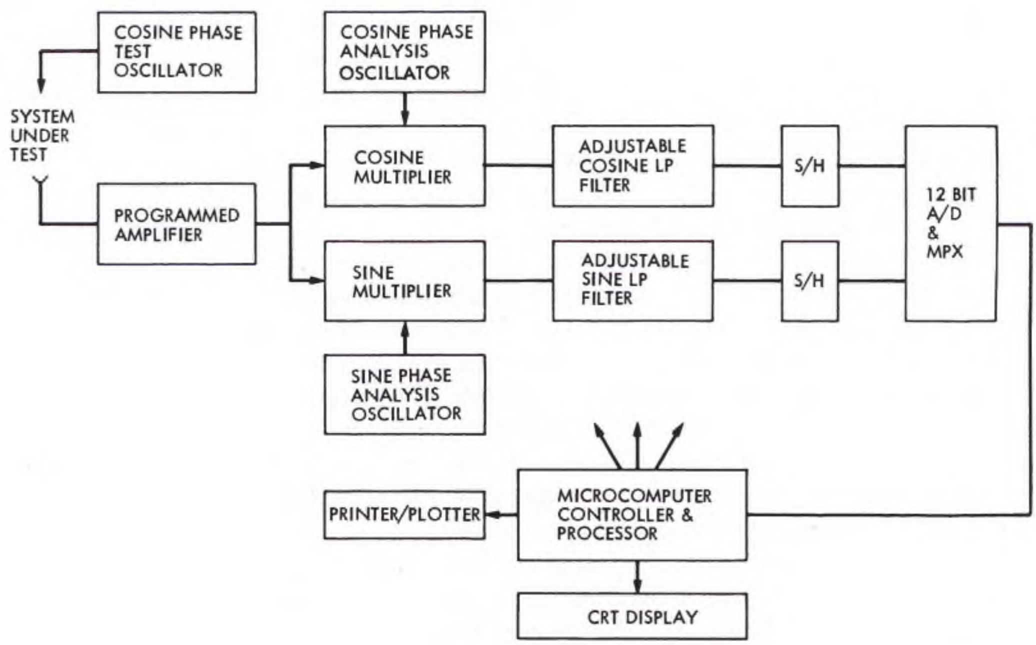


Figure 10 Extraction of desired signals in TDS system.



(a)



(b)

Figure 11 Implementation of two commercial TDS systems. (a) Bruel and Kjaer Model 5842, (b) Tecron Model 10.

chosen by Bruel and Kjaer in the Model 5842, whereas Figure llb is the implementation chosen by Tecron in their Model 10 analyzer, both of which are commercial versions of time delay spectrometers now being used in acoustic measurements.

Both of these commercial instruments are practical physical measuring devices which produce results that can readily be interpreted by the acoustic experimenter in terms with which he is familiar. Both instruments are quite simple to reconcile in terms of the time domain and frequency domain measurements which they perform; and without the foregoing abstract it might normally be presumed by an experimenter that he understands all there is to know about time delay spectrometry solely by knowledge of time domain and frequency domain properties.

By presenting some of the underlying multi-domain theory of TDS it should now be apparent why some of the practical measurements produced by these instruments are of a type not normally considered in acoustic analysis. For example, the time domain response of a network can be displayed by TDS as its traditional impulse response; however, it can also be presented as a complex time response, consisting of the traditional impulse response and a conjugate term which is the Hilbert transform of the impulse response. The full complex time response, which was introduced as the analytic signal by Gabor,[4] is a more useful acoustic characterization than the simple impulse response since, as we have seen, it relates to energy density and its partitioning into the two terms which must exist for causal measurements. Because of the partial confusion already existing concerning this particular measurement, a more thorough description of the basis of what I have called the energy time curve (ETC) is in order.

Energy-S Curves

In the most general sense, h(s), relation (7), contains all the energy information as a function of coordinate s. h(s) is the energy function in terms of s, and we could call it by that name, but such a title is a bit awkward. A more convenient title would be the *energy-s function*, with the acronym ESF. The energy function takes on a particularly useful form if its logarithm is displayed as a function of s, called the *energy-s curve* (ESC). The reason I call it a curve, as opposed to function, relates to its method of display. When s is the frequency coordinate, we have the EFC and when s is the time coordinate, we have the ETC. In those anticipated cases where s is some other coordinate(s) we can still use the acronym with whatever coordinate names we choose to use. The ESC (which includes the ETC) is a full complex representation consisting of a logarithmic amplitude $\alpha(s)$ and a phase $\phi(s)$ under the relation

$$\begin{gathered} h(s) = f(s) + ig(s) = \exp[\,\alpha(s)\,] \cdot \exp i\big[\phi(s)\big] \\ \ln[\,h(s)\,] = \alpha(s) + i(\phi)s \end{gathered} \tag{73}$$

We already know the EFC through conventional analysis. The logarithmic amplitude of the frequency function is commonly called amplitude, while the phase of the frequency function is simply referred to as its phase. Even with the newer energy interpretation, referring to the frequency response as an EFC would, in my opinion, be a needless neologism.

The time response, however, has not had a traditional interpretation as a complex function. In that case, the use of ETC is warranted since there is no available competing terminology.

I must point out that the ETC has both an amplitude component and a phase component. Because conventional FFT analyzers, of the type used in present TDS instrumentation, principally display the amplitude of the spectrum, the present ETC displays are that of the amplitude component, and a number of misinterpretations have already been published in which the ETC was considered only to be the amplitude of the analytic signal. This is not so.

Some concern has also been expressed that the analytic signal, which we have shown is an energy functional, is noncausal, requiring the use of a Hilbert transform to compute the quadrature of the component g(t) from f(t). This point requires some elaboration.

Hilbert Transform in the ESC

The seemingly abstract concept of multiple dimension alternative has led us to a practical set of relations. Under linear conditions, no matter what dimensionality of coordinate system we might choose, the observation of energy density must have two quadrature related components along each coordinate. Furthermore, the relationship between each quadrature pair of components will be that of the Hilbert transform.

Consideration of alternative frames of reference allows us to place an interpretation on these two components which might otherwise seem mysterious. The Hilbert transform utilizes a kernel which extends over the entire range of coordinate. Thus, when applied to the traditional time domain, the appearance of a Hilbert transform relationship is often misinterpreted to represent some nature of noncausality in the quadrature component. We can understand what is involved by returning to the concept of frame of reference. Having set up a frame of reference for observations of f(s) we can observe f(s) and *only* f(s). But the *support* of f(s), the *global* distribution in f(s) that is required in order for a local value to be what it is, requires a conjugate entity g(s), which we cannot observe in the V(s) frame of reference.

It is not difficult, therefore, to appreciate the significance of the complex representation in the time domain (or in any other domain). At each moment in time there is the equivalent of an imbedded total energy content associated with "real" observation. The "real" observation is a one-dimensional projection of a rotating two-dimensional phasor which carries this embedded total content. We, who observe this lower-dimensional "real" projection, are like "Flatlanders" who can see a higher-dimensional entity only when it passes through the framework of our lower-dimensional space, and whose total aspects we can fully discern only when it lies wholly in our space and nowhere else. Nor can we tell from a moment's glance whether what we see represents all of this entity or only that fragment caught passing through our lower-dimensional vision. But, there are rules governing the higher-dimensional movements of this entity, and if we carefully note what we see, moment by moment, we can determine, from everything that has gone before, where this entity lies and how large it is. Thus, the fact that an acoustic pressure is at its equilibrium value at a moment in time is not sufficient to state that there is no sound at that moment; all we can say is that there is no potential energy density at that moment, much as a snapshot of a scene cannot tell us whether objects are in motion. In order to determine how much kinetic energy is required to present this null pressure value at this moment, we must look at the distribution of sound pressure at other moments. If we start from the time domain, we must use the Hilbert transform to compute this quadrature component. If we start from some other domain and map to the time domain, the quadrature component is automatically available to us since the alternative domain includes all time in its representation. In the case of TDS measurements, the finite spread of bandwidth in the frequency measurement translates to a finite spread of interval in the alternative time measurement, and the analytic signal is smeared to this extent.

SECTION IV

SOME PRACTICAL APPLICATIONS OF TDS

No theory, however structured, could be expected to be of interest to an analyst unless practical applications could be made of that theory. In this section I will briefly discuss a number of practical applications of TDS which have already been made in the field of acoustics. These particular examples are chosen for two reasons. First, they illustrate the use of TDS to solve specific problems in acoustics which, in many cases, would be difficult to solve using traditional methods. Second, the data which they contain is presented here for the first time and cannot be found elsewhere in the literature.

Loudspeaker Response

The measurement of the free field response of a loudspeaker, while that loudspeaker is positioned in a room with acoustically reflecting boundaries and with substantial ambient noise, represents a microcosm of the difficulties of general transducer measurement. What is said about a loudspeaker in such an environment may also be said about sonar projectors in realistic situations.

For the purpose of equivalent freefield measurements, not only must the direct sound from loudspeaker to microphone be isolated from interfering wall reflections, but the experimenter may be confronted with additional technical difficulties. For example, there may be extraneous acoustic signals which interfere with the measurement. Even a good anechoic chamber may not be a quiet room. It may even be desired to measure the response of a loudspeaker while that loudspeaker is simultaneously reproducing other program material, or, for cross modulation distortion measurements the interfering program may be related to the test signal itself.

The experimenter must also recognize that phase, as well as amplitude, needs to be measured in order to characterize the frequency response of the loudspeaker. Thus, the speed of propagation of the sound wave and the effective acoustic location of both the loudspeaker and measuring microphone must be corrected for air path time delay. Since there is no guarantee that a loudspeaker (or, for that matter, a measuring microphone) is of minimum phase characteristic in its frequency response, the experimenter must exercise caution in the choice of air path delay for a phase response measurement. This problem could be the subject of a technical paper in its own right and will not be further elaborated here, but one aspect of the problem is discussed by Heyser.[31]

There is the additional consideration that either, or both, of the transducers may be in effective motion while the measurement is in process. In a normal static room measurement performed in real time this effect may be quite minute, caused principally by thermal gradients and wind. But playback from a previous recording which has time base errors, such as analog magnetic tape or disk records, may contain such errors; and measurement of a sonar transducer which is mounted on a movable platform will certainly contain such errors.

I mention these problems not to discourage experimenters, but to point out that proper measurement of a loudspeaker may require more of the experimenter than merely placing the microphone in the proper position and pushing a button.

Electroacoustic Transfer Response

For the past decade this author has been the reviewer of loudspeakers for AUDIO magazine and has published TDS measurements of the ETC and of the amplitude and phase response, corrected for air path time delay, of those loudspeakers. Those measurements could, with appropriate facilities, be duplicated by testing modalities other than TDS. However, there are classes of measurement which are rather uniquely handled by TDS. One of these will now be discussed.

The electroacoustic transfer response of a loudspeaker may be defined as the ratio of direct, first path sound pressure level that is produced by an electrical signal applied to the driving point terminals of that loudspeaker. In most cases the electrical signal is characterized in terms of voltage applied to the terminals. One aspect of this transfer response is the so-called frequency response, in which sound pressure is measured as a function of frequency for an equivalent constant amplitude sine wave voltage.

With the advent of relatively inexpensive Fast Fourier Transform (FFT) instrumentation, it has become common practice to apply linear theory concepts to the measurement of the frequency response of a loudspeaker. In linear theory, frequency response may be obtained as the Fourier transform of impulse response. A short duration stimulus is used to excite a broad spectrum of frequencies. The Fourier transform of the time domain response of this stimulus is corrected for the spectrum of that applied stimulus and used to obtain the frequency response of the loudspeaker. Problems arise when the loudspeaker is nonlinear.

A type of nonlinearity which can lead to an incorrect frequency response measurement with this technique is illustrated in the measurement results of Figure 12. An increment of, say, one decibel in voltage should produce precisely one decibel increment in sound pressure level. In this example, a high quality three-way loudspeaker system has been tested at four different constant amplitude sweep voltage levels. These levels correspond to 10 milliwatts, 1 watt, and 10 watt average power into a constant resistance of 8 ohms. The free field frequency response at the lowest power level, 10 milliwatts, was digitally recorded and used as a reference for the free field response at the three higher levels.

The measurement of Figure 12 is not the frequency response of the loudspeaker, but the change in frequency response caused by an increase in drive level. Three measurements are shown, corresponding to the response,

$$Q_{P/10}(\omega) = \frac{10}{P}\,\frac{F_P(\omega)}{F_{10}(\omega)}$$

where $F_{10}(\omega)$ is the frequency at 10 milliwatts, and $F_P(\omega)$ is the frequency response at P-milliwatts.

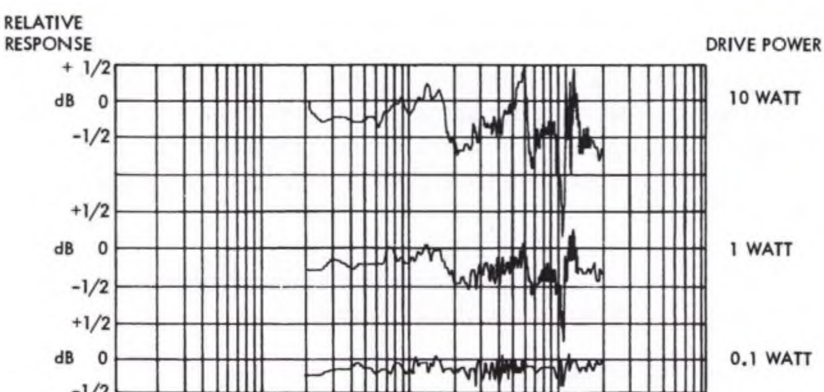


Figure 12 Loudspeaker test results at three different average power levels referenced to their response at the 10 milliwatt level.

If the loudspeaker were perfectly linear, then $Q_{P/10}(\omega)$ would be unity. Figure 12 is a plot of Q for 10 watts, 1 watt and 100 milliwatts drive; each presented relative to its own unity reference of 0 dB. The scale factor is 0.5 dB per division. Bit level quantization for these measurements corresponds to 0.09 dB per bit, and the lowest level fluctuations are caused by room background noise in the reference measurement which was made at 10 milliwatts drive level. The reference SPL is approximately 75 dB relative to 20 micropascals and the sweep rate is 19 kHz per second. The loudspeaker was allowed to come to near thermal equilibrium at the room temperature of 22 degrees Celsius and then a single sweep was made at the test level. Even though this particular loudspeaker is rated for safe continuous drive in excess of 10 watts, this method assures a minimum effect due to transducer temperature rise since none of the three drivers receives more than 5 watt-seconds at the highest test level. In other words, the effects which this test reveals are not thermal and are highly repeatable. Furthermore, this particular loudspeaker system was not selected for its nonlinear properties, but was chosen at random and represents a commonly encountered problem.

Curve A shows a response degradation even at 100 milliwatts. In this case the sound pressure level is approximately 0,15 dB below the level it should have.

Curve B shows a substantial change due to an increase to 1 watt drive level. Average sound pressure level is about 0.25 dB below its expected level and there is a peak-to-peak excursion of nearly 1.5 dB relative to perfection. Curve C, taken at 10 watts, continues this trend with a peak-to-peak excursion of 2.25 dB. The change in response due to the midrange driver can clearly be seen in the 2 kHz to 9 kHz range, while the tweeter characteristics show up in the 9 kHz to 20 kHz range.

The tracking filter of TDS allows us to measure the 10 watt response of this loudspeaker at each value of frequency while other parts of the spectrum are at far lower power levels. It is what one would get for each frequency if a 7 millisecond duration Gaussian shaped packet of single frequency sine wave was to be applied for a single measurement.

It should be clear that it would be most difficult to certify the frequency response of this particular loudspeaker system for any applied voltage whose peak to average ratio was greatly in excess of unity.

The foregoing TDS power-difference measurement technique can be quite valuable in situations when both amplitude and frequency dependent nonlinearities are in evidence, such as some cases of cone rub and certain hysteretic nonlinearities that give rise to an acoustic response which has been called "cone cry". When substantial transfer nonlinearities are known to be present in an acoustic process, such as those of relaxation and adiabatic nonlinearity in the medium, then it would also seem reasonable to exercise caution in the application of testing methodologies based upon linear theory.

Microphone Response

Acoustic measuring situations will often require the use of an intervening material, such as microphone windscreen, or involve adjacent physical objects which can interfere with an otherwise open field measurement. TDS can be used to measure the acoustic effect on a microphone measurement in such situations, even though the intent may be the study of naturally generated sounds. Furthermore, the very high processing gain (high time-bandwidth product) of TDS can allow such measurements to take place even in open air situations involving substantial ambient noise.

A loudspeaker can be used as a coherent sound source. This is placed a small distance from the microphone to be tested and in the direction for which the acoustic correction is to be determined. The loudspeaker does not have to be of high quality.

Two TDS measurements should then be performed; one with the object whose effect is to be determined, in place, and the second with the object either removed or displaced in an appropriate manner, depending upon the type of interference. The incremental change between these two measurements can then be used to determine the acoustic effect.

If the object is a windscreen through which the sound must pass in order to reach the microphone, then a simple difference measurement is all that is required. In effect, the acoustic transfer function of the windscreen is in cascade with the acoustic transfer function of the loudspeaker and microphone. If $L(\omega)$ is the frequency transfer function of the loudspeaker, $W(\omega)$ the windscreen transfer function, and $M(\omega)$ the microphone transfer function, then a measurement with the windscreen in place produces a response $O_1(\omega)$ of,

$$0_1(\omega) = L(\omega) \cdot W(\omega) \cdot M(\omega) \tag{74}$$

Removing the windscreen produces,

$$0_2(\omega) = L(\omega)\ M(\omega) \tag{75}$$

The desired information is obtained as the quotient

$$0_1(\omega)/_{0_2(\omega)} = L(\omega) \cdot W(\omega) \cdot M(\omega) / [L(\omega) \cdot M(\omega)] = W(\omega) \tag{76}$$

If the measurements are expressed in conventional logarithmic form, decibels (or nepers) for amplitude and degrees (or radians) for angle, then the attenuation in dB of the windscreen is the difference between the dB responses of the two measurements. A similar calculation will give the value for angle.

As an example of this procedure, Figures 13, 14, and 15 show the steps used to determine the acoustic effect of the protective grid on the frequency response of a laboratory microphone. The protective grid of a Bruel and Kjaer condenser microphone capsule, type 4163, was removed. A three-way consumer-produced "hi-fi" loudspeaker system was then placed at normal incidence angle with respect to the microphone diaphragm and 1.25 meters distant. Figure 13 is the TDS amplitude measurement of the direct sound from that loudspeaker with the microphone protective grid removed. It should be noted that the frequency response of the test loudspeaker is quite irregular; this is of no consequence, so long as the test loudspeaker can produce sound of sufficient intensity at all frequencies of interest.

Next, the protective grid was replaced on the condenser microphone and a second loudspeaker measurement was performed. The difference in the amplitude of these measurements is the amplitude response change caused by the grid. Figure 14 is this measured change caused by the protective grid. The data of Figure 14 is the actual unsmoothed difference measurement and is in good agreement with the curves published by Bruel and Kjaer for the correction to be applied due to the protective grid at this incident angle. Figure 15 is the change in phase response caused by the protective grid.

Another class of interference in an acoustic measurement situation occurs when a nearby object reflects sound back toward the measuring microphone. The simple post-demodulation subtraction process, the "difference of the dB's", of the previous example cannot be used. Instead, a vector difference procedure should be used which can be characterized as the "dB of the difference". In this case the response due to the loudspeaker, microphone and reflecting object,

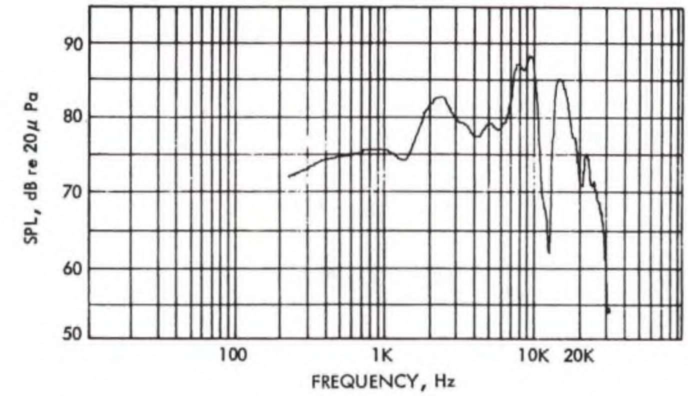


Figure 13 Reference response of loudspeaker - microphone combination with protective grid of microphone removed.

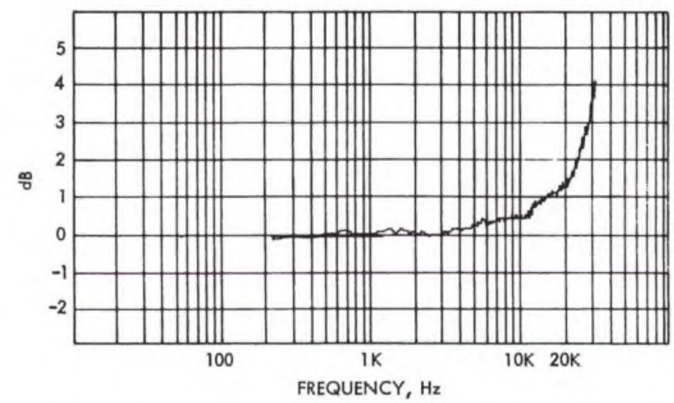


Figure 14 Measured change in amplitude response caused by microphone grid.

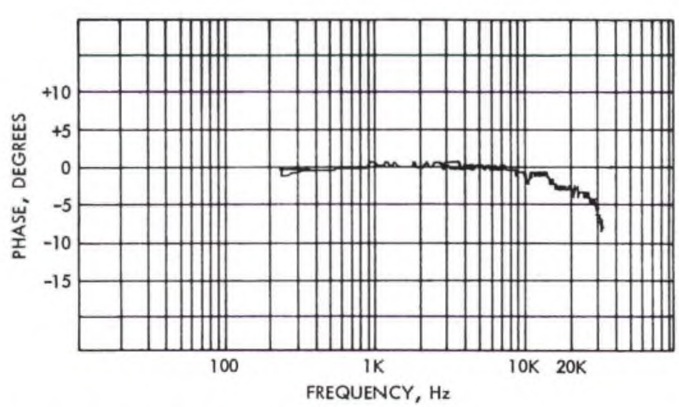


Figure 15 Measured change in phase response caused by microphone grid.

with frequency response $R(\omega)$ is the summation of direct and reflected sound which can be written as,

$$0_1(\omega) = L(\omega)[1 + R(\omega)] \cdot M(\omega) \tag{77}$$

A first measurement is made with the reflecting object in place. Then the object is removed and a new measurement made, producing,

$$0_2(\omega) = L(\omega) \cdot M(\omega) \tag{78}$$

A vector difference of these two measurements produces,

$$0_3(\omega) = 0_1(\omega) - 0_2(\omega) = L(\omega) \cdot R(\omega) \cdot M(\omega) \tag{79}$$

Figure 16 is a TDS measurement of such a situation. In this case a half-inch laboratory condenser microphone was being used in conjunction with a ribbon velocity microphone, type RCA 44BX, to measure both pressure and pressure gradient of a sound source. The ribbon microphone was placed adjacent to the condenser microphone such that the condenser diaphragm and ribbon lay in a plane normal to the direction of the sound source to be measured. The bulkier external windscreen of the ribbon microphone was spaced 5 cm from the frame of the condenser capsule. Due to the substantial size of this ribbon microphone and its close proximity to the condenser microphone, there was some concern about the validity of the pressure measurement. The question was: how much sound was being reflected off the windscreen of the ribbon microphone and picked up by the pressure microphone?

The upper curve of Figure 16 is the pressure response measurement of a test loudspeaker when the ribbon microphone was in place. The vector TDS signal, before being demodulated for amplitude or phase, was digitized and stored in memory. Then a second measurement was made with the same setup and the two vector signals were subtracted then demodulated. As expected, the result was null; the two signals precisely cancelled in both amplitude and phase. This second measurement was a verification test, to assure the stability of the test situation. Then the ribbon microphone was removed and a third measurement taken. The first and third measurements now

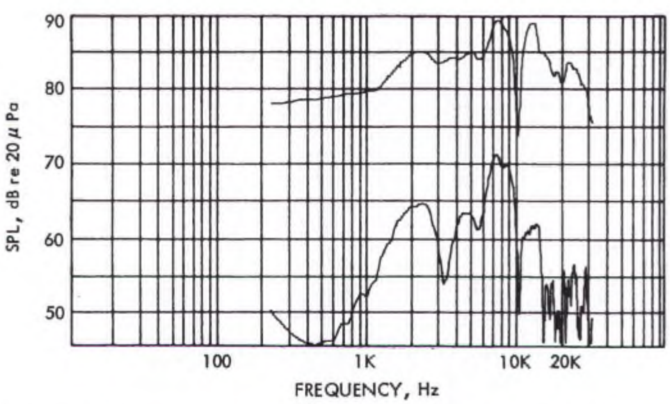


Figure 16 Vector difference measurements showing effects of reflections from a ribbon microphone.

did not cancel precisely since the vector signal components, due to the reflections off the ribbon microphone, were absent in the third measurement. The lower curve in Figure 16 is the demodulated amplitude response of the vector difference between the first and third measurement. The upper curve corresponds to the measurement $0_1(\omega)$, while the lower curve corresponds to $0_3(\omega)$. The difference between these two curves, $R(\omega)$ represents the amount of sound being scattered from the ribbon microphone and picked up by the pressure microphone.

As a practical matter, it must be pointed out that this vector difference technique is exceedingly sensitive to small changes between two successive measurements. For example, if two measurement situations are geometrically identical, but the second measurement differs from the first by 0.1 dB, then the vector difference will not be null, but will lie 39 dB below the first measurement. If, on the other hand, the two measurements are identical in amplitude, but differ in arrival time, then the vector difference will assume a frequency dependence of the form,

$$|R(\omega)| = 2\,|\sin(\pi\, T\, f)|, \tag{80}$$

where T is the difference in arrival time and f is frequency in Hertz. For example, an acoustic measurement with a 1 microsecond difference in arrival time produces a signal that lies only 18 dB below the undifferenced signals at 20 kHz, and 44 dB below at 1 kHz.

The vector difference technique is essentially an interferometric measurement that can process selected signals and exclude earlier or later arrivals. With a processed signal-to-noise ratio of 60 dB, which is commonly achieved in normal measurement situations, the vector difference of two signals at 10 kHz will be 7 dB above background level for a time difference of 32 nanoseconds. This corresponds to an equivalent spatial offset of 11 micrometers at the speed of sound in air. This technique, with micrometer drive on a microphone and fixed sound source, can be used to measure the speed of sound in a given measurement situation if desired.

Underwater Sound Measurements

An experimental towed acoustic sounder, capable of deep submergence, was assembled by the Jet Propulsion Laboratory of the California Institute of Technology (JPL) and used in a series of successful deep water cruises in the years 1977-1979. Figure 17 is a photograph of this "fish".

The assembled "fish" is approximately 2.5 meters in length. Low voltage sweep signals are passed from the TDS system in the surface tow ship through a multiconductor tow cable to a linear power amplifier on the "fish". The amplifier signals are acoustically projected from a flooded ring crystal transducer mounted forward of the circular tail fin. The useful frequency range of this transmitted signal, over which sound pressure level is essentially uniform, extends from 1.5 to 4.5 kHz. The polar pattern of projected sound is toroidal, with the axis of minimum response aligned along the axis of the "fish".

A six-element hydrophone assembly, having nearly the same polar acceptance response as that of the projector, is housed in the section immediately aft of the nose. Signals intercepted by the hydrophone are amplified within the "fish" and passed back up another pair of wires contained in the tow cable.

There are several features which set this JPL "fish" apart from more conventional towed sounders and which are a direct consequence of its use of TDS. First, since TDS is essentially a spread spectrum methodology, the projector and hydrophone have as wide a frequency bandwidth as possible. Second, the directivity patterns of both projector and hydrophone are deliberately made as broad as possible in the direction of the objects to be imaged and measured. Third, since a very high time-bandwidth product signal is utilized, the vehicle is designed to receive low level acoustic reflection signals during the same time that the transmitter is operating, Fourth, all aspects of signal processing are maintained in a linear mode with no attempt made for time-variable gain, limiting, or other aspects of more conventional pulsed sonar systems.

This "fish" is specifically designed to use TDS and was intended to obtain backscatter frequency spectrum properties of the bottom and subbottom. In order to obtain stable acoustic directional properties, independent of pitch angle or yaw of the vehicle, the received signal was mapped within the TDS processing so as to utilize vehicle forward motion, in conjunction with the associated Doppler frequency shifts of received signals, to produce electronic beam forming. That is, the phase coherent transmitted signal, which is used as a reference for mapping the received signal, can be time delayed and corrected for the spectrum shift corresponding to a particular Doppler offset. For the material reported in this paper, a zero Doppler component was utilized, corresponding to a maximum response abeam of vehicle translation. For simple scatterers, the beam angle, formed by this process, is inversely proportional to the product of the speed of translation and the time-bandwidth product of the signal that is received. At a tow speed of 7 knots (3.5 m/s) and the JPL time-bandwidth product of 1000, the theoretical 3 dB half angle produced by this process is 11 degrees. The processed data from these cruises is in agreement with this theory.

Attenuation of Compression Waves in Sediment

The case where a uniform sediment layer lies over an acoustically stiffer medium can be modelled, as shown in Figure 18. It is assumed that ocean bottom sediment lies at a thickness of d_3 above a much stiffer acoustic reflecting boundary. A projector and hydrophone assembly, shown as T, is towed a distance d_1 above the sediment and d_2 beneath the air-water interface. If T insonates the water, then the first order reflections are as shown in Figure 18a. The anticipated magnitude EFC for the reflected signals is shown in Figure 18b for the case where spherical spreading loss and frequency dependent seawater attenuation has been removed. The air-water interface is to be used

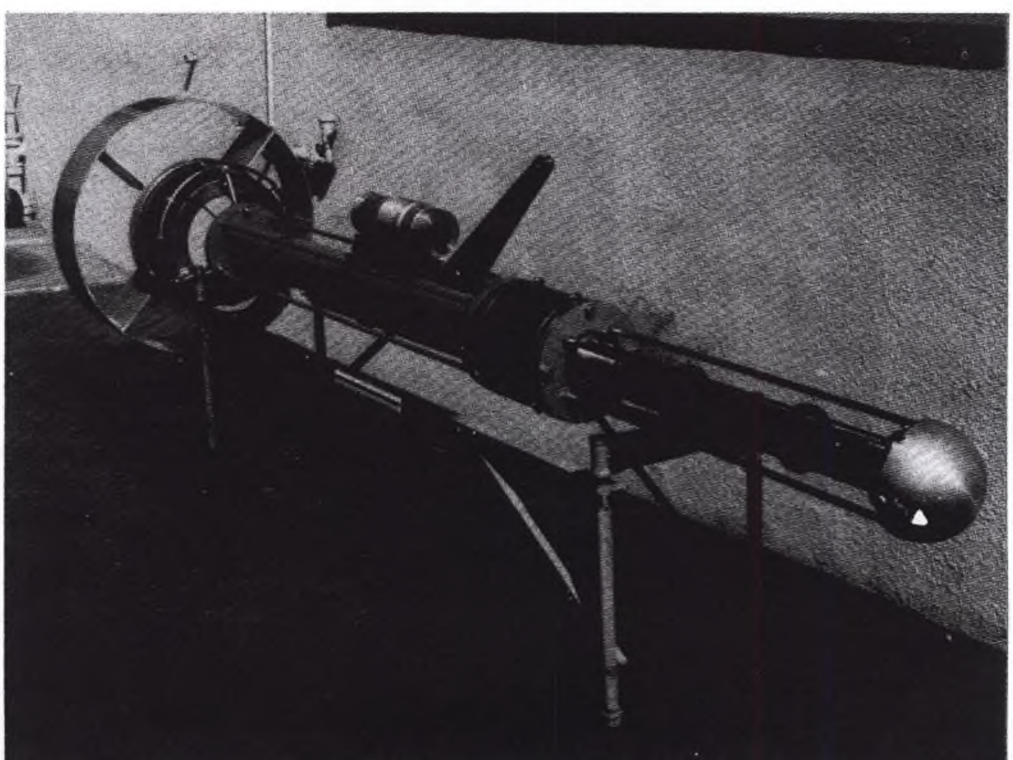
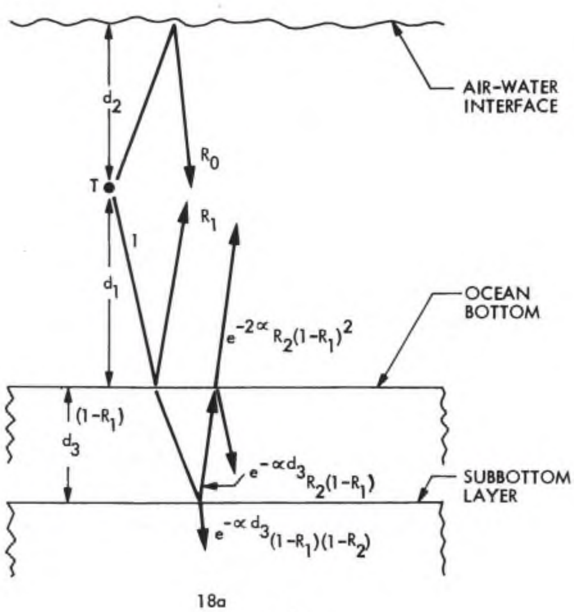


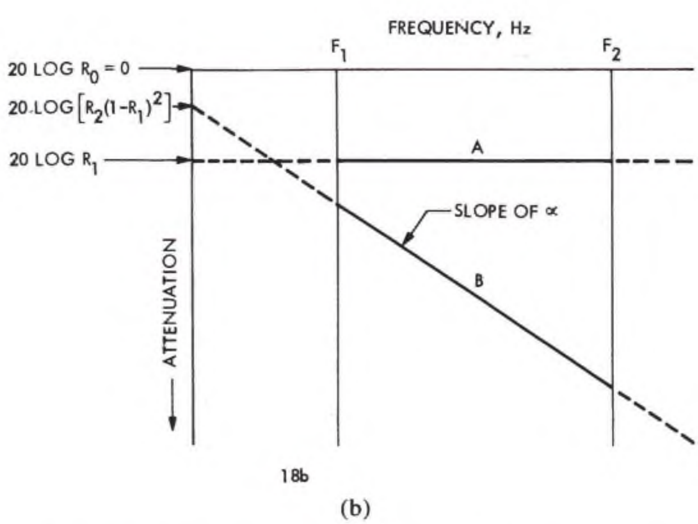
Figure 17 Photograph of an experimental towed acoustic sounder.

as the reference for reflected intensity measurements under the condition where T radiates as much sound upward as downward toward the ocean floor.

The air-water reflection coefficient is essentially constant with frequency in the range of frequencies used for these experiments and is considered to be the reference of 0 dB on the vertical logarithmic scale of Figure 18b. Similarly, the reflection coefficient R_1 of a smooth ocean sediment interface, shown as curve A, is independent of frequency and lies below that of the air-water interface. The value of R_1 can be directly determined at location T by comparing the ratio of reflected sound received from the first echo off the sediment to that of the first echo received off



(a)



(b)

Figure 18 (a) First order underwater reflections, (b) Schematic attenuation versus frequency curve for reflected signals.

the air-water interface, after removing the known attenuation properties due to spherical spreading of the sound wave and attenuation rate due to absorption in seawater. In the case where the projector and hydrophone operate only in frequency range from F_1 to F_2, the reflection at other frequencies can be extrapolated as projections of this linear dependence beyond F_1 and F_2 This is shown by the dashed lines.

If the sediment is essentially homogeneous and overlies a much harder subbottom, then the anticipated frequency dependence of reflected intensity in the homogeneous sediment, due to absorption, will follow a logarithmic decrement that is of constant decibels per kilohertz per meter of sediment path length. The value of the reflection coefficient R_2 can be determined by the zero frequency intercept of B, whereas the attenuation ratio (alpha) is determined from the slope of B.

Figure 19 is a TDS measurement of this type of subbottom measurement made in a region known as KAYAK TROUGH in the Gulf of Alaska. The data of the Figure 19 are the TDS filtered responses corresponding to curve B of Figure 18b. Five sequential measurements, each of one-third-second duration, and repeated on four-second intervals, is shown. Spherical spreading loss has been removed, but no other correction or smoothing has been applied in this display. The fall-off of response below 1.8 kHz and above 4.3 kHz is due to the deliberately restricted frequency response of the JPL "fish". Of engineering significance is the fact that these measurements were made "on the fly". Due to approaching storms, a plan for acquiring detailed acoustic soundings at each of several ocean bottom coring stations had to be abandoned. Instead, the ship could only remain on station for a time long enough to acquire the physical cores and then move to the next station some distance away. The "fish" was left in the water but could only acquire data while the ship was in motion, due to its designed towing ballistics. The measurements of Figure 19 were taken in relatively rough seas of seastate 5 at a speed of approximately 7 knots.

Figure 20 shows the computed attenuation versus frequency obtained from those measurements. Curves labelled 1, 2 and 3 were obtained for KAYAK TROUGH, with sediment depths ranging from 30 to 40 meters, while curve 4 was obtained from a region known as the COPPER RIVER DELTA in the Gulf of Alaska. The computed reflection coefficient for KAYAK TROUGH obtained from this TDS data is 0.138, which, according to Hamilton,[32,33] places it in a category of clayey silt, silty clay. The data of Figure 20 show good agreement with published results for this class of sediment but, interestingly, fill in a gap in the 2 to 4 kHz of published data.

<u>Bottom and Subbottom Imaging in the Gulf of Alaska</u>

TDS ETC measurements of the ocean bottom were simultaneously obtained with the EFC measurements. The magnitude ETC was corrected for spherical spreading loss and plotted as a function of reflection time (for range measurements). Figure 21 is a sample taken of such a bottom and subbottom plot in the Gulf of Alaska. Again, the region was KAYAK TROUGH, but this time the bottom is severely jumbled due to slumping from prior seismic activity. Figure 21 shows this effect, with tilted subsurface layers in evidence. Again, this data was acquired while the ship was in transit under moderately rough seas.

Figure 21 is a traditional bottom image of depth versus horizontal position. All data were recorded on magnetic tape and could be later reduced in a laboratory environment. Figures 22 and 23 show such a data reduction. The single line in Figure 22 represents a TDS sweep at one ship position. This particular sweep occurred at 21:17:30 Z. Figure 22 shows an expanded 500 point image of the ETC of 21:17:30 Z. For data reduction purposes the TDS data were sampled with a window corresponding to 37.5 meters (at 1500 m/s sound speed) and commencing approximately 25 meters above the bottom. The first few meters show a distinct double response, as though there were two layers.

Figure 23 is an expanded plot of the first meter of sound penetration shown in Figure 22.

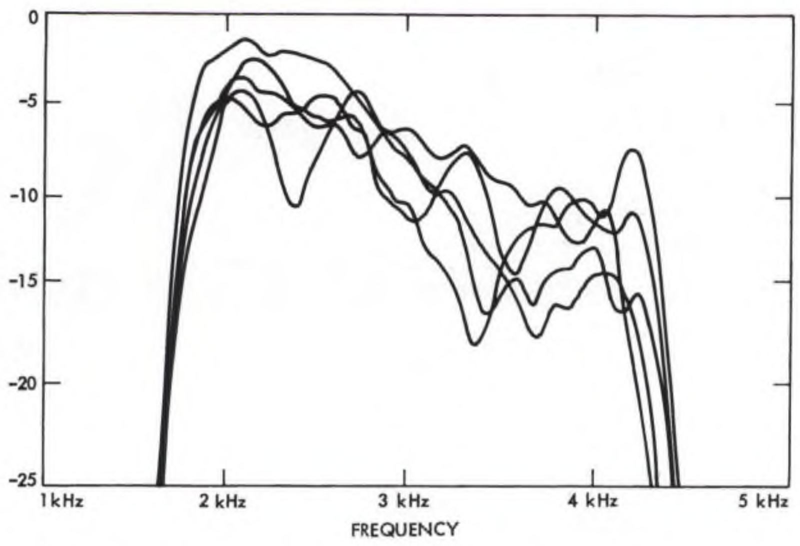


Figure 19 TDS measurements of the subbottom in the Kayak Trough in the Gulf of Alaska.

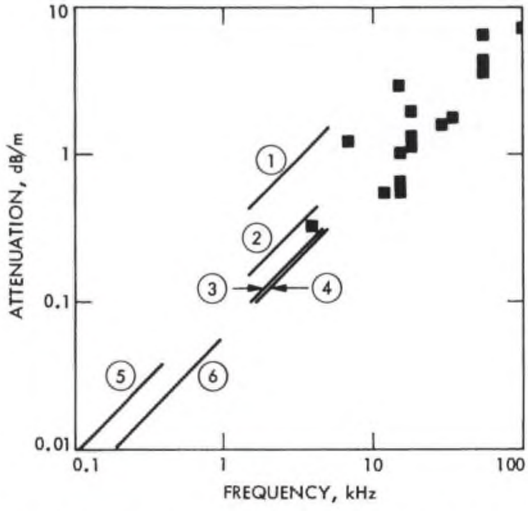


Figure 20 Computed attenuation versus frequency data for the Kayak Trough and for the Copper River Delta.

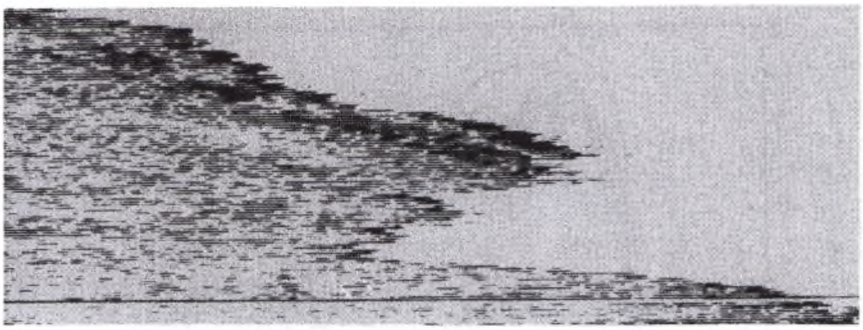


Figure 21 ETC of bottom and subbottom measurements of the Kayak Trough.

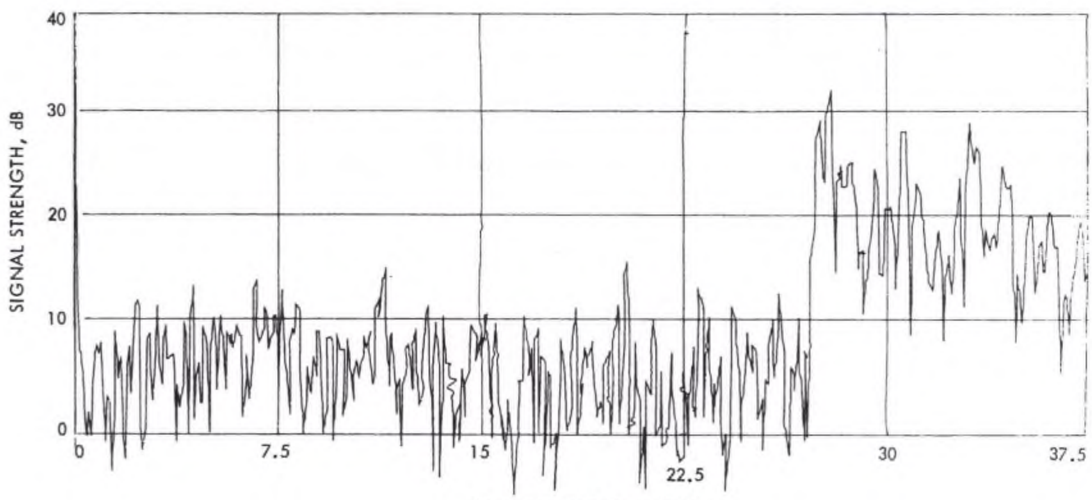


Figure 22 Typical TDS sweep measurements at one ship position.

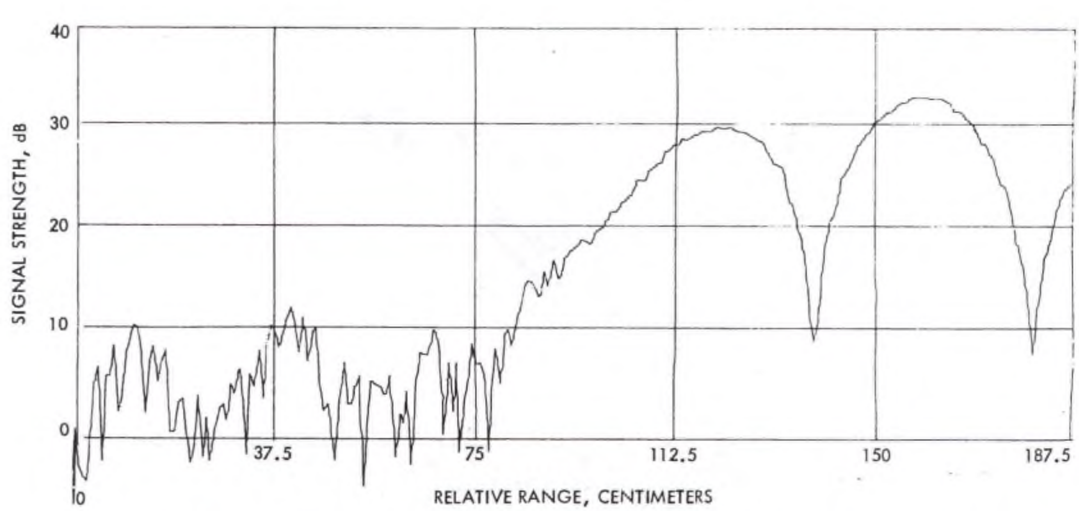


Figure 23 Expanded plot of the first 1.875 - meters of sound penetration.

The double layer response is now clearly in evidence. The limiting resolution of this particular TDS system for soft sediment sound speeds of 1500 meters/sec is approximately 25 cm, due to the 3 kHz bandwidth. This resolution is closely approached in this data which was taken from a towed "fish" at a translational speed of 3.5 meters/sec and flying 274 meters above the ocean bottom. The advantage of synthetic beam forming is clearly in evidence.

Imaging the San Andreas Fault

It has been pointed out in Section I that the complex h(s), relation (7), is an energy functional. This energy functional expresses energy density and its partitioning in terms of frame of reference s. The plot of the logarithm of this energy functional is what I have referred to as the energy curve. When acoustic soundings are taken of the ocean bottom and subbottom, the magnitude of this energy curve can be used to image submerged structures in terms of the amount of acoustic energy which they reflect back toward the receiver.

Traditional acoustic imaging depends upon the backscatter of energy that occurs when the elastic wave encounters a change in acoustic impedance, usually at the boundary of two different materials. Edges are thus clearly outlined, while the bulk scatter from regions on either side of the boundary are seldom imaged. There are situations, however, where interest lies in the nature of volumetric or bulk backscatter of energy as the elastic wave passes through a nearly homogeneous structure. In that case, the full energy functional can be of value.

In March 1979, a set of TDS measurements were taken of the bottom and subbottom trace of the San Andreas Fault in the waters off Cape Mendocino, California. One of the most famous geologic fault structures in the world, the San Andreas Fault cuts northward through southern and central California and passes into the Pacific Ocean near San Francisco. There it continues its northward course along the ocean bottom to a place west of Cape Mendocino, then arcs toward the west into much deeper ocean waters.

The soft sedimentary nature of the ocean bottom immediately north of San Francisco, together with a shallow depth that is of the order of a wavelength of long rolling surface waves, creates a situation in which the ocean bottom trace of this fault is smoothed and in some places obliterated by scrubbing action of bottom currents. The intent of the JPL's TDS measurement, taken under the sponsorship of the National Oceanic and Atmospheric Administration, was to determine whether the fault could be discerned by means of subbottom sediment disturbances.

Figure 24 is one example of San Andreas Fault data which was acquired on this cruise. The fault line is clearly evident as the transition in acoustic reflection for sediment lying immediately beneath the surface, as well as for a sediment layer lying 30 meters beneath the surface. The fault trace is indicated by the arrows in this figure.

The image of Figure 24 is a vertical slice taken for a path which crossed normal to the San Andreas Fault. The ocean bottom is the dark trace at the top of the image, with the light portion above this being water. The vertical scale has been expanded to a factor of nearly 100 times the horizontal scale. The coast of California lies some 30 km to the right of this image.

In this particular location the fault has split into two nearly parallel structures. Arrow A shows a trace which is distinguished by severe disturbance of the entire sediment layer, while Arrow B shows a more dramatic trace which involves a complete discontinuity of subbottom sediment, indicating perhaps a substantial lateral displacement of the material on either side of this particular trace.

There were a number of imaging pulse sonar transducers mounted in the hull of the oceanographic tow ship. These did not see this subbottom disturbance in sediment which was caused by the San Andreas Fault. The energy-time curve, however, could distinguish the bulk properties of backscattered sound since it measured the average energy of reflection. The data of

Figure 24 are corrected for spherical spreading loss and the image has been contrast enhanced to show the small change in subbottom acoustic backscatter. The slight line at 37.5 meter sediment depth is an artifact caused by the FFT from which this image was processed.

Synthetic Aperture Imaging

The TDS system used for ocean bottom and subbottom measurements utilizes a sweeper whose phase, as well as phase rate, is under digital control. A high stability clock is counted down and used as control for all sweep functions, including start time and programmable time delay. Since the clock data was recorded along with the projector sweep and hydrophone signal returns, it should be possible to utilize this fully phase-coherent set of spatial returns to form a wide-base synthetic aperture image.

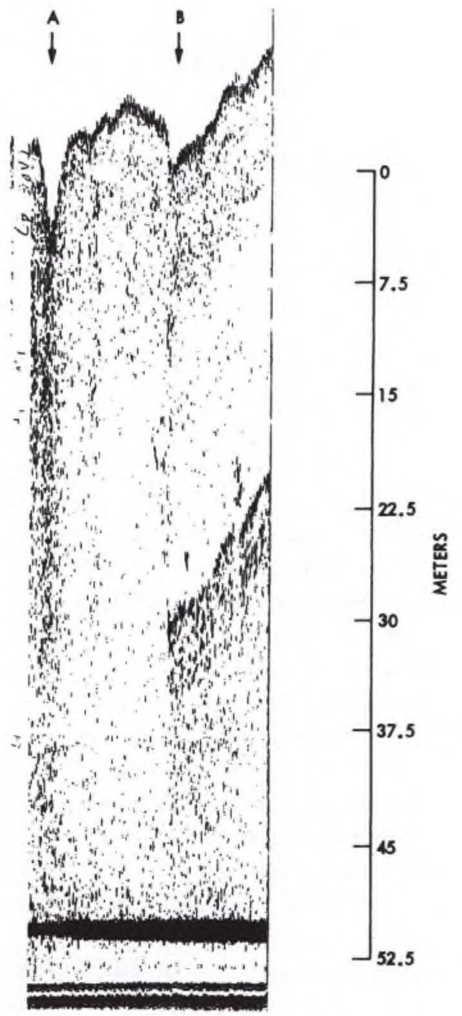


Figure 24 Imaging of a portion of the San Andreas fault.

The Doppler beam forming technique, which produced the results of Figure 21, is essentially a narrow-base synthetic aperture methodology whose base width is equal to the distance travelled by the "fish" during the duration of sweep. This base width is of the order of one meter for speeds of 3.5 meters per second and a sweep duration of one-third second. The corresponding half-angle resolution is slightly less than 0.25 radians for the 3 kHz bandwidth.

It should be possible to increase this base width through the coherent processing of a large number of successive TDS sweeps. In order to check this hypothesis, analysis was performed on a deep ocean data set which was taken off Monterey, California, in 1977. The Monterey test was performed in waters whose depth is in excess of 1000 meters, and the JPL "fish" was towed at a distance of 200 meters beneath the air-water interface. This means that the underside of the sea surface, as "seen" by the "fish", is at a sufficient distance to present a good test object for synthetic aperture analysis. The ocean bottom is also far enough away to prevent interference in such an experiment.

The underside of the sea surface provides a good test object to demonstrate synthetic aperture for two reasons. First, there is no possibility of imaging anything but the surface; "subsurface" penetration artifacts will not exist in the data for the air-water interface. Second, at the frequencies used, sound reflections are essentially specular in nature. The seas were moderate, with large rolling waves and few breaking crests.

Figure 25 is a digital diffraction pattern (hologram) made by mixing the TDS hydrophone received signal with a steady tone coherently derived from the clock which generated the sweep. The sweep extended from 1.5 to 4.5 kHz, and by multiplying the received data against 3.0 kHz, a diffraction pattern was generated whose fringes extended 1.5 kHz on either side of "zero beat". This pattern is compatible with the optical processor used by JPL to demodulate the Seasat Synthetic Aperture Radar image data. Figure 25 is a photographic record made from the digital processing of successive returns for approximately 8 minutes of translation. The horizontal displacement of Figure 25 corresponds to about 1.7 km of ship travel.

Figure 26 is the demodulated image produced by the JPL optical correlator from the data

Figure 25 Digital diffraction pattern produced by mixing hydrophone-received signal with a steady coherent tone.

Figure 26 Demodulated image produced by an optical correlator from the data of Figure 25.

of Figure 25. The nature of this reconstructed image is in good agreement with the anticipated synthetic aperture that could be expected from such rolling seas. Since no left-right distinction was made in the recorded data, both sides of the sea are folded together to produce imaging on either side of nadir. The aspect ratio of the synthetic aperture image is such that the vertical distance (azimuth) is expanded 3.5 times relative to the along-track distance. Slow variations in azimuth are due to changes in the depth of the "fish" caused by variations in the speed of the surface tow ship.

Second Harmonic Emission from Insonified Bubbles

The integral transform of relation (37) is capable of handling certain nonlinear situations. In particular, if a nonlinear system introduces a one-to-many mapping of an excitation into a response, it is often possible to isolate individual components of that multiple response through appropriate filter procedures and then map the resultant back to the excitation frame of reference. One such nonlinearity occurs when the system introduces responses which can be interpreted as harmonic multiples of the applied frequency spectrum.

This is the case when fluid immersed gas bubbles are insonified at their frequency of elastic resonance within that fluid. Both the absorption at fundamental frequencies at resonance and the second harmonic spectrum of backscatter sound from air bubbles in fluid have been measured by TDS methods. Using the technique of Miller and Nyborg,[34] air was trapped in the pores of immersed hydrophobic filter membranes to produce stable bubbles whose diameters lay in the 3 to 4 micrometer range. The membrane was insonified with a wideband transmitting crystal driven from a linear amplifier, and the backscattered sound was received by a second wideband ultrasound crystal. Transmission in the 0.5 to 10 MHz range was accomplished by a digitally controlled sweeping generator operating at a rate of 500 MHz per second.

Figure 27 is a measurement of bubble resonance in whole blood using this technique. Curve A is the measured frequency spectrum of the fundamental component of an ultrasound signal reflected from a uniform plane surface. Curve B is the resultant frequency component of reflection which was obtained when the membrane was placed in front of the reflecting surface. According to theory, resonating bubbles should cause substantial absorption of sound at their frequency of resonance. This effect is clearly in evidence in the fact that frequencies near the computed resonance frequency range of 1.2 to 1.6 MHz are substantially removed from the sound which passes through the membrane. This absorption phenomenon was experimentally determined to be independent of the angle of incidence between the ultrasound beam and the plane of the membrane, in accordance with theory. A fact which makes this measurement of significance is that it clearly shows that bubble acoustic resonance absorption exists in whole blood. These data are due to Rooney and Heyser.[35]

Similar experiments, in which the reflection of sound from the membrane was measured, verified that the resonating bubbles have a very high acoustic scatter cross section at their resonance frequency.[36] However, this fact cannot be used as a unique basis for detecting bubbles in fluids which contain other particulate material, such as whole blood, since this particulate material will also reflect energy at the fundamental frequency of an insonifying ultrasound beam.

Because of nonlinear resonance properties, bubbles have a unique backscatter property that might be used to distinguish them from competing ultrasound scatterers. Bubbles produce a second harmonic emission which is dependent in a specific way upon the energy of insonification.[37] In order to detect this emission, the TDS equipment was modified in its receive demodulation circuitry. Second harmonic reflection of energy was detected by a tracking receiver whose sweep program was driven synchronously with the transmitting program, but with a tracking rate of 1 GHz per second and a delay offset corresponding to the arrival time of backscattered sound from the membrane.

Figures 28 and 29 are frequency spectra obtained using this modified apparatus. Figure 28 is the spectrum of energy reflected as a fundamental component from a membrane containing bubbles which should resonate around 1.25 Hz. A narrower bandwidth crystal was used in this experiment in order to concentrate the energy around the frequency of bubble resonance.

Figure 29 is the measured second harmonic reflection from the bubbles within the membrane; the membrane itself having been verified to be an object which does not create measurable second harmonic distortion and backscatter. Both Figures 28 and 29 are oscilloscope images obtained from a time delay spectrometer and the horizontal axis of Figure 29 has been corrected such that the second harmonic component will appear at the same horizontal position on the screen as that of the corresponding fundamental measurement. The vertical scale in Figures 28 and 29 is relative amplitude in decibels, with a full scale deflection of 40 decibels. The display gain of the second harmonic component is increased 20 dB relative to that of the fundamental component in order to display both on the same screen display. The second harmonic from the bubbles lies approximately 30 dB below the fundamental component, which is principally due to the membrane in which the bubbles are immersed.

Selected combinations of transmission and reception sweep rate established that the observed second harmonic signal was due to resonance of bubbles and not due to incidental transmitter or receiver distortion components.

Figure 30 is an ETC of reflected fundamental energy. The membrane was physically placed 22 cm away from the transmitting and receiving transducers. Vertical scale factor is 10 dB per division, corresponding to 60 dB full scale. The peak at 22 cm corresponds to the first reflection from the membrane, with subsequent sound scattering from the structure supporting the membrane for these measurements. The signals centered around 18 cm are scattered sound from supporting structures. The small signal at 23.5 cm is the reflection from the rear wall of the tank.

Figure 31 is an ETC of the second harmonic reflection at precisely the same scale as the fundamental of Figure 30. All reflecting objects have disappeared except a cluster of energy emanating from a location known to be occupied by the bubbles. The membrane, its supporting structure, and the rear wall of the tank are all gone from the second harmonic measurement, as anticipated. The second harmonic emission from the bubbles is 45 dB above the baseline noise in this measurement.

Second harmonic emission from bubbles is a quadratic function of the insonifying level.[37] Figure 32 shows the relative amplitude of detected second harmonic ETC energy as a function of voltage applied to the transmitting crystal. The measured data are in excellent agreement with theory and each 5 decibel increase of fundamental excitation closely produces a 10 decibel increase in detected second harmonic emission up to the maximum test level which could be obtained from the linear power amplifier used for this test.

Ultrasound Imaging

TDS has been utilized in a successful and ongoing program of medical ultrasound measurements at JPL since 1971. One aspect of this work which has not been published in the literature relates to the use of the full ETC for the purpose of transmission ultrasound imaging of tissue. One example of this early imaging work will now be presented.

Figures 33 and 34 are transmission ultrasound images of a one centimeter thick section of

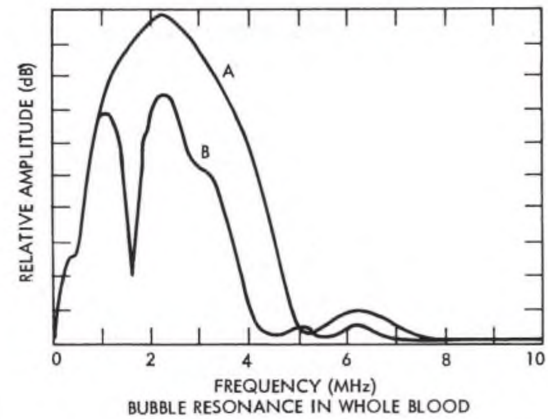


Figure 27 Measurement of bubble resonance in whole blood.

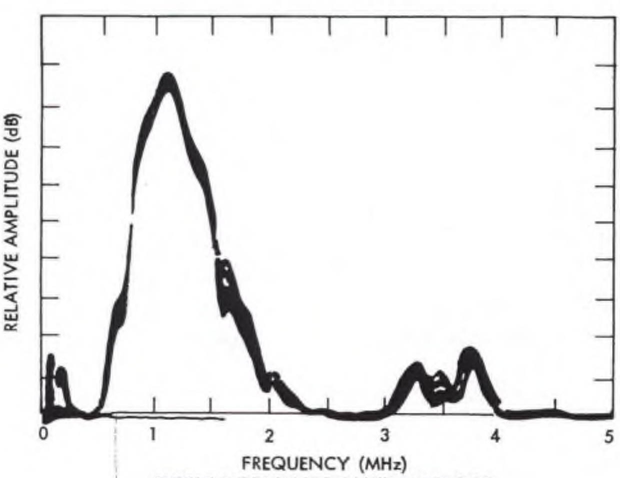


Figure 28 Frequency spectrum of energy reflected from bubbles in whole blood (Fundamental mode resonance).

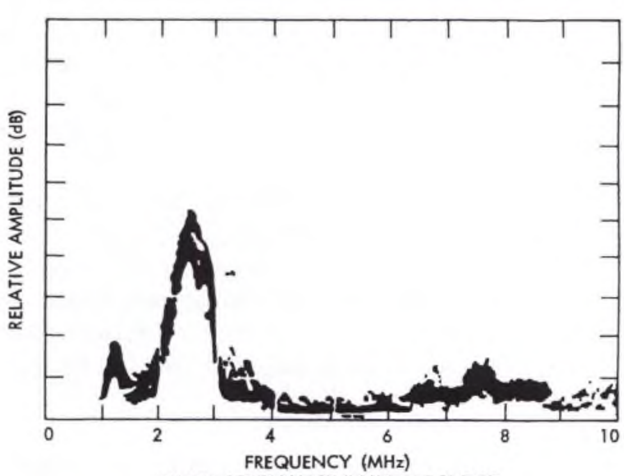


Figure 29 Frequency spectrum of energy reflected from bubbles in whole blood (Second harmonic mode excitation).

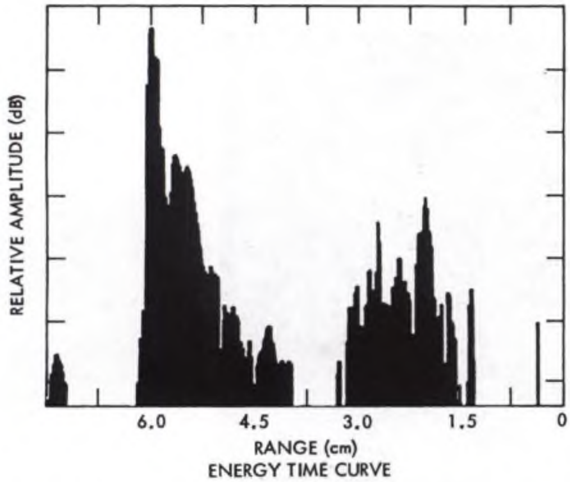


Figure 30 ETC of reflected energy from bubbles in whole blood (Fundamental mode resonance).

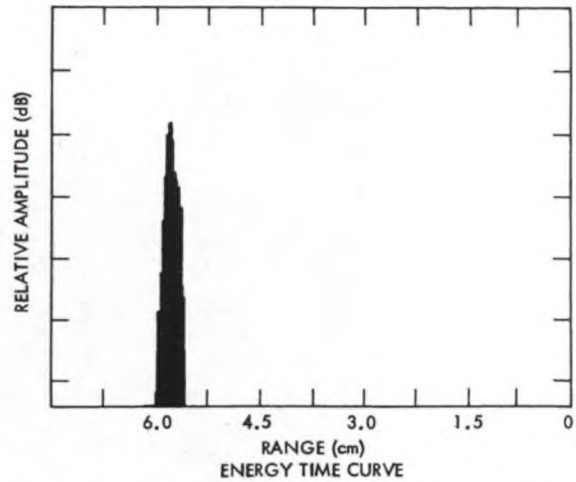


Figure 31 ETC of reflected energy from bubbles in whole blood (Second harmonic mode excitation).

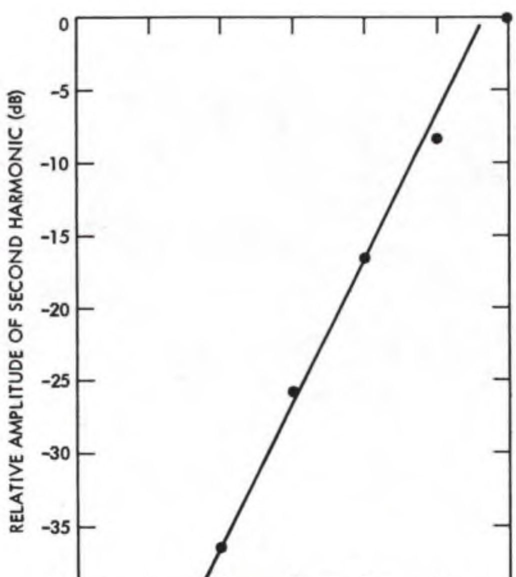


Figure 32 Relative amplitude of detected second harmonic ETC energy as a function of voltage applied to transmitting crystal.

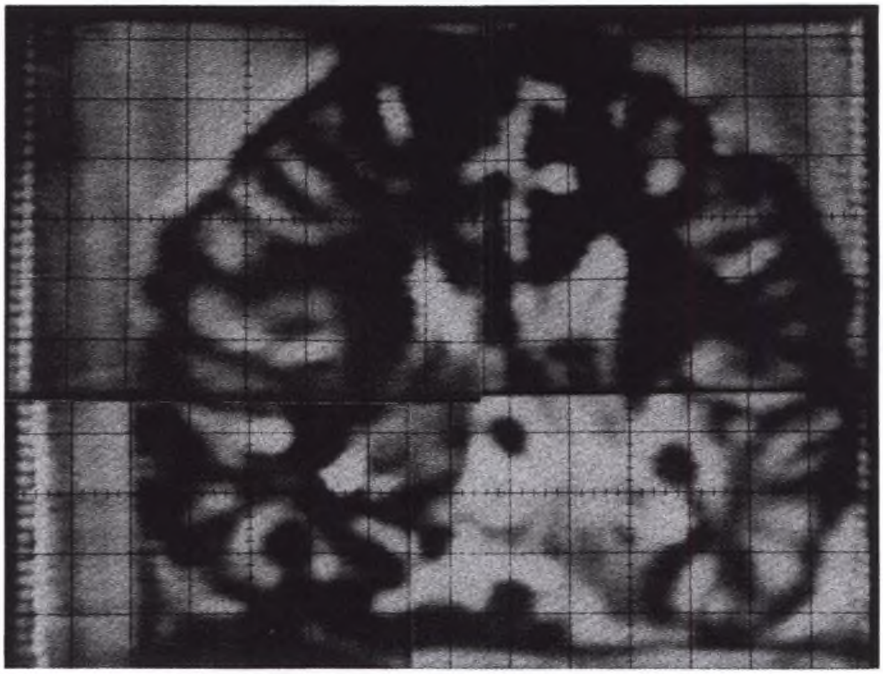


Figure 33 Transmission image of the amplitude of the first-sound ETC for a one-centimeter thick, formalin-fixed brain section.

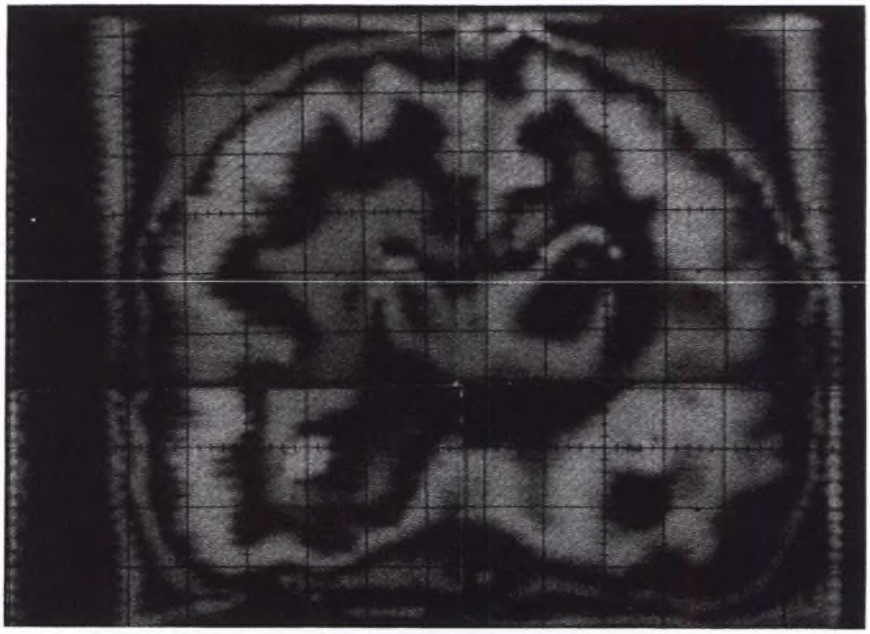


Figure 34 Transmission image of the amplitude of the first-sound ETC for a one-centimeter thick, formalin-fixed brain section.

formalin fixed brain. These images were obtained with the use of two opposing ultrasound crystals. The ultrasound crystals were simple unbacked barium titanate discs, 5 mm in diameter and with natural resonance frequencies of 2.5 MHz. One crystal was used for transmission of ultrasound energy and the other used for reception These crystals were heavily loaded with a coat of epoxy in order to broaden their resonance properties as required for the spread spectrum signals of TDS.

The coated crystals were rigidly mounted on an inverted U- shaped yoke assembly which placed them 20 cm apart and aligned their axes of maximum ultrasound sensitivity. This yoke was itself rigidly fixed to the pen carriage of an X-Y chart recorder drive-oriented for vertical/lateral motion above a water tank which contained the specimen to be scanned.

The images were taken with a sweep frequency bandwidth of 1 MHz and with the system operating between 2 and 3 MHz. That component of the ETC corresponding to the first microsecond of sound which passed through the specimen was demodulated for amplitude and phase. The demodulated signal was then used to modulate the intensity of a Tektronix Model 603 monitor, whose spot position was synchronously driven by the same scanning signals which were used to drive the mechanical yoke assembly.

Figure 33 is the transmission image of the amplitude of the first-sound ETC, while Figure 34 is the corresponding phase image. Since the screen size of the Model 603 monitor could not accommodate the full 125 mm width of the specimen, and since it was desired to display the image at full scale, it was necessary to reposition the scanner for separate measurements for each complete image. These four measurements were overlaid to present the full images shown in Figures 33 and 34. Screen reticle marks accurately correspond to 1 cm lateral and vertical displacement. Apparent lack of registry in the composite reticles is due to imprecise repositioning of the specimen for separate measurements.

These images were taken as part of an analysis relating to the ultrasound detection of a class of brain tumor known as glioblastoma. Two such tumors reside in the lower right quadrant of these images and the result of this analysis was reported in the literature,[16] although these particular images, and many more such images, have not heretofore been published.

The features which are of interest in this present discussion of TDS relate to the nature of image information which can be obtained from the full ETC, amplitude and phase. It must be emphasized that Figure 34 is not a plot of the phase angle of an ultrasound carrier which passes through the object being imaged; it is a plot of the phase of the energy-time curve corresponding to a particular moment of arrival of ultrasound energy at the receiving crystal. The ultrasound TDS signal was coherently spread over a 1 MHz spectrum centered at 2.5 MHz, in this particular case. The received signal was despread and computed as a complex time-dependent quantity (Equation 7). The magnitude of the ETC (Equation 73) was analyzed and an epoch of time was chosen which corresponded to the moment when the maximum amount of first-sound energy passed through the specimen. This was the direct sound; subsequent sounds were due to later signal arrivals, such as those caused by reflections and reverberation of sound within the object being scanned. The magnitude ETC for this first sound is plotted in Figure 33 and the phase ETC is plotted in Figure 34.

It should be recognized that the lateral resolution of these images is of the order of 2 mm, even though the ultrasound crystals from which the images were obtained were unfocussed, The use of TDS effectively narrowed the beamwidth of the crystals by coherently rejecting non-direct-path signals which took a longer transit time than the direct-path signals. This is a general property, and has been extensively utilized in ultrasound imaging at JPL. And, while the pattern of Figure 33 relates to the logarithmic attenuation of energy at each pixel location, the pattern of Figure 34 relates to the exchange relationship between the two quadrature components of energy density making up the total energy density. For a simple structure, such as this brain section, which does appreciably alter the shape of the separate components of the energy density packet, the pattern of Figure 34 relates to the change in arrival time relative to the epoch chosen for the moment of energy arrival. Each dark-to-white-to-dark transition in Figure 34 corresponds to an arrival time range of 1 microsecond, which is approximately 1.5 mm of path length in water. By comparing the amplitude and phase ETC images, it is possible to gain more information about the specimen than would be possible from observing only one image. For example, the contours of the formalin filled plastic bag, in which the specimen was placed, can be clearly discerned in the fringes at the bottom of the image, Also, the gray and white matter of the brain show a distinctly different attenuation, whereas there is little distinction in the relative speeds of sound, The glioblastoma, which show a significant attenuation, do not cause a change in speed of sound. It is apparent that there are features which reveal themselves in phase but do not show in amplitude, and conversely.

Although these represent ETC transmission images, those who are familiar with acoustic and electromagnetic reflection images should recognize that a significant benefit can also result in such cases when phase images are combined with amplitude images.

SUMMARY

Einstein said, "It is the theory which determines what we can observe". That is true, even in acoustics, for it is our theory that establishes how a particular process may be characterized in terms of some frame of reference. Having established a frame of reference, and, from our theory, a basis of relationships that ought to exist on such a frame, we can postulate the manner in which a stimulus might elicit a response. And from this we can set up apparatus which interacts with the process in some small but finite way so as to provide a measurement of the energetic exchanges involved in this frame of reference.

But self-consistency of the theory and of measurements performed on its frame of reference should not be taken to mean that this is the only possible frame of reference that can be used for this theory. We can recast the theory in terms of alternative frames of reference, an infinite number of which are available for our use. Equally valid, but independent characterizations under some conditions C, is what I have referred to as alternatives. The condition of Lebesgue square measure provides a very good basis for a particularly useful set of alternatives that can be utilized for acoustics and signal analysis. This paper outlines the theory of alternatives and sets out a few of the properties that such alternatives must possess.

When these concepts are applied to the very practical world of acoustics, we can begin to see some of our present endeavors in a slightly different light. It is clear that the attribute which we call time and the attribute which we call frequency are but two of the available alternative frames of reference. The reason why a description cannot be completely coprecise in "frequencyness" and "timeness" is because they are alternative representations of the same thing and are related through a mapping relationship that can be characterized by a zero curvature hyperplane.

It is my hypothesis that our ceaseless search for some description, be it Wigner distribution or other, which can somehow present frequency and time (or any Fourier transform related alternatives) on a common coordinate basis, is tied to perceptions which are cast in higher dimensional frames of reference. If the same process can be viewed from alternative frames of reference having different dimensionality, then it is not unreasonable for a person who was unaware of the existence of such alternatives to attempt "glueing together" two lower-dimensional alternatives in such a manner as to mimic certain attributes found in such a higher-dimensional

perception. This improperly assembled system is not a valid higher-dimensional system, but is an attempt to make sense out of two lower- dimensional alternative properties.

The Fourier transform, as powerful as it may be, can be seen to be nothing more than a specific recipe which can map a relationship from one frame of reference to a special alternative frame of reference having the same dimensionality. It cannot be used to map upward or downward in dimensionality, and hence can never be used to investigate the source of our anguish concerning joint behavior of attributes that map downward to time and frequency.

If the Fourier transform utilizes a mapping kernel that is characterized by a zero curvature hyperplane, then there should exist transforms utilizing higher curvature hypersurfaces. One such transform has been described in this paper and called the TDS transform because of its development from analysis of time delay spectrometry.

Time delay spectrometry may be broadly characterized as a method of system measurement in a stimulus-response situation. This stimulus utilizes a signal that has a defined total energy density and a defined exchange relationship between the two quadrature components of this total energy density. The exchange relationship determines the effective curvature of the mapping hypersurface as expressed in the host frame of reference. The response of this stimulus may be quite complicated in terms of the host frame of reference, but proper interpretation of this response in an alternative frame of reference may result in significant simplifications. The specific curvature of stimulus hypersurface in the host frame of reference is chosen such that certain properties of interest in the response to that stimulus may, in some measure, be optimally clustered in the alternative frame of reference and separated from undesired components. Selective isolation of the properties of interest, a process of spatial filtering within the alternative frame of reference, is to be performed within the TDS processor. Once extracted, the desired property may then be mapped back down to the host frame of reference for interpretation.

A number of practical TDS acoustic measurements have been presented as examples of this process. These examples are admittedly quite simple, inasmuch as the results have been directed toward system characterizations that can be expressed as well-known time domain or frequency domain properties.

ACKNOWLEDGEMENTS

I wish to thank the many persons who have provided moral support and encouragement during the early development of this still evolving theory. I am particularly grateful to the late Dr. Harry F. Olson, upon whose recommendations the first papers on TDS were published in the Journal of the Audio Engineering Society and who provided early encouragement for further development of the measurement process. My good friend, Dr. Ludwig Sepmeyer has given me continuing technical support and encouragement from the very beginning of my work in this field and has been of inestimable value to me. I wish to thank Mr. Cecil Cable, who was the first to apply TDS measurements to architectural acoustics, and to Mr. Don Davis, whose enthusiastic support and educational programs have significantly expanded the acceptance of TDS for acoustic measurements. And, finally, I wish to thank Professor Richard Waterhouse, who encouraged the writing of this paper; and the patient support of my professional colleagues at the Jet Propulsion Laboratory, Dr. Dennis Le Croissette and Dr. James Rooney, who gave me the needed support to accomplish the task.

REFERENCES

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

2. R. C. Heyser, A Signal Biasing Output Transformerless Power Amplifier, J. Audio Eng. Soc., vol. 8, p. 185, (1960).

3. G. W. Stewart. Problems Suggested by an Uncertainty Principle in Acoustics, J. Acoust. Soc. Am., vol. 2, p. 235, (1931).

4. D. Gabor, Theory of Communication, Proc. Inst. of Elect. Eng., vol. 93, p. 425, (1946).

5. R. C. Heyser, Loudspeaker Phase Characteristics and Time Delay Distortion: Part I, J. Audio Eng. Soc., vol. 17, p. 30, (1969): Part II, J. Audio Eng. Soc., vol 17, p. 130, (1969).

6. R. C. Heyser, Determination of Loudspeaker Signal Arrival Times: Part I, J. Audio Eng. Soc., vol. 19, p. 734, (1971); Part II, J. Audio Eng. Soc., vol. 19, p. 829, (1971); Part III, J. Audio Eng. Soc., vol. 19, p. 902, (1971).

7. R. C. Heyser, The Delay Plane, Objective Analysis of Subjective Properties: Part I, J. Audio Eng. Soc., vol. 21, p. 690, (1973); Part II, J. Audio Eng. Soc., vol. 22, p. 786, (1973).

8. R. C. Heyser, Geometric Consideration of Subjective Audio, J. Audio Eng. Soc., vol. 22, p. 674, (1974).

9. R. C. Heyser, Geometry of Sound Perception, Preprint No. 1009, 51st. Convention, Audio Eng. Soc., May 1975.

10. R. C. Heyser, Perspectives in Audio Analysis: Changing the Frame of Reference: Part I, J. Audio Eng. Soc., vol. 25, p. 660, (1976); Part II, J. Audio Eng. Soc., vol. 25, p. 742, (1976).

11. R. C. Heyser, Concepts in the Frequency and Time Domain of Loudspeakers, Proc. Inst. Radio and Electronic Eng. (Australia), vol. 37, p. 67, (1976).

12. R. C. Heyser, Fuzzy Alternatives, J. Audio Eng. Soc., vol. 26, p. 151, (1978).

13. D. H. Le Croissette and R. C. Heyser, Transmission Ultrasound Soft Tissue Visualization, Proc. 3rd. Int. Conf. on Med. Phys., p. 35.5, 1972.

14. R. C. Heyser and D. H. Le Croissette, Transmission Ultrasonography. 1973 IEEE Ultrasonics Symposium Proceedings, Cat # CHO 807-8SU, p. 7, 1974.

15. R. C. Heyser and D. H. Le Croissette, A New Ultrasonic Imaging System Using Time Delay Spectrometry, Ultrasound in Med. and Bio., vol. 1, p. 119, (1974).

16. L. S. Fishman, R. C. Heyser and D. H. Le Croissette, Ultrasonic Transmission Measurements on Human Brain, Radiology, vol. 112, p. 211, (1974).

17. D. H. Le Croissette and R. C. Heyser, Attenuation and Velocity Measurements on Tissue using Time Delay Spectrometry, NBS spec. Pub. 453, p. 81, (1976).

18. R. C. Heyser, Ultrasonic Imaging, Intrauterine Fetal Visualization: A Multidisciplinary Approach, Exerpta Medica, Amsterdam, p. 226, 1976.

19. D. H. Le Croissette, R. C. Heyser, J. D. Hestenes, P. M. Gammell and J. A. Rooney, A Swept-Frequency Medical Ultrasonic Imaging System, Proc. of 35th Ann. Conf. on Eng. in Med. and Bio., vol. 24, p. 202, Alliance for Eng. in Med. and Bio., Bethesda, MD, 1982.

20. P. M. Gammell, J. A. Roseboro, R. C. Heyser, D. H. Le Croissette, and R. L. Wilson, Swept Frequency Ultrasonic Measurements of Tissue Characteristics, Proc. of 30th. Ann. Conf. on Eng. in Med. and Bio., vol. 19, p. 229, Alliance for Eng. in Med. and Bio., Bethesda, MD 1977.

21. D. H. Le Croissette, R. C. Heyser and R. L. Wilson, The Characteristics of Soft Tissue using Ultrasonic Energy, Proc. Am. Assoc. of Med. Instrum., 12th Annual Meeting, p. 241, 1977.

22. D. H. Le Croissette, R. C. Heyser, P. M. Gammell, J. A. Roseboro and R. L. Wilson, The attenuation of selected Soft tissue as a Function of Frequency, 2nd. Int. Symposium on Ultrasonic Tissue Charaterisation, NBS Spec. Pub. 525, p. 101, 1979.

23. P. M. Gammell, D. H. Le Croissette and R. C. Heyser, Temperature and Frequency Dependence of Ultrasonic Attenuation in selected Tissues, Ultrasound in Med. and Bio., vol. 5, p. 269, (1979).

24. H. Biering and O. Z. Pederson, System Analysis and Time Delay Spectrometry Tech. Rev. of Bruel and Kjaer, (Part I), No. 1, (1983) and (Part II), No. 2, (1983).

25. T. S. Kuhn, The Structure of Scientific Revolutions, Int. Encyclopedia of Unified Science, vol. 2, No. 2, Univ. of Chicago Press, Chicago, IL, 1970.

26. E. C. Titchmarsh, Fourier Integrals, p. 128, Clarendon Press, Oxford, 1948.

27. P. Butzer and R. Nessel, Fourier Analysis and Approximation, pp. 310-315, Academic Press, NY., 1971.

28. E. Stein and G. Weiss, Introduction to Fourier Analysis on Euclidean Spaces, pp. 186-188, Princeton University Press, Princeton, NJ, 1971.

29. G. F. Simmons, Introduction to Topology and Modern Analysis, pp. 16-17, McGraw Hill, New York, NY, 1963.

30. M. Jammer, The Philosophy of Quantum Mechanics, pp. 2-9, John Wiley and Sons, New York, NY, 1974.

31. R. C. Heyser, Determining the Acoustic Position for Proper Phase Response of Transducers, J. Audio Eng. Soc., vol. 32, p. 23, (1984).

32. E. L. Hamilton, The Elastic Properties of Marine Sediments, J. Geophysi. Res., vol. 76, p. 579 (1971).

33. E. L. Hamilton, Compression-Wave Attenuation in Marine Sediments, Geophysics, vol. 37, p. 620, (1972).

34. D. L. Miller and W. L. Nyborg, Theoretical Investigation of the Response of Gas-filled Micropores and Cavitation Nuclei to Utrasound, J. Acoust. Soc. Am., vol. 73, p. 1537 (1983).

35. J. A. Rooney and R. C. Heyser, Use of Swept-frequency Ultrasonic Techniques for Quantitative Bubble Detection, Proc. Aerospace Med. Assoc., p. 120, (1983).

36. C. S. Clay and H. Medwin, Acoustic Oceanography: Principles and Applications, pp. 194-203, John Wiley and Sons, New York, NY, 1977.

37. A. Prosperetti, Nonlinear Oscillations of Gas Bubbles in Liquids: Steady State Solutions, J. Acoust. Soc. Am., vol. 56, p. 878, (1974).

United States Patent Office

3,466,652
Patented Sept. 9, 1969

1

3,466,652

TIME DELAY SPECTROMETER

Richard C. Heyser, Tujunga, Calif., assignor to California Institute of Technology, Pasadena, Calif., a corporation of California

Filed Jan. 15, 1968, Ser. No. 697,824
Int. Cl. G01s 9/23
U.S. Cl. 343—14 15 Claims

ABSTRACT OF THE DISCLOSURE

Apparatus for measuring the amount of sound or electromagnetic radiation from a test object at various frequencies while the test object is in an environment containing radiation reflecting objects whose output could easily be confused with output from the test object. In one application for measuring sound reflections from a test object, the apparatus comprises a loudspeaker driven by a sweep frequency oscillator and a microphone whose output is filtered by a sweep frequency filter. The passband frequency of the filter follows the frequency of the oscillator by a delay equal to the time required for sound to travel from the loudspeaker directly to the object and directly back to the microphone. Accordingly, the filter is always tuned to pass the frequency of sound waves following this direct path, and to rejection sound waves which arrive at a later time when the filter has already passed on to a new frequency.

BACKGROUND OF THE INVENTION

This invention relates to apparatus and methods for measuring the spectral response of an object.

It is often necessary to measure the amount of radiation emanating from a body at various frequencies. For example, the output of a loudspeaker at various frequencies, or the acoustical reflection or transmission characteristics of a door or other object at various frequencies must often be determined. It might appear that the acoustical output of an object could be determined by energizing it with a sine wave at one frequency and measuring the sound radiation at a location near the object, repeating this procedure for various frequencies within the band of interest. However, the object is generally surrounded by many other objects which reflect the sound waves and prevent the determination of the output of the particular object under investigation.

Special anechoic test chambers have been constructed for providing an environment free of extraneous reflecting objects, for testing a single object. However, these facilities are expensive, particularly in the case of anechoic chambers for electromagnetic radiation such as radio waves. Furthermore, there are many situations where the characteristics of an object in a particular environment must be measured, where an anechoic test area cannot be used. For example, the object may comprise a wall area of an auditorium, and it may be desired to measure the reflection characteristics of that wall independently of the floor, ceiling or other wall areas.

OBJECTS AND SUMMARY OF THE INVENTION

One object of the present invention is to provide apparatus which measures the acoustical characteristics of an object or area located at a predetermined distance.

Another object of the invention is to provide apparatus for measuring the amplitude of radiation from an object at various frequencies.

Still another object of the invention is to provide apparatus for measuring the phase shift of radiation from an object at various frequencies.

2

In accordance with the present invention, there is provided apparatus for measuring the radiation from a test body by energizing it to produce radiation which varies rapidly in frequency. A detector is located a fixed distance from the test body. The detector has a narrow passband filter whose passband frequency tracks the frequency of radiation at the test body. The filter passes only that radiation which is received directly from the test body. This occurs because the passband of the filter reaches the frequency of radiation which previously emerged from the test body at the same time as the radiation, traveling through the air or other medium, reaches the detector and filter. Radiation reflected or otherwise made to emerge from any other object in the environment takes a shorter or longer path and arrives at a time when the filter is not tuned to pass it.

In one application of the invention, the response of a loudspeaker is tested in an ordinary room containing sound reflecting surfaces, by energizing the loudspeaker with signals from a sweep frequency oscillator. A microphone is positioned a fixed distance from the loudspeaker to detect sound from it. The output of the microphone is passed through a tunable bandpass filter. The filter is tuned or driven to track the frequency of the sweep frequency oscillator, the offset in tracking being equal to the time required for sound to travel from the loudspeaker through the air directly to the microphone. Thus, when the sound waves of a particular frequency emerging from the loudspeaker reach the microphone, the filter is tuned to that frequency. Sound reflected from objects in the room reach the microphone at a later time when the filter is tuned to a different frequency, and they do not pass through the filter. The amplitude of the filter output can be displayed as a function of the frequency of the oscillator at each instant, to show the spectral response of the loudspeaker.

The invention can be used to measure the spectral response of a passive test object in an environment of other objects by projecting radiation at the objects and measuring the amplitude of radiation emerging from just the test object. For example, it may be desired to measure the acoustical reflection or transmission coefficients of a door at various frequencies when the door is in place. This requires that reflections or transmissions from the sealing areas around the door be eliminated. In accordance with this invention, a loudspeaker and microphone are set up on directly opposite sides of the door to measure sound transmission through it. The loudspeaker is energized by a sweep frequency oscillator and the microphone output is filtered by a tunable filter. The passband of the filter is offset from the oscillator so that the filter reaches a previous oscillator frequency after a delay equal to the transit time of sound directly from the loudspeaker to the door and then to the microphone. The sweep is fast enough and the passband narrow enough to eliminate sound taking a longer path, such as a path passing through the seal around the door.

The invention can be used to detect objects such as pipes lying in the ground at any particular depth, to detect defects in materials at particular depths, and in many other applications. Furthermore, electromagnetic radiation may be used in the same manner as sound radiation, by replacing a loudspeaker and microphone by antennas.

In one embodiment of the invention, the sweep frequency oscillator comprises a linear voltage controlled oscillator which is driven by a circuit that generates ramp voltages. The output of the sweep frequency oscillator, whose frequency varies at a constant rate, is delivered to a loudspeaker directed at a test object. The received radiation is detected by a microphone. The output of the microphone is delivered to a multiplier or modulator circuit which also receives the output of the sweep frequency oscillator. The output of the multiplier is equal to the difference in frequency between the transmitted and received radiation. This difference frequency is constant for a given radiation path.

The output of the multiplier is delivered to a narrow band filter, whose center frequency is equal to that difference frequency which is obtained for sound waves following a particular path. The filter output is rectified and delivered to the Y or vertical input of an oscilloscope for indicating the amplitude of the filtered signals at every instant. The X or horizontal sweep of the oscilloscope is driven by the same ramp voltage which drives the variable frequency oscillator. The oscilloscope shows the amplitude response to the object being tested at each frequency.

The novel features of the invention are set forth with particularity in the appended claims. The invention will best be understood from the following description when read in conjunction with the accompanying drawings.

BRIEF DESCRIPTION OF THE DRAWINGS

FIGURE 1 is a simplified block diagram of a time delay spectrometer constructed in accordance with the invention;

FIGURES 2A through 2D illustrate several applications of the invention;

FIGURE 3 is a more detailed block diagram view of a time delay spectrometer of the type shown in FIGURE 1; and

FIGURE 4 is a block diagram of a time delay spectrometer constructed in accordance with the invention which indicates phase delay at various frequencies.

DESCRIPTION OF THE PREFERRED EMBODIMENTS

FIGURE 1 illustrates a simplified time delay spectrometer, showing its principles of operation. The apparatus is used for determining the coefficient of reflection at various frequencies of a test object 10 which is surrounded by various interfering objects, one of which is shown at 12. The reflection characteristics of the test object 10 are determined by generating sound waves of various frequencies with a loudspeaker 14. The sound waves reflected by both objects 10 and 12 are received by a microphone 16. In order to measure the reflection characteristics of the test object 10, it is necessary to differentiate the radiation it reflects from the radiation reflected by other objects such as interfering object 12.

In accordance with the present invention, sound waves which do not represent direct reflections from the test object are differentiated by the fact that they arrive at a different time (generally later) from the sound waves received from the test object. The manner by which the radiation which traverses a shorter or longer path than the radiation to be detected is excluded is a principal novel feature of the invention.

The loudspeaker 14 is energized by the output of a sweep frequency oscillator 18 whose frequency varies continuously with time and at a constant rate. The oscillator 18 may contain a sweep voltage generator 11 which energizes a voltage controlled oscillator 13. The output of the sweep frequency oscillator begins at a low level such as 200 Hz. and increases to a high level such as 20 kHz. during a one second interval, and then begins again at 200 Hz. The time required for the sound waves to travel from the loudspeaker 14 to the tested object 10 and then to the microphone 16 is a period Δt; in one application, this period may be on the order of 4 milliseconds during which time the oscillator output frequency changes by approximately 80 Hz. Accordingly, the output of the microphone 16 includes a signal 16' representing reflections of the test object, which tracks the oscillator output frequency with an offset, such as 80 Hz., representing the frequency change during the interval Δt.

The output of the microphone 16 is delivered to a tunable filter circuit 19 having a narrow passband, the passband center frequency varying with time. The tunable filter 19 comprises a mixer or modulator 20 which mixes the transmitted frequency from oscillator 18 with the output from the microphone. The mixer output includes the difference frequency Δf which is equal to the change in frequency of the oscillator output during the interval Δt required for sound to pass from the loudspeaker 14 to the test object 10 and then to the microphone 16. Only this difference frequency Δf passes through a narrow bandpass filter 21 of the tunable filter circuit. The tunable filter 19 therefore acts as a filter which tracks the sweep oscillator output $f(t)$ by an offset frequency Δf, i.e., it acts as a filter of a frequency $f(t)-\Delta f$. The output of the filter 19 therefore has an amplitude proportional to the amount of sound reflected from the tested object.

Sound waves reflected from an interfering object 12 travel along a path which is different from the path of waves received from the test object. As a result, waves from the interfering object reach the microphone 16 after a delay of more than Δt. When these waves arrive at the microphone 16, the tunable filter 20 has already been tuned to a different frequency, so that the reflections from the interfering object do not pass through the filter. Accordingly, they do not affect the output of the filter. In a similar manner, waves traveling directly from the loudspeaker 14 to the microphone 16 arrive too early to pass through the filter and affect its output.

The output of the tunable filter 19 represents the amplitude of sound reflections from the test object, the amplitude at any instant representing the coefficient of reflection for a particular frequency. If a direct current meter 24 with very slow response time (e.g., less than a second for 1 second sweeps) is energized with the filter output, the meter indicates the average coefficient of reflection of the object for the range of frequencies. If it is desired to determine the reflection coefficient for any particular frequency, this can be done by using the filter output to drive the vertical or Y input of an oscilloscope 26. The horizontal or X axis of the oscilloscope is swept in synchronism with the change in frequency of the sweep oscillator 18, as by driving it with the output of the sweep voltage generator 11.

The apparatus of FIGURE 1 can be used in a number of ways in addition to determining the reflection coefficients of a known object. As shown in FIGURE 2A, a sonic transducer such as a vibrator or shaker 15 can be directed into the ground, and a geophone 17 positioned to receive reflected sound waves to locate an object 28 such as a pipe having a known frequency response characteristic. A more precise discrimination between various objects can be made than has been possible heretofore. The approximate velocity of sound in typical ground compositions, water, and other media, is known, and the delay Δt can be adjusted for a desired depth in the particular medium.

Another use of the apparatus, illustrated in FIGURE 2B, is to enable the testing of a loudspeaker 14B or other radiating device without the necessity for an anechoic chamber. This can be done by energizing the loudspeaker 14B with the output from the sweep frequency oscillator and adjusting the delay Δt to the time required for sound to travel directly from the loudspeaker to a microphone 16B. The angle A between the axis of the loudspeaker and the location of the microphone can be varied to measure off-axis output.

Still another use of the apparatus, indicated in FIGURE 2C, is to measure the velocity of propagation of waves in an unknown medium 21. For example, the type of composition of a ground area is indicated by the velocity of propagation of sound waves of various frequencies. A sonic transducer 23 is driven by a sweep frequency oscillator 18C for introducing sound waves of various frequencies into the ground at one location. A detector 25 is located a predetermined distance from the transducer 23. The output of the detector is delivered to a tunable filter 19C. The passband of the filter is the difference frequency $f(t)-\Delta f$ between the frequency $f(t)$ of the oscillator 18C and an offset frequency Δf of a filter tuning means 27. The tuning means 27 can be manually adjusted to vary the offset. The offset frequency Δf is slowly increased from a low value, and the amplitude at each frequency is observed on an oscilloscope. The time required for sound to travel the shortest path between the transducer 23 and the detector 25 at various frequencies is indicated by the offset required to obtain a considerable amplitude output for that frequency.

While examples have been given for sound radiation, it should be recognized that the same principles apply to electromagnetic radiation. For radio waves, antennas are substituted for the loudspeaker and microphone, or other sonic transducers, and in many situations higher frequencies are used. FIGURE 2D shows a setup for measuring the output of a transmitting antenna 29 at a particular angle D by detecting the radiation with a receiving antenna 31 positioned at that angle. This can be done in the presence of a reflecting body 33 in the environment.

FIGURE 3 is a more detailed block diagram of a circuit for generating a sweep frequency to energize a transmitter to control a tunable filter. The sweep frequency oscillator 30 comprises a sweep voltage generator 32 that generates ramp voltages. The sweep frequency oscillator also includes a high frequency voltage controlled oscillator, or VCO 36 which generates a sinusoidal output 36' having a frequency linearly proportional to the voltage input thereto. The output of the VCO repeatedly sweeps between a lower frequency, such as 100 kHz. and a higher frequency such as 120 kHz. in a sweep time such as one second, the frequency changing at a constant rate during each sweep.

The output of the VCO 36 is delivered to a mixer circuit 38, where it is modulated by the output of a manually tunable oscillator 39. The manually tunable oscillator 39 has a constant frequency output 39' of a frequency $100k-F_0$. The mixer 38 generates an output 38' having a frequency equal to the difference between the outputs of the VCO 36 and the manually tunable oscillator 38 (the sum frequencies are filtered out by a filter, not shown). The difference frequency sweeps between F_0 and $20k+F_0$ in synchronism with the sweep output of the VCO. This difference frequency output 38' is delivered to a loudspeaker 40 to drive it.

The loudspeaker 40 directs its output to an object 42 which reflects some of the sound waves to a microphone 44. The length of the direct path from the loudspeaker 40 to the object 42 and from thence to the microphone 44 is accurately known. During the time Δt required to traverse this path, the loudspeaker frequency changes by an amount Δf equal to F_0 (F_0 is chosen to be this value). The output of the microphone 44 is delivered to a tunable filter 46 for passing only those received frequency components which represent reflections from the object 42.

The microphone output is first delivered to a modulator or mixer circuit 48 of the tunable filter the mixer circuit 48 also receives a signal from the VCO 36 to the sweep frequency oscillator. The mixer circuit 48 delivers a signal which contains the difference between the frequencies of its inputs. The portion 48' of this signal which represents reflections from the test object 42 is at a known frequency, such as 100 kHz. This difference signal 48' is passed to a narrow band filter 54 which passes only a very narrow frequency component, such as the 100 kHz. component. The filtered output passes through a rectifier 58 to the tunable filter output 60. The output at 60 has an amplitude at all times proportional to those inputs to microphone 44 which represent reflections from the object 42.

In considering the operation of the circuit of FIGURE 3, the tunable filter 46 acts like a narrow passband filter whose frequency can be varied. An output from the sweep frequency oscillator performs this variation so that the tunable filter passband tracks the frequency of the loudspeaker output by the required offset. This is accomplished in the actual circuit by mixing the transmitted and received frequencies and passing the difference frequency through a constant frequency filter which is set to pass only the required offset frequency.

The output signals of the tunable filter can be viewed on an oscilloscope 62 by connecting the output 60 to the Y, or vertical input of the oscilloscope. The X or horizontal sweep is taken as the voltage output of the voltage sweep generator 32 in the sweep frequency oscillator.

It should be noted that at any given instant, the X axis sweep of the oscilloscope 62 represents the frequency being transmitted, while the Y axis represents the amplitude of the signals transmitted at an earlier time. Because of this offset, the highest frequencies from the test object 42 are received at a time when the oscilloscope begins a new sweep, and the object response to these frequencies cannot be measured. The response at a high frequency can be measured by increasing the maximum frequency within the sweep band, or by sweeping in a reverse direction, that is, by starting each sweep at the highest frequency and decreasing linearly to the lowest frequency.

The rejection of sound radiation from other objects in the environment, by the circuit of FIGURE 3, is based on the different time or arrival of radiation from these other objects. The output of the microphone 44 may contain additional frequency components, such as that shown at 44" which represents radiation following a longer path. This will give rise to a component 48" at the mixer output, which is too high in frequency to pass through the narrow band filter 54. In order to adjust the apparatus to a different radiation path length between the loudspeaker, object and microphone, the output 39' of the manually tunable oscillator is adjusted. The path length for which radiation will be detected varies in direct proportion to the component F_0 of the output 39' from the manually tunable oscillator.

While information about the amplitude of radiation from a particular body is useful, information about the phase of radiation from the body is also of importance. Knowledge about the exact phase shift indicates the material at the surface of the test body, or the exact range of the body. For example, a precise determination of the phase shift of reflected radiation can indicate the distance of an object within a small fraction of a wavelength of the highest frequency radiation.

FIGURE 4 illustrates apparatus for accurately measuring the phase shift of radiation from a body. The apparatus is similar to the apparatus of FIGURE 3 in that it contains a VCO 70 which is driven by a sweep voltage generator 72. The VCO output is modulated by a constant frequency signal in a mixer 74, whose output is delivered to a loudspeaker 76. The loudspeaker radiates sound which reaches a microphone 78, and the microphone output is delivered to another mixer 80. The second mixer modulates the microphone output with the VCO output, and the difference passes to a narrow band filter 82. The filter output can be used to measure the frequency response of a test object in the manner described above in connection with FIGURE 3, or it can be used to measure phase shift using the rest of the circuit of FIGURE 4.

In the circuit of FIGURE 4, the output of the narrow bandpass filter 82 is passed through a limiter 84 to remove amplitude variations. The limiter output is delivered to a phase detector 86 for detecting the phase difference between the transmitted and received radiation. The phase of the transmitted radiation is derived from a divider or countdown circuit 88 whose frequency is a fraction of the frequency of a crystal reference oscillator 90, and in a controlled phase relationship. The crystal oscillator output also is divided by a frequency synthesizer 92 to obtain the mixer signal which is delivered to the loudspeaker. The reason why the frequency synthesizer 92 is used is to enable the generation of any frequency within a range, with a very accurately controlled frequency and with a predetermined phase relationship to the crystal oscillator output.

The phase difference between the transmitted and received radiation is indicated by the output from the phase detector 86. The phase difference for any frequency within the band of transmitted frequencies (except for a small frequency band at the low end of each sweep, as described above) can be measured on an oscilloscope. The phase detector output 86' is delivered to the Y-axis input of an oscilloscope 94, whose X axis is swept by the output of the sweep voltage generator 72. The oscilloscope then displays the phase shift along the frequency band of transmitted frequencies.

As described above, the invention allows for the detection of radiation of known frequency from a particular object while rejecting radiation of the same frequency from other objects, by reason of the different path lengths. While this can be accomplished with a continually swept frequency, it can also be performed with radiation which repeatedly steps in frequency. In such a case, the filter can also step in frequency, or be swept. The problem with such an arrangement is that the radiation at any particular frequency has a starting transient. It requires an appreciable time for the starting transient to die down, and this time often may exceed the difference in time for different radiation path lengths in the testing environment. On the other hand, a continuously swept frequency (during sweep periods) generally eliminates transients, and better discrimination is possible.

The rate of frequency variation must be high enough so that radiations following substantially different paths, taking them to two different bodies in the environment, arrive at times when the filter is at two substantially different frequencies. The difference in center frequency of the filter passband at the different arrival times must be sufficient to exclude the unwanted radiation. The required difference depends upon the width of the filter passband. For example, for objects close enough to each other to provide a difference in path length of eight feet, and a filter of 140 Hz passband width for 6 db rejection, a sweep rate of 20 kHz per second will be sufficient. A passband approximately equal to the square foot of the sweep rate yields optimum special discrimination while retaining optimum frequency discrimination. For a greater frequency discrimination, i.e., to measure response at frequencies close to each other, a filter of narrower bandwidth must be used, together with a lower sweep rate. This reduces the ability to discriminate against objects close to the desired object, i.e., spacial discrimination. Alternatively, spacial discrimination can be increased at the cost of frequency discrimination.

Although particular embodiments of the invention have been described and illustrated herein, it is recognized that modifications and variations may readily occur to those skilled in the art, and consequently, it is intended that the claims be interpreted to cover such modifications and equivalents.

What is claimed is:

1. Apparatus for time delay spectrometry comprising:
transmitting means for transmitting radiation which varies in frequency with time;
receiving means for receiving said radiation;
means coupled to said transmitting and receiving means for indicating the amplitude of frequency components of radiation received by said receiving means which have a frequency equal to the frequency of said radiation at said transmitting means at a predetermined previous time; and

means responsive to the instantaneous frequency of radiation transmitted by said transmitting means for indicating the relationship between the amplitudes of said frequency components and the frequency of said rotation;

2. Apparatus as defined in claim 1 wherein:
said transmitting means comprises means for varying the frequency of said radiation substantially continuously during sweep periods, whereby to avoid transient signals.

3. Apparatus as defined in claim 1 wherein:
said transmitting means comprises a sweep frequency oscillator for generating a signal whose frequency varies substantially at a constant rate during sweep intervals and a transducer driven by the output of said oscillator; and
said means for detecting comprises tunable filter means coupled to said receiving means, and means coupled to said oscillator for controlling the passband frequency of said filter means.

4. Apparatus as defined in claim 1 wherein:
said transmitting means comprises a sweep frequency oscillator whose output varies at a substantially constant rate during sweep intervals; and,
said means for detecting frequency components comprising:
modulator means having a first input coupled to said receiving means, a second input coupled to said transmitting means for receiving signals of a frequency which varies at the same rate as the frequency of transmitted signals during at least a portion of said sweep intervals, and an output; and
a narrow passband filter coupled to said output of said modulator means.

5. Apparatus for measuring the amplitude of radiation from a first body, independently of radiation of the same frequency from another body, comprising:
means for energizing said first body to produce radiation which continuously varies in frequency during predetermined intervals;
radiation detecting means spaced from said bodies for detecting, at each instant, substantially only the components of said radiation which have a frequency equal to the frequency of radiation emerging from said first body at a time which is previous by a period equal to the time required for said radiation to traverse the distance between said first body and said detecting means; and
means coupled to said radiation detecting means for indicating substantially the relationship between the amplitude of said components detected by said detecting means and their frequency.

6. Apparatus as defined in claim 5 wherein:
said means for energizing comprises means spaced a predetermined distance from said first body for generating radiation; and
said radiation detecting means comprises transducer means for converting radiation passing a particular area into electrical signals, tunable filter means coupled to said means for energizing for passing only frequency components representing the frequency of said radiation at said first body at a time which is previous by a period substantially equal to said time required for said radiation to traverse the distance between said means for generating radiation and said first body plus the time required for said radiation to traverse the distance between said first body and said detecting means, and means for indicating the amplitude of said frequency components.

7. Apparatus as defined in claim 5 wherein:
said means for energizing includes means for varying said frequency at a constant rate; and
said radiation detecting means comprises filter means

coupled to said means for generating to pass only components of received radiation having a frequency which differs from the frequency of said radiation at said means for energizing by a predetermined difference frequency, and means for indicating the amplitude of components of said difference frequency.

8. A method for indicating the radiation characteristics of an object in an environment of other bodies which can radiate comprising:

applying radiation which varies rapidly in frequency, to said object;

detecting radiation at a point spaced from said object; and

filtering said detected radiation to determine the amplitude of only the components having a frequency equal to the frequency of radiation at said object at a predetermined time previous to its detection.

9. A method as defined in claim 8, including:

indicating the relationship between the amplitude of said components in said detected radiation and the frequency of radiation at said object which produced those components.

10. A method as defined in claim 8, including:

indicating the phase relationship between said component in said detected radiation and the phase of said applied radiation.

11. Apparatus for time delay spectrometry comprising:

transmitting means for transmitting radiation which varies in frequency with time;

receiving means for receiving said radiation; and

means coupled to said transmitting and receiving means responsive to frequency components of radiation received by said receiving means which have a frequency equal to the frequency of said radiation at said transmitting means at a predetermined previous time, for indicating the phase shift between said frequency components and said radiation at said transmitting means at a predetermined previous time.

12. Apparatus for time delay spectrometry comprising:

transmitting means including:

means for generating a sweep signal which varies linearly with time,

means for generating an offset signal of constant offset frequency,

a mixer for mixing said sweep and offset signals, and

means for transmitting an output of said mixer;

receiving means for receiving said output transmitted by said means for transmitting; and

means coupled to said transmitting and reeciving means including:

modulator means for modulating signals from said receiving means with signals from said transmitting means which vary at the same rate as said sweep signal,

band pass filter means coupled to said modulator means for passing a particular freqeuncy, and

phase detection means coupled to the output of said band pass filter means and said means for generating an offset signal, for generating a signal indicating the phase difference between the output of said band pass filter means and a signal derived from said offset signal.

13. Apparatus for measuring the amplitude of radiation from transducer means, independently of radiation of the same frequency from another body, comprising:

means for generating an electrical signal which varies in frequency with time;

means for applying said electrical signal to said transducer means; and

radiation detecting means spaced from said transducer means and said another body for detecting substantially only the components of said radiation which have a frequency equal to the frequency of radiation emerging from said transducer means at a time which is previous by a period equal to the time required for said radiation to traverse the distance between said tarnsducer means and said detecting means.

14. Apparatus for measuring the response characteristics of a body comprising:

a variable frequency oscillator, including means for sweeping its frequency along a band of frequencies at a predetermined constant rate with respect to time;

means for applying the output of said oscillator to said body to cause the emanation of radiation from said body;

transducer means for detecting said radiation;

modulator means for modulating the output from said transducer with a signal which varies at said predetermined constant rate;

pass band filter means coupled to said modulator means for passing frequency components of a particular frequency; and

means coupled to said filter means and said oscillator means for indicating the relationship between the frequency of said radiation which was applied to said body at a predetermined instant and the amplitude of said frequency components, whereby to obtain the amplitude response of said body at various frequencies.

15. Apparatus for time delay spectrometry comprising:

first oscillator means for generating signals which vary in frequency at a constant rate during predetermined periods;

second oscillator means for generating a signal which has a constant frequency;

first mixing means for mixing the outputs of said first and second oscillator means to obtain a signal for radiating which is offset in frequency from the output of said first oscillator means by a predetermined frequency dependent upon the output of said second oscillator means;

second mixing means for mixing detected radiation signals with the output of said first oscillator means;

filter means coupled to said mixing means for passing only components of a predetermined frequency;

means coupled to said second oscillator means for deriving a signal having a frequency which is a precisely constant multiple of the freqeuncy delivered by second oscillator means to said first mixing means; and

phase detection means for indicating the phase between said components from said filter means and the signal from said means for deriving a signal.

References Cited

UNITED STATES PATENTS

2,553,907	5/1951	Fleming-Williams et al.	343—14
3,016,513	1/1962	Van Dyke	340—3
3,343,167	9/1967	Rademacher	343—17.5

RODNEY D. BENNETT, JR., Primary Examiner

M. F. HUBLER, Assistant Examiner

U.S. Cl. X.R.

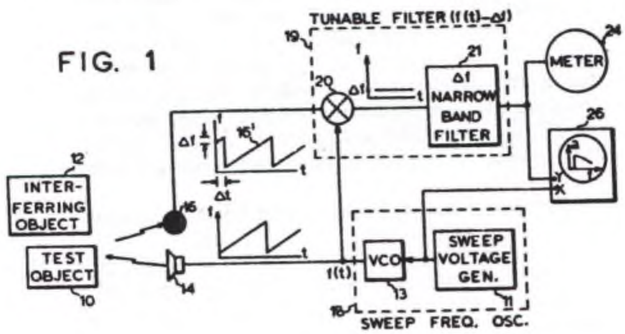
340—3; 343—17.5

Sept. 9, 1969 R. C. HEYSER 3,466,652

TIME DELAY SPECTROMETER

Filed Jan. 15, 1968 3 Sheets-Sheet 1

FIG. 1



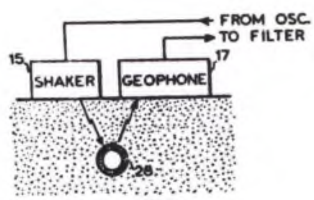


FIG. 2A

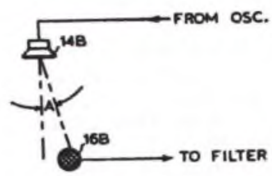
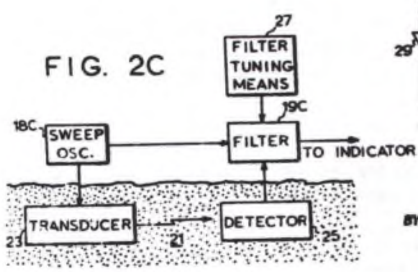


FIG. 2B

FIG. 2C



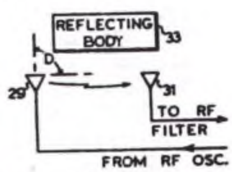


FIG. 2D

INVENTOR.
RICHARD C. HEYSER

BY Lindenberg & Freilich

ATTORNEYS

Sept. 9, 1969 R. C. HEYSER 3,466,652

TIME DELAY SPECTROMETER

Filed Jan. 15, 1968 3 Sheets-Sheet 3

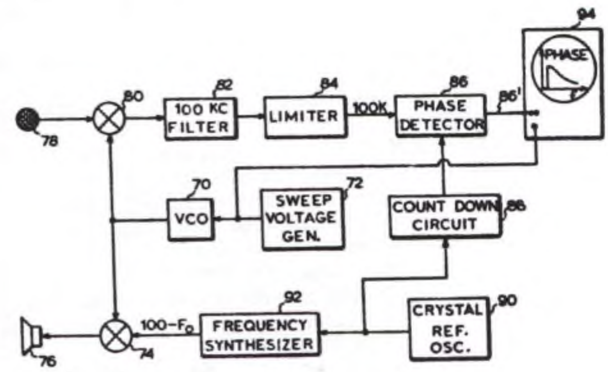


FIG. 4

INVENTOR.
RICHARD C. HEYSER

BY Lindenberg & Freilich

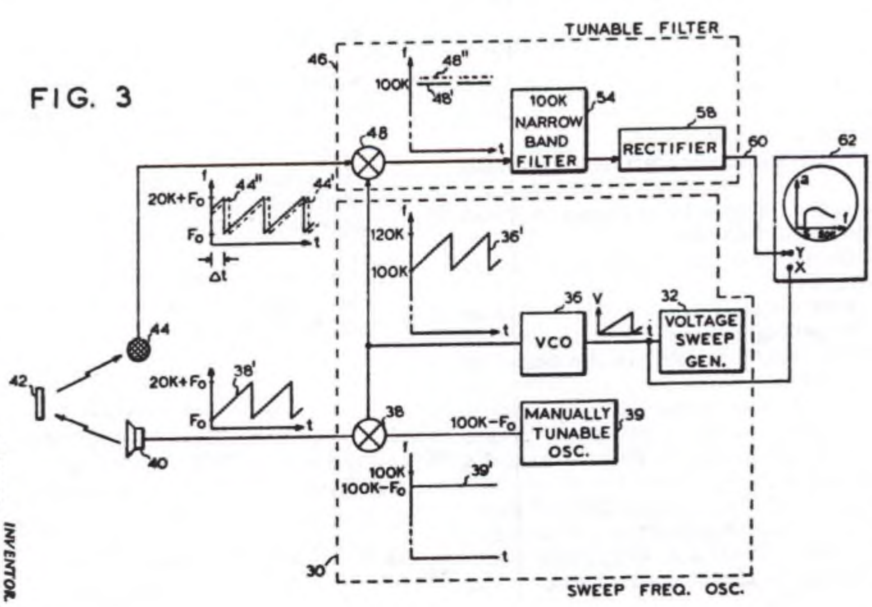
ATTORNEYS

Sept. 9, 1969 R. C. HEYSER 3,466,652

TIME DELAY SPECTROMETER

Filed Jan. 15, 1968 3 Sheets-Sheet 2

FIG. 3



INVENTOR.
RICHARD C. HEYSER

BY Lindenberg & Freilich

ATTORNEYS

United States Patent [19]

Heyser

[11] **4,287,578**

[45] **Sep. 1, 1981**

[54] **METHOD FOR SHAPING AND AIMING NARROW BEAMS**

[75] Inventor: **Richard C. Heyser, Tujunga, Calif.**

[73] Assignee: **The United States of America as represented by the Administrator of the National Aeronautics and Space Administration, Washington, D.C.**

[21] Appl. No.: **92,143**

[22] Filed: **Nov. 7, 1979**

[51] **Int. Cl.³** **G01S 15/04;** G01S 15/89

[52] **U.S. Cl.** **367/88;** 367/102; 367/100

[58] **Field of Search** .... 367/101, 88, 100, 102, 367/103; 343/17.2 PC

[56] **References Cited**

U.S. PATENT DOCUMENTS

3,302,162	1/1967	Rowlands	367/100
3,484,737	12/1969	Walsh	367/88
3,798,590	3/1974	Jacobson et al.	367/100
3,805,222	4/1974	Pekau	367/88
3,827,022	7/1974	Booth	367/104
3,890,618	6/1975	Speiser	367/100
3,969,725	7/1976	Couvillon et al.	343/17.2 PC
4,088,978	5/1978	Gilmour	367/88

OTHER PUBLICATIONS

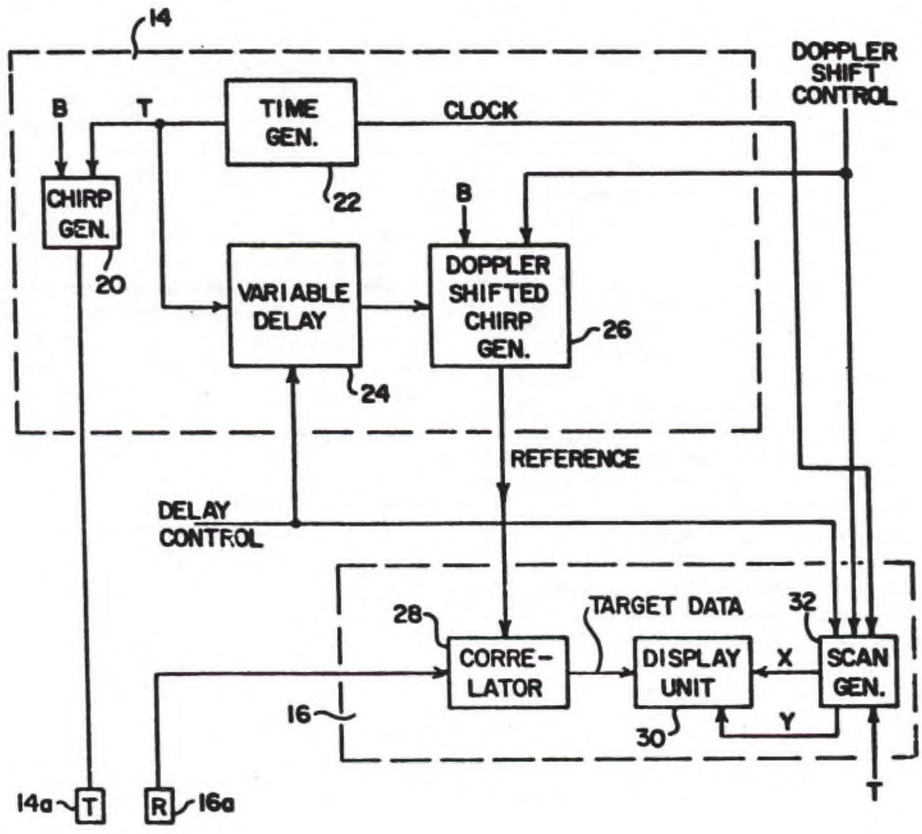
Nawar et al., "A Pulse Compression Echo Sounder for Ocean Bottom Surveys", *Journal of Geophysical Research*, Nov. 15, 1966, pp. 5279–5282.

Primary Examiner—Richard A. Farley
Attorney, Agent, or Firm—Monte F. Mott; John R. Manning; Paul F. McCaul

[57] **ABSTRACT**

A method and apparatus is disclosed for use of a linear frequency chirp in a transmitter/receiver (14/16) having a correlator to synthesize a narrow beamwidth pattern from otherwise broadbeam transducers when there is relative velocity between the transmitter/receiver (14/16) and the target. The chirp is so produced in a generator (20) in bandwidth, B, and time, T, as to produce a time-bandwidth product, TB, that is increased for a narrower angle. A replica of the chirp produced in a generator (26) is time delayed and Doppler shifted for use as a reference in receiver (16) for correlation of received chirps from targets. This reference is Doppler shifted to select targets preferentially, thereby to not only synthesize a narrow beam but also aim the beam in azimuth and elevation.

12 Claims, 8 Drawing Figures



METHOD FOR SHAPING AND AIMING NARROW BEAMS

ORIGIN OF THE INVENTION

The invention described herein was made in the performance of work under a NASA contract and is subject to the provisions of Section 305 of the National Aeronautics and Space Act of 1958, Public Law 85-568 (72 Stat. 435, 42 USC 2457).

TECHNICAL FIELD

This invention relates to a method and apparatus for synthesizing a narrow beamwidth pattern from otherwise broad beamwidth transducers when there is a velocity difference between the transmitter/receiver and a return source of wave energy, and for preferentially aiming the beam to select targets.

BACKGROUND ART

In ocean floor surveying and subbottom mapping with sonar for different commercial and noncommercial applications, it would be desirable to form and aim a narrow sonar beam from a submerged vessel having both a transmitter and receiver towed at some speed, such as 7 knots. It has been discovered that the Doppler effect produced in a frequency chirp may be used to advantage to aim a beam, and that control of the time-bandwidth product of the chirp can be used to form a narrow beam thereby to increase the effectiveness of sonar mapping and surveying.

Mapping is, of course, only one application of a sonar system embodying the invention. It may also be used in other applications, such as in selective target identification. For example, assume a command ship wishes to be able to determine the presence of a particular submarine in an area having many other targets in the water at about the same range. The echo returns from all the targets may make it very difficult to determine the presence of the submarine if all targets have about the same relative velocity as the submarine. If the particular submarine returns a transponder produced echo that is Doppler shifted more than expected echo returns from any other target in the water, the submarine can be easily identified by correlating the signal received from the transponder with its replica at the receiver. Still other applications will occur to those skilled in the art.

Although the invention will be first described in terms of synthesizing and aiming the narrow beam in real time, it can be readily appreciated that in practice the transmitted and received sonar signal may be recorded, such as on magnetic tape, for later processing. In the later processing, the technique to be described can be applied as though the data were being received in real time. An advantage is that the same data can be processed repeatedly, each time effectively aiming the beam in the direction of a different target. It should also be understood that the basic concept of the invention may be used with radar, since electromagnetic wave energy will respond to the same principles in respect to the Doppler effect occurring when the source and the receiver are in motion relative to each other, as with acoustical wave energy.

It will be shown that the ratio of the velocity of the relative motion to the speed of sound in water is an important factor in synthesizing a narrow sonar beam. The corresponding ratio of the relative velocity to the speed of light is likewise an important factor in synthesizing a narrow radar beam. As will be explained more fully hereinafter, the half angle of the beam decreases as that ratio increases, but since there is a practical limit in producing relative velocity that is commensurate to the speed of light, it is not generally practical to try to reduce the width of a radar beam for earth bound applications. However, spaceflight and satellite applications may produce a relative velocity that is commensurate to the speed of light to permit the present invention to be utilized. One application may be for a spaceship (or earth station) to track the position of another spaceship or satellite.

STATEMENT OF INVENTION

In accordance with the present invention, a linear frequency chirp is transmitted for a known interval and received as a Doppler shifted chirp from targets through a suitable transducer as the transmitter/-receiver moves relative to the target. When the Doppler shift is due solely to motion of the transmitter/-receiver, echo returns of the transmitted chirp will have a positive Doppler shift for targets in front of the transmitter/receiver, and a negative Doppler shift for targets behind the transmitter/receiver. Any one Doppler shift can be selected at the receiver by correlating a replica of the transmitted chirp that is delayed and Doppler shifted. Thus, by properly delaying and Doppler shifting the transmitted chirp for use as a replica in the receiver for correlation, the beam may be aimed in azimuth and elevation to targets selected preferentially. To form a narrow beamwidth pattern, the time-bandwidth product, TB, of the transmitted chirp is increased. An increase of one order of magnitude, such as from 1,000 to 10,000, decreases the beamwidth by one order of magnitude, such as 12.50° to 2.5°. The operation can be performed in a single sweep for real-time operation by selecting the time-bandwidth product at the time of transmitting the chirp, and selecting the time delay and Doppler shift for the chirp replica to be used in the correlation. However, having selected the beamwidth (by preselecting the time-bandwidth products for successive sweeps), and having recorded the echo returns, it is possible to preferentially aim the beam at targets in different directions at different times by performing the correlation with a replica time delayed and Doppler shifted as required on different data processing passes.

The novel features that are considered characteristic of this invention are set forth with particularity in the appended claims. The invention will best be understood from the following description when read in connection with the accompanying drawings.

BRIEF DESCRIPTION OF DRAWINGS

FIG. 1 illustrates a submerged vessel towing a sonar transmitter/receiver.

FIG. 2 shows a typical sonar beam pattern.

FIG. 3 shows the beam of FIG. 2 made narrower and aimed in accordance with the present invention.

FIG. 4 illustrates in general terms the methodology of the present invention.

FIG. 5 graphically illustrates the technique for aiming a beam considering only Doppler shift.

FIG. 6 illustrates the technique of FIG. 5 considering both Doppler shift and time delay.

FIG. 7 illustrates a block diagram of a sonar system embodying the present invention.

FIG. 8 illustrates a block diagram of the system of FIG. 7 for off-line data processing.

BEST MODE FOR CARRYING OUT THE INVENTION

Referring now to the drawings, FIG. 1, illustrates a submerged vessel 10 towed by a ship 12. The vessel contains a sonar transmitter 14 and receiver 16. The transmitter/receiver apparatus is used in ocean floor surveying and subbottom mapping for scientific or commercial applications with the ship moving at about 7 knots. The sonar beam pattern is typically of the shape shown in FIG. 2, having a main lobe of 3 db half angle ϕ, and a plurality of side lobes.

Since the sonar vessel is moving through the water at a significant velocity, there is a substantial Doppler shift produced in the sonar echo signals received. A method and apparatus for using this Doppler effect to synthesize a narrow beam, or to aim the beam, can increase the effectiveness of the sonar surveying and mapping apparatus. FIG. 3 illustrates both aiming the beam in elevation and reducing the beamwidth angle θ. As will be appreciated from the discussion that follows, either or both effects of a Doppler formed beam can be achieved as a spatial directional pattern for preferential transmission of signals based upon relative motion between the transmitter (source of signals) and the receiver (detector of sonar echos) when the transmitter and receiver are collocated, and objects in the resulting beam are at a distance.

Briefly, objects are preferentially selected in azimuth and elevation from a multiplicity of echo returning objects at the same or nearby range during each successive sweep using a correlator at the receiver for the Doppler shifted return of the selected objects, and the narrow beam is formed by control of the time-bandwidth product, TB, at the transmitter. This technique is illustrated in general terms by FIG. 4 which indicates for the transmitter 14 the general function of transmitting a known frequency chirp, i.e., a chirp of known bandwidth, B, for a specified time, T, and for the receiver the general function of correlating a time delayed replica of the transmitted chirp with the Doppler shifted return from the target.

The correlator is controlled to select the Doppler shifted return from targets in a desired direction, thereby aiming the beam in azimuth and elevation. FIG. 5 illustrates the technique graphically. Consider a chirp having a frequency, f, that varies linearly with time, t. For a target at 90° from the direction of vessel motion, the frequency of the return signal will follow a line of the same slope as the transmitted signal, indicated in the graph of FIG. 5 as ZERO DOPPLER. For a target at less than 90° from the direction of motion, the frequency of the return signal will follow a line of greater slope labeled POS. DOPPLER, and for a target at greater than 90° from the direction of motion, the frequency of the return signal will follow a line of smaller slope labeled NEG. DOPPLER. In each case, the Doppler selected is determined by the correlator using known digital data processing techniques, either on a real time basis, or off-line.

The essential process by which the correlator functions consists of the steps of multiplication of a delayed reference with the received signal, over the entire chirp interval, followed by summation of the results of multiplication. In a particular embodiment used to verify the process of Doppler beam aiming, the received signal was multiplied by the chirp signal transmitted and delayed as well as Doppler shifted. Only those received signal components having the same Doppler frequency change as that of the delayed and Doppler shifted chirp produced a multiplication product that can be characterized as a steady tone. All other Doppler shifted signals produced products characterized by a varying tone. The particular received signal, whose path delay between the transmitter and receiver was equal to the delay of the Doppler shifted chirp, produced a zero frequency (DC) signal upon multiplication with the delayed and Doppler shifted chirp. Integration of the multiplier output enhances the amplitude of the zero frequency signal, and diminishes all other signals in a manner well understood as a matched filter process. The output of this summation (integration) is the target data that may be displayed.

There are a number of different ways to implement a correlator. The foregoing technique is only one presented here by way of example, and not by way of limitation.

The foregoing discussion is general, and is easily understood by assuming for simplicity aiming in elevation only in a vertical plane passing through the vector of vessel motion as shown in FIG. 3. However, it can be readily appreciated that aiming can be extended in azimuth as well since any change in azimuth will alter the Doppler shift established by a beam at the same elevation, but at zero azimuth. However, a problem does arise in the ambiguity between a target with a negative azimuth angle and a target with a positive azimuth angle, when both have the same elevation. This ambiguity can be resolved by using a sonar that is looking only to one side. Then aiming can be carried out in azimuth and elevation without ambiguity.

To look to only one side, the sonar apparatus is installed in the vessel with a transmitter/receiver canted to one side. Synthesizing a narrow beam by controlling the time-bandwidth product of the transmitter then assures that the look is to only one side. Alternatively, arrays of transmitting and receiving transducers may be employed for electronically canting the beam as in an electronically steered radar array.

Once the cant angle is set, aiming the beam is accomplished by operation of the correlator for the desired Doppler shift. Thus by properly using the correlator, the angular position of targets in azimuth and elevation are selected preferentially from a multiplicity of targets at the same nearby range. The operation can be performed for each single sweep, and by control of the transmitted time-bandwidth product of each chirped sweep, a narrower angular selection is made than would otherwise be possible, i.e., a narrower beam is formed.

To understand the technique employed to synthesize a narrow beam, consider that the beam angle (as measured from the beam axis to the first null point of the main lobe as shown in FIG. 3) is determined by the relationship Sine $\theta=[2\nu(v/c)TB]^{-1}$ where v is the relative velocity between vessel and target, and c is the speed of sound in water. The ratio of velocity to speed of sound in water is multiplied by two for the round trip to produce a value proportional to the Doppler frequency shift, and the product is multiplied by the time-bandwidth product TB. The sine of the angle θ is then equal to the reciprocal of that value. Thus it can be seen that by increasing the chirp period T, or the frequency bandwidth B of the chirp, or both, the angle θ is decreased. In order to generate the narrow beam angle θ, it is preferred that the original sonar angle ϕ be wider than θ. This is a condition contrary to the conventional sonar practice of using a narrow projected angle ϕ.

Now to better understand the technique of aiming the beam, consider FIG. 6, which shows the effect of simple time delay and of time delay plus Doppler shift. The presumed frequency chirp transmitted is indicated by the line a-b in the graph, and the return signal delayed by only the transit time to the target and back by a line c-d of the same slope.

The transit time delay plus Doppler shift of frequency produces a signal represented by the line e-f of a different slope. In this case, the Doppler shifted signal slope is increased, representing a target being approached by the vessel.

A correlator detects the Doppler shifted signal for the duration of the sweep interval, T. Thus, for aiming the beam, the correlator will use as a replica a presumed frequency-time slope corresponding to the Doppler direction in which maximum beam strength is desired.

If there is a frequency difference between the presumed slope and the Doppler shifted slope, the output of the correlator will diminish. In the simplest case, without time apodization, the output of the correlator will follow a (sin x)/x form, where x is a parameter related to the accumulated frequency difference for the duration of the received signal.

In a practical embodiment of this concept, a sonar projector is used with an essentially uniform sound pressure amplitude over a frequency range from 1.5 to 4.5 kHz. A transmitted chirp is used which has a slope of 10,000 Hz per second. This yields a time-bandwidth product of 900.

$$\frac{(4500 - 1500)\text{ Hz.}}{10{,}000\text{ Hz/sec}} (4500 - 1500)\text{ Hz} = 900\text{ Hz-sec}$$

A Doppler shift of three Hertz during the three kilohertz sweep will yield the first null of the (sin x)/x response. This Doppler offset requires a velocity of 0.7605 meters per second.

$$\frac{2v}{c} = \frac{2v}{1521\text{ m/sec}} = \frac{3\text{ Hz}}{3000\text{ Hz}}$$

$$v = 0.7605\text{ meters per second.}$$

Presuming a ship forward speed of 7 knots, or 3.6008 meters per second, a forward angle of 12.19 degrees produces the necessary 0.76 meter per second closing rate. As an example only, and not as a limitation, the transmitted chirp repeats every 4 seconds. At a speed of 7 knots the positional offset of each sonar chirp is 14.4 meters along track. The angle θ of the Doppler formed beam is approximately 10 degrees, which means that adjacent objects closer than 67 meters from the location of the transmitter/receiver will be resolved without range ambiguity.

Referring now to FIG. 7, a system for real time selection of targets by beam aiming is comprised of a transmitting transducer 14*a* driven by a frequency chirp generator 20 in the transmitter that is controlled to produce a frequency signal of selected bandwidth B for an interval T established by a timing generator 22. A variable delay 24 delays the interval T a specified time. The delayed interval T is then used to control a Doppler shifted chirp generator 26, which produces a replica of the chirp transmitted, but Doppler shifted, for use as a reference in a correlator 28.

Both chirp generators are linear sweep frequency generators with a bandwidth B either designed into the generators, or selectively set into the generators. However, the linear sweep of the chirp generator 26 is not identical to that of the chirp generator 20, except in bandwidth, since its slope and frequency offset is set by a Doppler shift control input for the particular pointing (aiming) of the beam required at the time of each sweep, or at least at the time of commencing a run of successive sweeps.

The output of the correlator is target data that may be displayed on a unit 30 as a function of range (or time). The X and Y drive for the display unit is derived from a scan generator 32. For example, assuming time is plotted on the X axis, the interval T is used to increment the starting point of each sweep on the X axis, and clock pulses occurring from one interval to the next are used to increment the Y axis in range. The delay control may be used to inhibit the Y deflection of the plot for the delay period indicated so as to plot only the range swath of interest. Since the sweeps may be in directions other than abeam, depending upon the Doppler shift control input, the slope of each sweep is modified accordingly. For example, one counter may be used to increase the X deflection one unit for each interval T to a new starting point for each sweep. and another counter is used to increase the Y deflection following the delay set in by the delay control. The outputs of the two counters are then converted from digital to analog, and multiplied by a scan converting factor introduced in the digital to analog converter. The factor is, of course, a function of the slope, and it increases from zero at the base of each sweep as a function of the Y counter. In that manner the X deflection is increased by a factor $\Delta(Y/S)$, where S is the slope determined by the Doppler shift control, and Y is the output of the Y deflection counter. For a positive Doppler shift control to aim the sonar beam at an angle less than 90° from the direction of motion, the sign of the slope S is positive, and for a negative Doppler shift control to aim the sonar beam at an angle greater than 90° from the direction of motion, the sign of the slope S is negative. Similarly, the Y deflection is increased by a factor $\Delta X \cdot S$, where the sign of the slope remains the same for both positive and negative Doppler shift.

An adaptation of the system of FIG. 7 for off-line data processing is shown in FIG. 8. The only essential difference is that a recorder 34, such as a tape recorder, is used to store: the signal from the receiving transducer 16*a*; the interval signal T; and the clock pulses, which may be accumulated and encoded to indicate real time relative to some starting point, with each code change recorded indicating a lapse of one unit of time, such as one second. The recorded data is then played back at a later time for processing as before, i.e., as in FIG. 7. An advantage of this arrangement is that once data is recorded, it can be replayed a number of times, each time aiming the beam differently through the Doppler shift control to look at different targets.

While the process of Doppler beam forming can utilize only a single transmitted chirp for each sweep, it is readily appreciated that combined processing of a successive multiplicity of chirps may be used to increase the time-bandwidth product, and thereby narrow the beam angle θ for targets present in successive chirps. The system of FIGS. 7 and 8 are intended to include this possibility, should the need present itself. Still other modifications and equivalents may readily occur to those skilled in the art. Consequently, it is intended that the claims be interpreted to cover such modifications and equivalents.

I claim:

1. In a process for transmitting a beam of frequency chirped wave energy, and for both receiving chirped wave energy from targets having relative motion and for correlating time delayed chirped wave energy with received chirped wave energy, a method of synthesizing a narrow beam from otherwise broadbeam wave energy by so controlling the frequency bandwidth, B, and the time, T, of transmitted chirps that for a narrower beamwidth under given conditions the time-bandwidth product, TB, is increased, and producing time delayed replicas of transmitted chirps for correlation with received chirps.

2. In a process as defined in claim 1, a method of preferentially selecting targets by so Doppler shifting said time delayed replica chirps as to approximately match Doppler shifted chirps returned by the targets.

3. In a process as defined in claim 2, the method of preferentially selecting targets, wherein Doppler shifted chirps correlated are transmitted chirps reflected by targets.

4. In a process as defined in claim 2, the method of preferentially selecting a target, wherein said target transponds by transmitting frequency chirped wave energy of controlled time-bandwidth product equal to that of said transmitted chirps, but with a distinct synthesized Doppler shift, and wherein Doppler shifted chirps correlated are chirps transponded by said target whereby said particular target may be preferentially selected by producing a reference signal for correlation with approximately the same distinct Doppler shift.

5. In a process as defined in claim 1, 2, 3 or 4, the method or methods described wherein said system is a sonar system, and said wave energy is comprised of sound waves transmitted through water.

6. Apparatus for transmitting to, and receiving from, targets having relative motion, a beam pattern of frequency chirped wave energy using relatively broadbeamwidth transmitting and receiving transducers, said apparatus comprising means responsive to control signals for controlling the frequency bandwidth, B, and the time interval, T, during which each chirp of wave energy is transmitted, thereby to synthesize a narrower beamwidth than is transmitted and received by so controlling the bandwidth and the time that under given conditions the time-bandwidth product, TB, in chirped wave energy transmitted is increased, means for producing a delayed and Doppler shifted replica of the transmitted chirped wave energy for use as a reference, means for correlating chirped wave energy received with said replica to produce target data, thereby synthesizing a narrower beamwidth wave energy pattern than the pattern of said transducers, and means for displaying target data.

7. Apparatus as defined in claim 6 wherein said means for producing said delayed and Doppler shifted replica includes a control for providing a selected Doppler shift thereby to preferentially select targets for display from a multiplicity of targets in different directions.

8. Apparatus as defined in claim 7 wherein said means for generating a delayed and Doppler shifted replica of the transmitted chirped energy wave is comprised of a chirp generator responsive to said control signals, and means for storing at least said time interval for use in said means for generating a delayed and Doppler shifted replica and storing received chirped wave energy for later correlation and display, said means for generating a delayed and Doppler shifted replica being responsive to control signals for controlling the frequency bandwidth and Doppler shift thereby to preferentially select a target for display at a later time.

9. Apparatus as defined in claim 6, 7 or 8 wherein said apparatus is a sonar system and said wave energy consists of sound waves transmitted through water.

10. Apparatus for locating targets with transmitted waves of energy using collocated broadbeam transmitting and receiving transducers, where said energy is transmitted as a linear frequency chirp of predetermined bandwidth, B, for a controlled interval, T, thereby to control the time-bandwidth product of transmitted energy for a beam of narrow width, comprising means for correlating waves of energy from targets with a time delayed replica of said energy transmitted, said received waves of energy also being linear frequency chirps of the same time-bandwidth product as transmitted chirps, and means for generating said replica with a predetermined Doppler shift, thereby to preferentially select a target for said beam of narrow width.

11. Apparatus as defined in claim 10 wherein said received energy is energy reflected by targets.

12. Apparatus as defined in claim 10 wherein said received energy is energy transmitted by a target.

* * * * *

U.S. Patent Sep. 1, 1981 Sheet 1 of 4 4,287,578

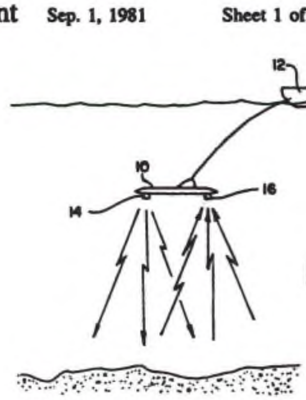


FIG. 1

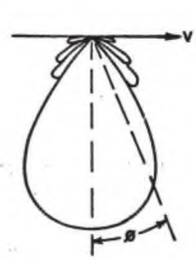


FIG. 2

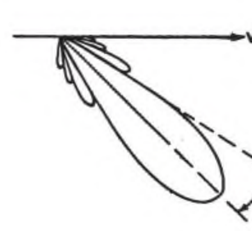


FIG. 3

U.S. Patent Sep. 1, 1981 Sheet 2 of 4 4,287,578

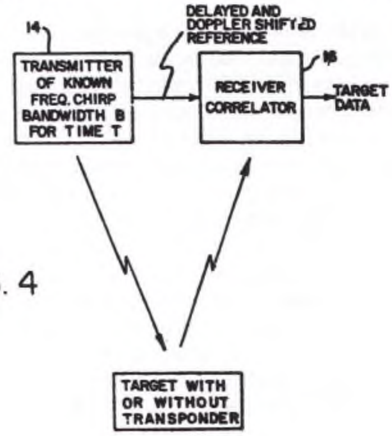


FIG. 4

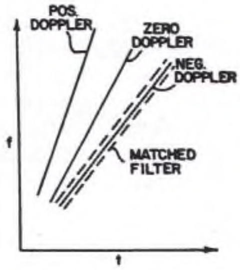


FIG. 5

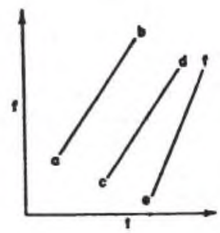


FIG. 6

U.S. Patent Sep. 1, 1981 Sheet 3 of 4 4,287,578

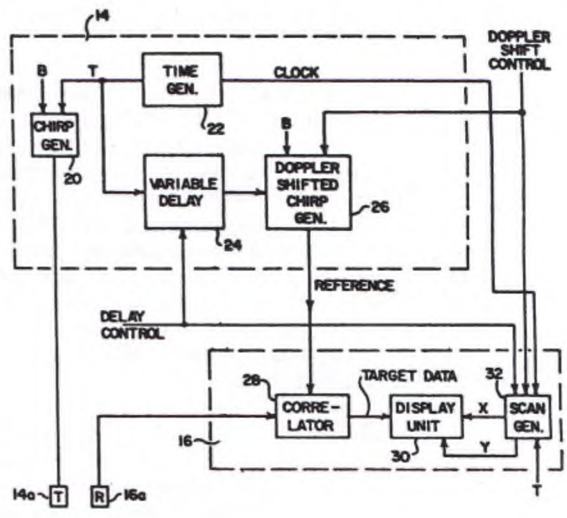


FIG. 7

U.S. Patent Sep. 1, 1981 Sheet 4 of 4 4,287,578

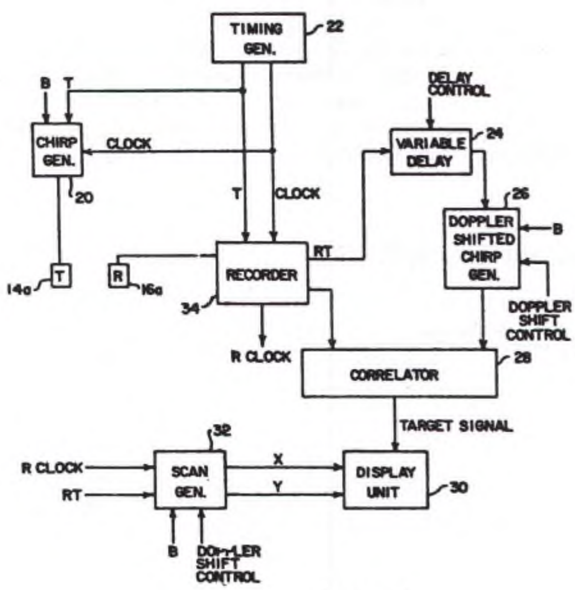


FIG. 8

bibliography

WORKS BY RICHARD C. HEYSER

TDS-Audio and Related Topics

"Acoustical Measurements by Time Delay Spectrometry," *J. Audio Eng. Soc.*, vol. 15, pp. 370–382 (1967 Oct.).

"Loudspeaker Phase Characteristics and Time Delay Distortion: Part 1," *J. Audio Eng. Soc.*, vol. 17, pp. 30–41 (1969 Jan.); reprinted in *Loudspeakers: An Anthology of Articles on Loudspeakers from the Journal of the Audio Engineering Society*, 1st ed., pp. 120–131 (1978); 2nd ed. (1980).

"Loudspeaker Phase Characteristics and Time Delay Distortion: Part 2," *J. Audio Eng. Soc.*, vol. 17, pp. 130–137 (1969 Apr.); reprinted in *Loudspeakers: An Anthology of Articles on Loudspeakers from the Journal of the Audio Engineering Society*, 1st ed., pp. 132–139 (1978); 2nd ed. (1980).

"Determination of Loudspeaker Signal Arrival Times: Part I," *J. Audio Eng. Soc.*, vol. 19, pp. 734–743 (1971 Oct.); reprinted in *Loudspeakers: An Anthology of Articles on Loudspeakers from the Journal of the Audio Engineering Society*, 1st ed., pp. 225–234 (1978); 2nd ed. (1980).

"Determination of Loudspeaker Signal Arrival Times: Part II," *J. Audio Eng. Soc.*, vol. 19, pp. 829–834 (1971 Nov.); reprinted in *Loudspeakers: An Anthology of Articles on Loudspeakers from the Journal of the Audio Engineering Society*, 1st ed., pp. 235–240 (1978); 2nd ed. (1980).

"Determination of Loudspeaker Signal Arrival Times: Part III," *J. Audio Eng. Soc.*, vol. 19, pp. 902–905 (1971 Dec.); reprinted in *Loudspeakers: An Anthology of Articles on Loudspeakers from the Journal of the Audio Engineering Society*, 1st ed., pp. 240–244 (1978); 2nd ed. (1980).

"The Delay Plane, Objective Analysis of Subjective Properties: Part I," *J. Audio Eng. Soc.*, vol. 21, pp. 690–701 (1973 Nov.).

"The Delay Plane, Objective Analysis of Subjective Properties: Part II," *J. Audio Eng. Soc.*, vol. 21, pp. 786–791 (1973 Dec.).

Author's Reply to M. A. Gerzon, "Comments on 'The Delay Plane, Objective Analysis of Subjective Properties,' " *J. Audio Eng. Soc.*, vol. 22, pp. 104–106 (1974 Mar.).

"Geometrical Considerations of Subjective Audio," *J. Audio Eng. Soc.*, vol. 22, pp. 674–682 (1974 Nov.).

"Some New Audio Measurements," presented at the 51st Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts)*, vol. 23, p. 498 (1975 July/Aug.), preprint 1008.

"Geometry of Sound Perception," presented at the 51st Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts)*, vol. 23, p. 494 (1975 July/Aug.), preprint 1009.

"Some Useful Graphical Relationships," *J. Audio Eng. Soc. (Project Notes/Engineering Briefs)*, vol. 23, pp. 562–565 (1975 Sept.).

Reply to J. Moir, "On Differential Time Delay," *J. Audio Eng. Soc. (Forum)*, vol. 24, p. 752 (1976 Nov.).

"Concepts in the Frequency and Time Domain Response of Loudspeakers," *Monitor—Proc. IREE* (Australia), vol. 37, pp. 67–76 (1976 Mar.).

"Perspective in Audio Analysis: Changing the Frame of Reference—Part I," *J. Audio Eng. Soc. (Communications)*, vol. 24, pp. 660–667 (1976 Oct.).

"Perspective in Audio Analysis: Changing the Frame of Reference—Part II," *J. Audio Eng. Soc. (Communications)*, vol. 24, pp. 742–751 (1976 Nov.).

"Perception and Geometry," *Audio*, vol. 61, pp. 52–60 (1977 June).

"Time and Frequency in Loudspeaker Measurements," *Audio*, vol. 61, pp. 80–83 (1977 July).

"Alternatives," *Audio*, vol. 62, pp. 50–52 (1978 Feb.).

"Fuzzy Alternatives," *J. Audio Eng. Soc.*, vol. 26, pp. 151–153 (1978 Mar.).

"Hearing vs. Measurement," *Audio*, vol. 62, pp. 46–50 (1978 Mar.).

"Imprecise Descriptors," presented at the 60th Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts)*, vol. 26, p. 582 (1978 July/Aug.), preprint 1350.

"A View through Different Windows," *Audio*, vol. 63, pp. 56–66 (1979 Feb.).

"Catastrophe Theory and Its Effect on Audio: Part I," *Audio*, vol. 63, pp. 52–62 (1979 Mar.).

"Catastrophe Theory and Its Effect on Audio: Part II," *Audio*, vol. 63, pp. 36–44 (1979 Apr.).

"Catastrophe Theory and Its Effect on Audio: Part III," *Audio*, vol. 63, pp. 42–55 (1979 May).

"Determining the Acoustic Position for Proper Phase Response of Transducers," *J. Audio Eng. Soc.*, vol. 32, pp. 23–25 (1984 Jan./Feb.).

Author's Reply to S. P. Lipshitz and J. Vanderkooy, "Comments on 'Determining the Acoustic Position for Proper Phase Response of Transducers,'" *J. Audio Eng. Soc. (Letters to the Editor)*, vol. 33, pp. 463–466 (1985 June).

"Instantaneous Intensity," presented at the 81st Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts)*, vol. 34, p. 1038 (1986 Dec.), preprint 2399.

"Specific Acoustic Admittance," presented at the 81st Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts)*, vol. 34, p. 1038 (1986 Dec.), preprint 2400.

"Fundamental Principles and Some Applications of Time Delay Spectrometry" in *Time Delay Spectrometry* (Audio Engineering Society, New York, 1988), pp. 205–262.

TDS-Audio Patents

"Time Delay Spectrometer," U.S. patent 3,466,652 (1969 Sept. 9).

"Method and Apparatus for Delay Analysis of Energy Transmitted through a Medium," U.S. patent 4,287,578 (1981 July 14).

General Patents

"Self-Biasing Transistor Amplifier," British patent 910079 (1961 July 27).

"Self-Biasing Transistor Amplifier," U.S. patent 3,035,233 (1962 May 15).

"Transistor Amplifier," U.S. patent 3,050,688 (1962 Aug. 21).

"Balanced Transistor Power Amplifier Connected in a Bridge Configuration," U.S. patent 3,124,757 (1964 Mar. 10).

"Temperature Control System with Pulse-Width Modulated Bridge," U.S. patent 3,733,463 (1973 May 15).

"Medical Tomograph System Using Ultrasonic Transmission," U.S. patent 4,083,232 (1978 Apr. 11).

"Method for Shaping and Aiming Narrow Beams," U.S. patent 4,279,019 (1981 Sept. 1).

TDS-Medical/Ultrasound Topics

R. C. Heyser, L. N. Besseling, and D. J. Le Croissette, "Transmission Ultrasound Soft Tissue Visualization," *Proc. 3rd Int. Conf. Med. Phys.* (1972), p. 35.5.

R. C. Heyser, L. N. Besseling, and D. H. Le Croissette, "Transmission Ultrasonograms Using Time Delay Spectrometry," *Proc. 25th Conf. Eng. in Biol. and Med.* (1972), vol. 14, p. 138.

R. C. Heyser and D. H. Le Croissette, "Transmission Ultrasonography," *Proc. 1973 Ultrasonics Symp. (IEEE 73 CHO 807-8SU)*, p. 709.

R. C. Heyser and D. H. Le Croissette, "A New Ultrasonic Imaging System Using Time Delay Spectrometry," *Ultrasound in Med. and Biol.*, vol. 1, pp. 119–131 (1974).

L. S. Fishman, R. C. Heyser, and D. H. Le Croissette, "Ultrasonic Transmission Measurements on Human Brain Sections," *Radiology*, vol. 112, pp. 211–213 (1974 July).

R. C. Heyser, "Ultrasonic Imaging, Intrauterine Fetal Visualization: A Multidisciplinary Approach," reprint Excerpta Medica International Cong. Ser. 371 (Excerpta Medica, Amsterdam, 1976), pp. 226–231.

D. H. Le Croissette and R. C. Heyser, "Attenuation and Velocity Measurements in Tissue Using Time Delay Spectrometry," *Proc. Conf. Ultrasonic Tissue Characterization* (1975), National Bureau of Standards, Spec. Pub. 453, pp. 81–95 (1976 Oct.).

P. M. Gammell, J. A. Roseboro, R. C. Heyser, D. H. Le Croissette, and R. L. Wilson, "Swept Frequency Ultrasonic Measurements of Tissue Characteristics," *Proc. 30th Ann. Conf. Eng. in Med. and Biol.*, vol. 19 (Alliance for Engineering in Medicine and Biology, Bethesda, MD, 1977), p. 229.

D. H. Le Croissette, R. C. Heyser, and R. L. Wilson, "The Characterization of Soft Tissue Using Ultrasonic Energy," *Proc. Am. Assoc. of Med. Instrum., 12th Ann. Meet.* (1977), p. 241.

R. C. Heyser, D. H. Le Croissette, R. Nathan, and R. L. Wilson, "Medical Ultrasonic Tomographic System," JPL Pub. 77-72 (1977 Oct.).

R. L. Wilson, D. H. Le Croissette, P. M. Gammell, J. A. Roseboro, and R. C. Heyser, "Measurement of Attenuation in Tissues Using Time Delay Spectrometry," presented at the 22nd Ann. Meet. of the American Institute of Ultrasound in Medicine (Dallas, TX, 1977 Oct./Nov.), in Denis White, Ed., *Proc. Ultrasound in Medicine*, vol. 4.

D. H. Le Croissette, R. C. Heyser, P. M. Gammell, and R. L. Wilson, "Tissue Identification by Ultrasound," JPL Pub. 78-90 (1978 Oct.).

D. H. Le Croissette, R. C. Heyser, P. M. Gammell, J. A. Roseboro, and R. L. Wilson, "The Attenuation of Selected Soft Tissue as a Function of Frequency," *Proc. 2nd Int. Symp. Ultrasonic Tissue Characterization* (1979), National Bureau of Standards, Spec. Pub. 525, pp. 101–108.

R. C. Heyser, "Detection of Thrombus on Heart Valve Cages by Reverberation," presented at the 25th Ann. Conv. of the American Institute of Ultrasound in Medicine (New Orleans, LA, 1980 Sept.), in Denis White, Ed., *Ultrasound in Medicine* (Pergamon, Oxford, 1981).

J. A. Rooney, R. C. Heyser, D. H. Le Croissette, J. D. Hestenes, and P. M. Gammell, "Imaging of Deep Arteries Using Time Delay Spectrometry," *J. Ultrasound in Med.*, vol. 1, no. 7, p. 84 (1982).

D. H. Le Croissette, R. C. Heyser, J. D. Hestenes, P. M. Gammell, and J. A. Rooney, "A Swept-Frequency Medical Ultrasonic Imaging System," *Proc. 35th Ann. Conf. Eng. in Med. and Biol.*, vol. 24 (Alliance for Engineering in Medicine and Biology, Bethesda, MD, 1982), p. 202.

J. A. Rooney and R. C. Heyser, "Use of Swept-Frequency Ultrasonic Techniques for Quantitative Bubble Detection," *Proc. Aerospace Med. Assoc.*, p. 120 (1983).

R. C. Heyser, J. D. Hestenes, J. A. Rooney, P. M. Gammell, and D. H. Le Croissette, "Medical Ultrasound Imager Based on Time Delay Spectrometry," *Ultrasonics* (In Press).

General Articles

"A Signal Biasing Output Transformerless Transistor Power Amplifier," *J. Audio Eng. Soc.*, vol. 8, pp. 185–191 (1960 July).

"Breakthrough in Speaker Testing," *Audio*, vol. 57, pp. 20–30 (1973 Nov).

"A Rosetta Stone for Audio?" *J. Audio Eng. Soc. (Editorial)*, vol. 22, p. 251 (1974 May).

"Speaker Tests: Impedance," *Audio*, vol. 58, p. 33 (1974 Sept.).

"Speaker Tests: Anechoic Frequency Response," *Audio*, vol. 58, pp. 44–48 (1974 Nov.).

"Speaker Tests: Phase Response," *Audio*, vol. 58, pp. 71–72, 82 (1974 Dec.).

"Speaker Tests: Room Test," *Audio*, vol. 59, pp. 66–68 (1975 Jan.).

"Speaker Tests: Frequency Response—Phase II (Addendum)," *Audio*, vol. 59, p. 31 (1975 May).

"Speaker Tests: Polar Response," *Audio*, vol. 59, pp. 30–31 (1975 May).

"Earthquake Prediction by Acoustic Methods," *J. Audio Eng. Soc. (Forum)*, vol. 23, pp. 469–470 (1975 July/Aug.).

"More on 'Recording Concert Hall Acoustics for Posterity,'" *J. Audio Eng. Soc. (Forum)*, vol. 23, pp. 571–572 (1975 Sept.).

"The Sound of One Clap Handing," *J. Audio Eng. Soc. (Forum)*, vol. 23, p. 644 (1975 Oct.).

"Speaker Tests: Harmonic Distortion," *Audio*, vol. 60, pp. 16–17 (1976 Feb.).

"IM Distortion in Speaker Systems," *Audio*, vol. 60, pp. 38–40 (1976 Mar.).

"Holomorph Recording," presented at the 54th Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts)*, vol. 24, p. 491 (1976 July/Aug.), preprint 1115.

"Audio Crescendo Test (Crescendo Test)," *Audio*, vol. 60, pp. 30–32 (1976 May).

"Energy-Time Test," *Audio*, vol. 60, pp. 74–80 (1976 June).

"Acoustic Rosetta Stone," *Audio*, vol. 63, pp. 66–68 (1979 Jan.).

"Polarity Convention (in 'The Forum')," *Audio*, vol. 63, pp. 18–20 (1979 Sept.).

"The Concept of Distortion," *Syn-Aud-Con Newsl.*, vol. 7, pp. 8–9 (1980 Summer).

"Speaker Impedance: More Complex than One Number," *Audio*, vol. 68, pp. 42–68 (1984 June).

Equipment Profiles

"Equipment Profile—RTR 280 DR Speaker System," *Audio*, vol. 57, pp. 76–79 (1973 Nov.).

"Equipment Profile—ESS amt-1 Speaker System," *Audio*, vol. 57, pp. 61–63, 72–75 (1973 Dec.).

"Equipment Profile—Audioanalyst Model A-200," *Audio*, vol. 58, pp. 55–57 (1974 Jan.).

"Equipment Profile—Altec Concept EQ-5," *Audio*, vol. 58, pp. 64–67 (1974 Feb.).

"Equipment Profile—Frazier Concerto Model F-10W-37," *Audio*, vol. 58, pp. 70–75 (1974 Mar.).

"Equipment Profile—Sherwood Evolution One Speaker System," *Audio*, vol. 58, pp. 58–64 (1974 Apr.).

"Equipment Profile—Magnavox Max 12," *Audio*, vol. 58, pp. 40–44 (1974 May).

"Equipment Profile—Applied Physics Laboratory APL-101 Speaker System," *Audio*, vol. 58, pp. 65–67 (1974 June).

"Equipment Profile—Frankmann Integrated Stereo Speaker System," *Audio*, vol. 58, pp. 46–51 (1974 July).

"Equipment Profile—Fairfax Model FX-350 Speaker System," *Audio*, vol. 58, pp. 68–74 (1974 Nov.).

"Equipment Profile—GTE Sylvania Model AS210A Speaker System," *Audio*, vol. 59, pp. 63–67 (1975 Aug.).

"Equipment Profile—Electro-Voice Interface: A Loudspeaker System," *Audio*, vol. 60, pp. 58–62 (1976 Mar.).

"Equipment Profile—Allison Acoustics Allison: One Speaker System," *Audio*, vol. 60, pp. 70–73 (1976 Apr.).

"Equipment Profile—Duntech Labs DL-15 Loudspeaker System," *Audio*, vol. 60, pp. 60–64 (1976 Aug.).

"Equipment Profile—Audioanalyst Model A-100X Loudspeaker System," *Audio*, vol. 60, pp. 62–65 (1976 Sept.).

"Equipment Profile—Dynaco A-25XL Loudspeaker," *Audio*, vol. 60, pp. 84–88 (1976 Nov.).

"Equipment Profile—Celestion UL-10 Loudspeaker," *Audio*, vol. 61, pp. 46, 50, 54, 73 (1977 Jan).

"Equipment Profile—Hartley Model Zodiac '76 Speaker System," *Audio*, vol. 61, pp. 75–80 (1977 June).

"Equipment Profile—Jensen Model 530 Three-Way Loudspeaker System," *Audio*, vol. 61, pp. 84–88 (1977 July).

"Equipment Profile—Tannoy Model Berkeley Speaker System," *Audio*, vol. 61, pp. 88–96 (1977 Dec.).

"Equipment Profile—B & W DM-6 Speaker System," *Audio*, vol. 62, pp. 74–80 (1978 Feb.).

"Equipment Profile—Bang & Olufsen Beovox Model M-70 Speaker System," *Audio*, vol. 62, pp. 64–68 (1978 May).

"Equipment Profile—Janis Audio Model W-1 Subwoofer," *Audio*, vol. 62, pp. 94–100 (1978 June).

"Equipment Profile—Yamaha NS-1000 Loudspeaker System," *Audio*, vol. 63, pp. 82–90 (1979 Jan.).

"Equipment Profile—Acoustic Research Model AR90 Loudspeaker System," *Audio*, vol. 63, pp. 88–94 (1979 Nov.).

"Equipment Profile—KEF 104 aB Loudspeaker System," *Audio*, vol. 63, pp. 88–96 (1979 Dec.).

"Equipment Profile—Fried Products Model Q Loudspeaker System," *Audio*, vol. 64, pp. 93–98 (1980 June).

"Equipment Profile—Fulton Nuance Loudspeaker," *Audio*, vol. 64, pp. 48–54 (1980 Dec.).

"Equipment Profile—Mordaunt-Short Signifier Loudspeaker," *Audio*, vol. 65, pp. 56–59 (1981 Feb.).

"Equipment Profile—Wharfedale E-90 Loudspeaker System," *Audio*, vol. 65, pp. 68–71 (1981 Sept.).

"Equipment Profile—KEF 105 Series II Loudspeaker," *Audio*, vol. 66, pp. 56–62 (1982 Feb.).

"Equipment Profile—M & K Satellite-1A and Volkswoofer-A Loudspeakers," *Audio*, vol. 66, pp. 63–68 (1982 Apr.).

"Equipment Profile—Black Acoustics Night Speaker," *Audio*, vol. 66, pp. 52–56 (1982 Aug.).

"Equipment Profile—Cambridge Physics Model 310 Loudspeaker," *Audio*, vol. 66, pp. 68–72 (1982 Nov.).

"Equipment Profile—Koss CM/1030 Loudspeaker," *Audio*, vol. 67, pp. 74–79 (1983 May).

"Equipment Profile—Boston Acoustics A-40 Speaker," *Audio*, vol. 67, pp. 78–82 (1983 July).

"Equipment Profile—Genesis Model 210 Speaker," *Audio*, vol. 67, pp. 67–71 (1983 Sept.).

"Equipment Profile—ESB 7/06 Speaker," *Audio*, vol. 67, pp. 72–78 (1983 Nov.).

"Equipment Profile—Acoustic Research AR98LS Speaker," *Audio*, vol. 69, pp. 94–99 (1985 Jan.).

"Equipment Profile—Quad ESL 63 Loudspeaker," *Audio*, vol. 69, pp. 110–122 (1985 June).

"Equipment Profile—Thiel CS3 Speaker," *Audio*, vol. 69, pp. 60–68 (1985 Nov.).

"Equipment Profile—Klipschorn Loudspeaker," *Audio*, vol. 70, pp. 64–72 (1986 Jan.).

RELATED PUBLICATIONS BY OTHER AUTHORS

G. W. Stewart, "Problems Suggested by an Uncertainty Principle in Acoustics," *J. Acoust. Soc. Am.*, vol. 2, p. 235 (1931).

Y. W. Lee, "Synthesis of Electrical Networks by Means of the Fourier Transforms of Laguerre's Functions," *J. Math. and Phys.* (M.I.T.), vol. 11, pp. 83–113 (1932).

D. Gabor, "Theory of Communication," *Proc. IEE*, vol. 93, p. 425 (1946).

E. C. Titchmarsh, *Introduction to the Theory of Fourier Integrals* (Oxford University Press, London, 1948).

E. C. Titchmarsh, *Fourier Integrals* (Claredon Press, Oxford, 1948), p. 128.

N. Wiener, *Extrapolation, Interpolation, and Smoothing of Stationary Time Series* (M.I.T. Press, Cambridge, MA, 1949).

N. Wiener, *The Fourier Integral and Certain of Its Applications* (Dover, New York, 1958).

G. F. Simmons, *Topology and Modern Analysis* (McGraw-Hill, New York, 1963), pp. 16–17.

I. M. Gel'fand and G. E. Shilov, *Generalized Functions*, vol. 1, *Properties and Operations* (Academic Press, New York, 1964).

R. R. Goldberg, *Fourier Transforms* (Cambridge University Press, Cambridge, 1965).

I. M. Gel'fand and G. E. Shilov, *Generalized Function*, vol. 2, *Spaces of Fundamental Generalized Functions* (Academic Press, New York, 1968).

W. F. Donoghue, Jr., *Distributions and Fourier Transforms* (Academic Press, New York, 1969).

T. Kuhn, "The Structure of Scientific Revolutions." *Int. Encyclopedia of Unified Science*, vol. 2, no. 2 (University of Chicago Press, Chicago, IL, 1970).

P. Butzer and R. Nessel, *Fourier Analysis and Approximation* (Academic Press, New York, 1971), pp. 310–315.

E. L. Hamilton, "The Elastic Properties of Marine Sediments," *J. Geophys. Res.*, vol. 76, p. 579 (1971).

E. Stein and G. Weiss, *Introduction to Fourier Analysis on Euclidean Spaces* (Princeton University Press, Princeton, NJ, 1971), pp. 186–188.

R. G. Douglas, *Banach Algebra Techniques in Operator Theory* (Academic Press, New York, 1972).

E. L. Hamilton, "Compression-Wave Attenuation in Marine Sediments," *Geophysics*, vol. 37, p. 620 (1972).

M. Jammer, *The Philosophy of Quantum Mechanics* (Wiley, New York, 1974), pp. 2–9.

M. A. Gerzon, "Comments on 'The Delay Plane, Objective Analysis of Subjective Properties: Part I,' " *J. Audio Eng. Soc.*, vol. 22, pp. 104–106 (1974 Mar.).

D. Davis, "A Rosetta Tone?" *J. Audio Eng. Soc. (Forum)*, vol. 22, p. 430 (1974 July/Aug.).

M. A. Gerzon, "On a 'Rosetta Stone' for Acoustic Recording," *J. Audio Eng. Soc. (Forum)*, vol. 22, p. 638 (1974 Oct.).

A. Prosperetti, "Nonlinear Oscillations of Gas Bubbles in Liquids: Steady State Solutions," *J. Acoust. Soc. Am.*, vol. 56, p. 878 (1974).

M. A. Gerzon, "Recording Concert Hall Acoustics for Posterity," *J. Audio Eng. Soc. (Forum)*, vol. 23, pp. 569–571 (1975 Sept.).

J. Moir, "Differential Time Delay," *J. Audio Eng. Soc. (Forum)*, vol. 24, p. 752 (1976 Nov.).

C. S. Clay and H. Medwin, *Acoustic Oceanography: Principles and Applications* (John Wiley and Sons, New York, 1977), pp. 194–203.

H. Biering and O. Z. Pederson, "System Analysis and Time Delay Spectrometry, Parts I and II," *Bruel and Kjaer Tech. Rev.*, no. 1 and 2 (1983).

D. L. Miller and W. L. Nyborg, "Theoretical Investigation of the Response of Gas-Filled Micropores and Cavitation Nuclei to Ultrasound," *J. Acoust. Soc. Am.*, vol. 73, p. 1537 (1983).

S. P. Lipshitz and J. Vanderkooy, "Comments of 'Determining the Acoustic Position for Proper Phase Response of Transducers,' " *J. Audio Eng. Soc. (Letters to the Editor)*, vol. 33, pp. 463–465 (1985 June).