Nonlinear Filtering of Multiplied and Convolved Signals

ALAN V. OPPENHEIM, MEMBER, IEEE, RONALD W. SCHAFER, MEMBER, IEEE, AND THOMAS G. STOCKHAM, JR., MEMBER, IEEE

Invited Paper

Abstract—**An approach to some nonlinear filtering problems through a generalized notion of superposition has proven useful. In this paper this approach is investigated for the nonlinear filtering of signals which can be expressed as products or as convolutions of components. The applications of this approach in audio dynamic range compression and expansion, image enhancement with applications to bandwidth reduction, echo removal, and speech waveform processing are presented.**

I. Introduction

IN THIS PAPER, a class of nonlinear filters is discussed. This class is based on an approach to the problem of synthesizing nonlinear systems from the same point of view as that used for linear system design and analysis. Specifically, there are many classes of nonlinear systems which obey a principle of superposition. This property can be exploited in much the same way as it is in characterizing linear systems.

The general theoretical structure for characterizing nonlinear systems in this way has been formulated and studied in detail by Oppenheim [1]–[3]. While the framework which this structure provides is quite broad, it has so far been pursued in depth for two specific cases: the synthesis of nonlinear filters for signals which can be expressed as a product of components and the synthesis of nonlinear filters for signals which can be expressed as a convolution of components.

The first part of the paper is directed toward a brief explanation of the notion of superposition as it applies to problems in nonlinear filtering. This explanation is followed by a detailed discussion of the analytical framework for the specific cases of the filtering of multiplied signals and the filtering of convolved signals. Following this analysis, the discussion is directed toward the applications of the theory which have thus far been pursued.

Four applications are presented, two involving multiplicative filtering and two involving convolutional filtering or deconvolution. The multiplicative applications, as developed by Stockham, involve audio dynamic range compression and expansion and image enhancement with applications to bandwidth reduction. The deconvolution examples involve echo removal and speech analysis as pursued by Schafer and Oppenheim, respectively. All four applications have progressed to the point where working models have been realized through computer simulation, and one to the point where specially designed hardware has been installed as part of an unrelated system.

The work was originally inspired to a large extent by the ideas and attitudes of Dr. M. V. Cerillo, and many readers will undoubtedly recognize the flavor of his thinking in some of the applications presented in the following.

II. Generalized Linear Filtering

When considering the problem of filtering signals that have been added, we often focus our attention on the use of a linear system. While this constraint does not always lead to a "best" choice for the filter, it has the advantage of analytical convenience. This analytical convenience is almost a direct result of the principle of superposition that linear systems satisfy. In contrast, when determining a filtering procedure to separate signals that have been nonadditively combined, such as through multiplication or convolution, it is usually more difficult, and in many cases less meaningful to use a linear system. However, we can imagine generalizing the notion of linear filtering in such a way that it encompasses this broader class of problems. Specifically, let us consider two signals $s_1(t)$ and $s_2(t)$ that have been combined according to some rule which we denote by $\circ$, so that the resulting signal $s(t)$ to be processed can be expressed as

$$s(t) = s_1(t) \circ s_2(t).$$

Let ϕ represent the transformation for the filter. Then in generalizing the notion of linear filtering, we require that ϕ have the property that

$$\phi[s_1(t) \circ s_2(t)] = \phi[s_1(t)] \circ \phi[s_2(t)]. \quad (1)$$

Manuscript received April 5, 1968; revised June 5, 1968. *This invited paper is one of a series planned on topics of general interest.—The Editor.*

A. V. Oppenheim is with the Department of Electrical Engineering and the Research Laboratory of Electronics (supported in part by the Joint Services Electronics Program [Contract DA 28-043-AMC-02536(E)]), Massachusetts Institute of Technology, Cambridge, Mass. He is now on a leave of absence at M.I.T. Lincoln Laboratory (operated with support from the U. S. Air Force and the U. S. Advanced Research Project Agency) where a portion of the work was performed.

R. W. Schafer was formerly with the M.I.T. Research Laboratory of Electronics. A portion of the work reported here was submitted to the M.I.T. Department of Electrical Engineering in partial fulfillment of the requirements for the Ph.D. degree. He is now with the Bell Telephone Laboratories, Inc., Murray Hill, N. J.

T. G. Stockham, Jr., was formerly with the M.I.T. Lincoln Laboratory. He is now with the Department of Computer Sciences, University of Utah, Salt Lake City, Utah.

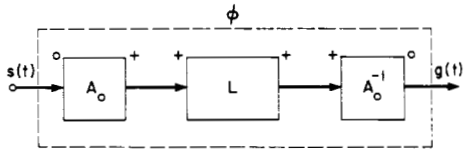


Fig. 1. The canonic representation for a homomorphic filter.

The formalism for representing systems having this property lies in interpreting the system inputs as vectors in a vector space with the rule ○ corresponding to vector addition, and the system transformation ϕ as an algebraically linear transformation on that space [1]. We must therefore restrict the operation ○ so that it satisfies the algebraic postulates of vector addition and associate with the set of inputs a rule for combining inputs with scalars, which we will call scalar multiplication and denote by :. To generalize the notion of linear filtering, then, we require that the class of systems, in addition to satisfying (1), also have the property that

$$\phi[c:s(t)] = c:\phi[s(t)]. \tag{2}$$

When the rule ○ corresponds to addition of the functions and the rule : corresponds to the product of the input with the scalar, then (1) and (2) reduce to the principle of superposition as it applies to linear systems. Systems in the class satisfying (1) and (2) have been referred to as homomorphic systems, emphasizing their interpretation as algebraically linear transformations between vector spaces.

The primary advantage in the restriction of the class of filters through (1) and (2) lies in the canonic representation for systems having this property. It has been shown [1] that if the system inputs constitute a vector space with the operations ○ and : corresponding to vector addition and scalar multiplication, then ϕ is representable as a cascade of three systems as shown in Fig. 1. The first system, $A_\circ$ in this representation, has the property that

$$A_\circ[s_1(t) \circ s_2(t)] = A_\circ[s_1(t)] + A_\circ[s_2(t)] \tag{3}$$

and

$$A_\circ[c:s(t)] = cA_\circ[s(t)]. \tag{4}$$

Furthermore, $A_\circ$ is characteristic of the class in the sense that it depends only on the operations ○ and : and not on the details of the system ϕ, and consequently is conveniently referred to as the *characteristic system*. The system L is a linear system and the system $A_\circ^{-1}$ is the inverse of the system $A_\circ$, i.e.,

$$A_\circ^{-1}\{A_\circ[s(t)]\} = s(t). \tag{5}$$

On the basis of this canonic representation we observe that generalized linear filtering corresponds to transforming the original problem to one in which the components are added, and after linear filtering, transforming the result back to the original space of inputs. Thus, once the characteristic system for the class has been determined, the problem reduces to a linear filtering problem.

III. Homomorphic Filtering of Multiplied Signals

One of the simplest examples of a rule of superposition satisfying the conditions above is that of ordinary multiplication. Of further interest is the fact that there exist several practical situations in which it is especially convenient to consider waveforms as products rather than as sums. Examples include problems involving fading channels, amplitude modulation, automatic gain control, audio dynamic range compression or expansion, and image processing. In these situations it is common to find two signals, one varying slowly and the other rapidly, combined as a product. In addition, it is frequently desirable to modify one signal and not the other or to process each according to separate objectives.

The product rule satisfies the algebraic postulates of vector addition. The companion rule for scalar multiplication is that of taking a signal to a scalar power. In terms of the symbols used earlier we have[1]

$$s_1(t) \circ s_2(t) = s_1(t) \cdot s_2(t)$$

and

$$c:s(t) = [s(t)]^c.$$

Equations (1) and (2) then become

$$\phi[s_1(t) \cdot s_2(t)] = \phi[s_1(t)] \cdot \phi[s_2(t)] \tag{6}$$

and

$$\phi\{[s(t)]^c\} = \{\phi[s(t)]\}^c. \tag{7}$$

Following the pattern of Fig. 1 we may construct Fig. 2 and, in analogy with (3), (4), and (5), we require that P have the property that

$$P[s_1(t) \cdot s_2(t)] = P[s_1(t)] + P[s_2(t)] \tag{8}$$

$$P[\{s(t)\}^c] = cP[s(t)] \tag{9}$$

and

$$P^{-1}\{P[s(t)]\} = s(t). \tag{10}$$

If we limit our consideration to include only positive real signals $s(t)$, and therefore real scalars c, the characteristic system P may be chosen as the ordinary logarithm function. It follows that P^{-1} is the corresponding exponential function. With this information this class of homomorphic systems can be represented more explicitly as in Fig. 3. An example in which we encounter only positive real signals is to be found in image processing in which the signals are formed of incoherent light. The physics of the situation guarantees the absence of negative or nonreal light intensities, while practical considerations almost certainly preclude zero intensity.

In the event that the signals to be processed cannot be restricted as above we may consider complex signals $s(t)$

[1] In these arguments we will assume that the signals involved are functions of time t. However, it is important for the reader to realize that there is nothing in the arguments to be presented which prevents the consideration of signals which are functions of space, frequency, or any other parameter. Neither is there any restriction to the consideration of one-dimensional signals.

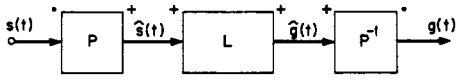


Fig. 2. The canonic representation for a multiplicative filter.

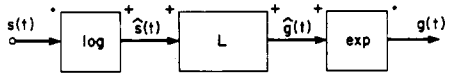


Fig. 3. The filter of Fig. 2 with P and P^{-1} specified as the logarithm and exponential transformations, respectively.

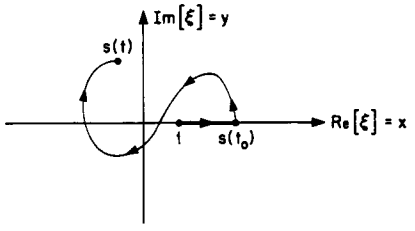


Fig. 4. A typical path of integration for defining the complex logarithm.

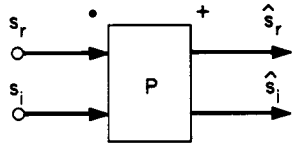


Fig. 5. The system P of Fig. 2 represented for complex signals.

and either real or complex scalars c. If we attempt to employ the complex logarithm function as the characteristic system in this situation, we encounter the immediate dilemma that the output $\hat{s}(t)$ of that system is not unique. The standard artifice of invoking the principal value of the complex logarithm cannot be used in this case, because the principal value of the logarithm of a product of complex signals is not always the sum of the principal values corresponding to the individual complex signals, violating (8).

There are restrictions which can be placed upon complex input signals such that a satisfactory characteristic system P closely related to the complex logarithm can be found. These restrictions require that complex inputs be continuous nonzero functions which attain a positive real value $s(t_0)$ at some prescribed instant of time t_0. In the case that complex scalars c are to be considered, $s(t_0)$ must be unity. The operation P is then taken to be

$$P[s(t)] = \int_1^{s(t)} \frac{d\xi}{\xi} = \hat{s}(t) \tag{11}$$

where the path of integration from the point $\xi = 1$ to the point $\xi = s(t)$ is constrained to be a straight line on the positive real axis from the point $\xi = +1$ to the point $\xi = s(t_0)$ followed thence to the point $\xi = s(t)$ via the continuous curve traced by $s(t)$ in the interval between t_0 and t. A typical path is shown in Fig. 4. The uniqueness of this transformation is assured by the fact that the path of integration is specified completely by the constraint placed upon it. While (11) is often used to define the complex logarithm function, only the end points of the contour are specified in that definition and thus multiples of $j2\pi$ may be added or subtracted by introducing arbitrary encirclements of the pole of unit residue at the origin of the ξ-plane. With our interpretation of the complex logarithm we may write $\hat{s}(t) = \log s(t)$. The inverse system P^{-1} is the complex exponential function.

Equation (11) serves as a formal definition for the transformation P, but requires further practical interpretation. Typically $\hat{s}(t)$ and $s(t)$ are realized as pairs of real signals:

$$s(t) = s_r(t) + js_i(t) \tag{12}$$

$$\hat{s}(t) = \hat{s}_r(t) + j\hat{s}_i(t). \tag{13}$$

These relationships are shown diagrammatically in Fig. 5. We now wish to establish an explicit relationship between the input and output signal pairs. Let us consider $\hat{s}_r(t)$ first. There is never any ambiguity about the real part of a complex logarithm. It is always the real logarithm of the magnitude of the complex argument. Thus we have

$$\begin{aligned} \hat{s}_r(t) &= \log |s(t)| = \log \sqrt{s_r^2(t) + s_i^2(t)} \\ &= (1/2) \log \left[s_r^2(t) + s_i^2(t)\right]. \end{aligned} \tag{14}$$

Next we consider $\hat{s}_i(t)$. Except for an ambiguity of multiples of 2π the imaginary part of the complex logarithm of a complex number is proportional to the angle of the complex number. The provisions we have made for resolving the ambiguity require more than a knowledge of $s_r(t)$ and $s_i(t)$ at a single instant. A complete history in the interval t_0 to t must be employed in the determination. We may accomplish this by noting that, from (11),

$$\frac{d\hat{s}_r(t)}{dt} + j\frac{d\hat{s}_i(t)}{dt} = \frac{d}{dt}\left[\int_1^{s(t)} \frac{d\xi}{\xi}\right] = \frac{1}{s(t)}\frac{ds(t)}{dt} \tag{15}$$

so that

$$\frac{d}{dt}\hat{s}_i(t) = \frac{s_r^2(t)}{s_r^2(t) + s_i^2(t)} \frac{d}{dt}\left[\frac{s_i(t)}{s_r(t)}\right] \tag{16}$$

and

$$\hat{s}_i(t_0) = 0.$$

Thus we can construct an expression for $\hat{s}_i(t)$ in integral form as

$$\hat{s}_i(t) = \int_{t_0}^{t} \frac{s_r^2(t)}{s_r^2(t) + s_i^2(t)} \frac{d}{dt}\left[\frac{s_i(t)}{s_r(t)}\right] dt \tag{17}$$

which becomes

$$\hat{s}_i(t) = \int_{t_0}^{t} \frac{1}{|s(t)|^2}\left[s_r(t)\frac{ds_i(t)}{dt} - s_i(t)\frac{ds_r(t)}{dt}\right] dt. \tag{18}$$

The inverse characteristic system P^{-1} is diagramed in Fig. 6. Explicit expressions for its input-output relations are

$$g_r(t) = e^{\hat{g}_r(t)} \cos \hat{g}_i(t) \tag{19}$$

$$g_i(t) = e^{\hat{g}_r(t)} \sin \hat{g}_i(t). \tag{20}$$

The canonic form for a multiplicative homomorphic system employing complex signals is depicted in Fig. 7. There is a

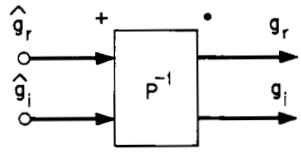


Fig. 6. The system P^{-1} of Fig. 2 represented for complex signals.

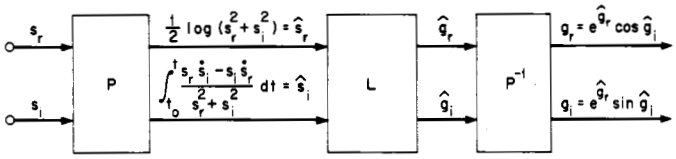


Fig. 7. The canonic representation for a multiplicative filter employing complex signals. $\dot{s}_i$ and $\dot{s}_r$ denote the time derivatives of s_i and s_r.

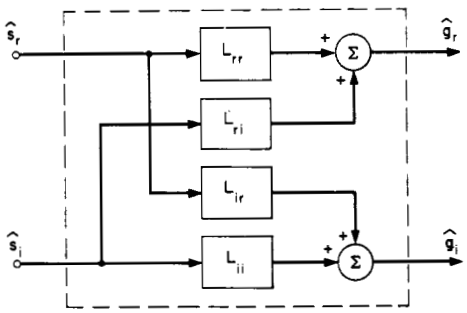


Fig. 8. The general topology for the system L of Fig. 7.

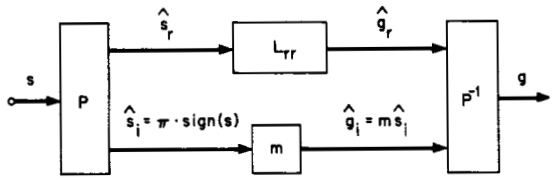


Fig. 9. The multiplicative filter of Fig. 7 specialized to employ real bipolar signals.

residual question concerning the form of the linear system L. If we restrict ourselves to real scalars c, then we place a different set of restrictions upon L than if we allow complex scalars. A completely general topology for the system L is presented in Fig. 8 in which the small internal systems are linear with real signals. In the case of real scalars c, the system L must obey superposition only when the input $s(t)$ is multiplied by a real scalar. Under these circumstances there are no restrictions on the four real linear systems. However, in the case of complex scalars the system L must obey superposition when the input $s(t)$ is multiplied by a complex scalar. This requires that

$$L_{rr} = L_{ii} \tag{21}$$

and

$$L_{ri} = -L_{ir} \tag{22}$$

which restricts the systems L that may be employed.

One of the important practical applications of the above ideas involves signals which are real and bipolar, that is, are sometimes positive and sometimes negative. Formally such signals do not fit within the above framework since the condition that the signals be nonzero and continuous simultaneously cannot be met. Consequently, if we want to apply the notion of homomorphic filtering to this case we must modify the bipolar signals so that they are complex. For example, we may treat the signals as real until they become smaller than some very small value at which time they assume complex values of fixed magnitude ε and varying angle. Another possible method involves adding a very small constant imaginary value to the signal, thus forcing it to be nonzero. All such methods are basically similar in nature and require that complex signals be considered. Formalizing this idea is difficult and seems to offer no real advantage. For the particular application to be discussed, where only one of the two multiplied signals is bipolar, no difficulties arise if we require that $L_{ir}=L_{ri}=0$ and $L_{ii}=m$ (an integer). In this case, the system of Fig. 7 is reduced to that of Fig. 9.

IV. Homomorphic Filtering of Convolved Signals

There are many waveforms of interest which can be considered as a convolution of component signals which we wish to separate. Often, for example, a waveform is corrupted through reverberation, that is, the introduction of echos, which we would like to remove. In speech processing, it is often of interest to isolate the effects of vocal tract impulse response and excitation, which at least on a short-time basis can be considered to have been convolved to form the speech waveform [4]. Another example lies in the separation of probability density functions which have been convolved by the addition of independent random processes.

A common approach to deconvolution is the technique of inverse filtering. In this case the unwanted components of the signal to be processed are removed by filtering with a linear system whose system function is the reciprocal of the Fourier transform of these components. Clearly this method is reasonable only for those situations in which we have a detailed model or description of the components to be removed. This approach has been successful, for example, in recovering the excitation function from the speech waveform since accurate models of the vocal tract have been developed [5]. Inverse filtering is analogous to removing the effect of noise in the additive case (i.e., signal plus noise) by subtraction. If the noise is known exactly except for a few parameters then we might reasonably expect to recover the signal by subtracting the noise from the sum. In many cases, however, we do not have available detailed information about the unwanted components of the signal, and consequently this method of subtraction in the additive case or inverse filtering in the convolutional case is no longer feasible.

In applying the notion of generalized linear filtering to the separation of convolved signals, we must first determine the characteristic system for this class of filters. While we may formulate the results either in terms of continuous or discrete (sampled) inputs, the processing to be described is

most easily realized on a digital computer. Consequently, the discussion will be phrased in terms of discrete time series. Thus, we consider a sequence $s(n)$ consisting of the discrete convolution of two sequences $s_1(n)$ and $s_2(n)$ so that

$$s(n) = \sum_{k=-\infty}^{+\infty} s_1(k)s_2(n-k)$$

or

$$s(n) = s_1(n) \otimes s_2(n) \tag{23}$$

where $\otimes$ denotes a discrete convolution. The canonic form for the class of filters is represented symbolically in Fig. 10 where D is the characteristic system for the class and has the property that

$$D[s_1(n) \otimes s_2(n)] = \hat{s}_1(n) + \hat{s}_2(n) \tag{24}$$

where $\hat{s}_1(n)$ and $\hat{s}_2(n)$ are the responses of D for inputs $s_1(n)$ and $s_2(n)$, respectively.

Let $S(z)$ and $\hat{S}(z)$ denote the two-sided z-transforms of $s(n)$ and $\hat{s}(n)$, respectively, so that

$$S(z) = \sum_{n=-\infty}^{+\infty} s(n)z^{-n} \tag{25a}$$

$$\hat{S}(z) = \sum_{n=-\infty}^{+\infty} \hat{s}(n)z^{-n} \tag{25b}$$

$$s(n) = \frac{1}{2\pi j}\oint_{C_1} S(z)z^{n-1}\,dz \tag{25c}$$

and

$$\hat{s}(n) = \frac{1}{2\pi j}\oint_{C_2} \hat{S}(z)z^{n-1}\,dz \tag{25d}$$

with c_1 and c_2 closed counterclockwise contours of integration in the z-plane. It will be assumed for notational convenience that c_1 and c_2 are always taken to be the unit circle $z = e^{j\omega}$. While this is somewhat restrictive the results obtained are easily modified to incorporate the general case.

It follows from (23) and the properties of the z-transform that

$$S(z) = S_1(z)S_2(z) \tag{26}$$

where $S_1(z)$ and $S_2(z)$ are the z-transforms of $s_1(n)$ and $s_2(n)$, respectively. Hence, from the results of the previous section, applied here to functions of frequency, we may relate $S(z)$ and $\hat{S}(z)$ through a suitably defined logarithmic transformation.

Let us require that both $S(z)$ and $\hat{S}(z)$ be analytic functions with no singularities on the unit circle. Letting

$$S(e^{j\omega}) = S_R(e^{j\omega}) + jS_I(e^{j\omega}) \tag{27a}$$

and

$$\hat{S}(e^{j\omega}) = \hat{S}_R(e^{j\omega}) + j\hat{S}_I(e^{j\omega}), \tag{27b}$$

we then require that S_R, $\hat{S}_R$, S_I, and $\hat{S}_I$ be continuous functions of ω. Since the z-transform is a periodic function of ω with period 2π we require in addition that $\hat{S}_R$ and $\hat{S}_I$ be periodic functions of ω. Furthermore, we may impose the constraint that $s(n)$ and $\hat{s}(n)$ be real functions so that S_R and $\hat{S}_R$ are even functions of ω and S_I and $\hat{S}_I$ are odd functions of ω. Then from (14) and (16) we define

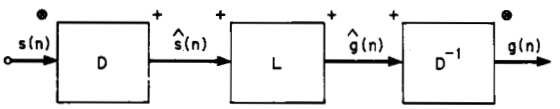


Fig. 10. The canonic representation for a deconvolution filter.

$$\hat{S}_R = \log|S| \tag{28}$$

and

$$\frac{d\hat{S}_I}{d\omega} = \frac{S_R^2}{S_I^2 + S_R^2}\frac{d}{d\omega}\left[\frac{S_I}{S_R}\right] \tag{29}$$

with

$$|\hat{S}_I(e^{j\omega})|_{\omega=0} = 0.$$

Thus the imaginary part of $\hat{S}$ is interpreted to be the angle of S considered as a continuous, odd, periodic function of ω. The response of the system D then corresponds to the inverse transform of the complex logarithm of the transform.

A similar transformation was introduced by Bogert, Healy, and Tukey in which the power spectrum of the logarithm of the power spectrum was proposed as a means for detecting echoes [6]. The result of this set of operations was termed the cepstrum. It is clear that $\hat{s}(n)$ bears a strong relationship to the cepstrum with the primary differences being embodied in the use of the *complex* Fourier transform and *complex* logarithm [7]. To emphasize the relationship while maintaining the distinction, it has been convenient to refer to $\hat{s}(n)$ as the complex cepstrum.

Properties of the Complex Cepstrum

While (28) and (29) define the complex cepstrum, it is possible to reformulate the relationship between $s(n)$ and $\hat{s}(n)$ in several ways which place more in evidence the properties of the complex cepstrum. From (28) and (29)

$$\hat{s}(n) = \frac{1}{2\pi j}\oint_C \log[S(z)]z^{n-1}\,dz. \tag{30}$$

Since the contour of integration is the unit circle and we have defined $\log[S(z)]$ so that it is a single-valued function, we may rewrite (30) as

$$\hat{s}(n) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \log[S(e^{j\omega})]e^{j\omega n}\,d\omega.$$

Integrating by parts and using the fact that $S_I(e^{j\omega})$ is restricted to be a continuous, odd, periodic function of ω, we obtain the result that

$$\hat{s}(n) = \begin{cases} -\dfrac{1}{2\pi jn}\oint z\dfrac{1}{S(z)}\dfrac{dS(z)}{dz}z^{n-1}\,dz & n \neq 0 \\ \dfrac{1}{2\pi}\displaystyle\int_{-\pi}^{\pi} \log|S(e^{j\omega})|\,d\omega & n = 0. \end{cases} \tag{31}$$

An example of a class of functions $S(z)$ satisfying the requirement that both $S(z)$ and $\hat{S}(z)$ be analytic is the class of the form

$$S(z) = |K| \frac{\prod_{i=1}^{M_0} (1 - a_i z^{-1}) \prod_{i=1}^{M_1} (1 - b_i z)}{\prod_{i=1}^{P_0} (1 - c_i z^{-1}) \prod_{i=1}^{P_1} (1 - d_i z)} \tag{32}$$

where the a_i and c_i are the zeros and poles inside the unit circle and $(1/b_i)$ and $(1/d_i)$ are the zeros and poles outside the unit circle. For this class of examples, we note that the poles of the integrand in (31) occur at the poles and zeros of $S(z)$. Consequently, $\hat{s}(n)$ will be composed of a sum of exponentials divided by n.

Equation (31) can be rewritten in a somewhat different form by noting that, from (15),

$$\frac{d\hat{S}(z)}{dz} = \frac{1}{S(z)} \frac{dS(z)}{dz}$$

or

$$S(z) \frac{d\hat{S}(z)}{dz} = \frac{dS(z)}{dz}. \tag{33}$$

Using the fact that $d\hat{S}(z)/dz$ is the z-transform of $-n\hat{s}(n)$, and $dS(z)/dz$ is the z-transform of $-ns(n)$, the inverse z-transform of (33) is

$$[n\hat{s}(n)] \otimes s(n) = ns(n)$$

or

$$\sum_{k=-\infty}^{+\infty} \frac{k}{n} \hat{s}(k) s(n-k) = s(n) \qquad n \neq 0. \tag{34}$$

In general, this is an implicit relation between $s(n)$ and $\hat{s}(n)$ and cannot be computed. However, if it is assumed that $s(n)$ and $\hat{s}(n)$ are zero for n negative and that $s(0) \neq 0$, then (34) becomes

$$\hat{s}(n) = \begin{cases} \dfrac{s(n)}{s(0)} - \displaystyle\sum_{k=0}^{n-1} \frac{k}{n} \hat{s}(k) \frac{s(n-k)}{s(0)} & n \neq 0 \\ \log s(0) & n = 0. \end{cases} \tag{35}$$

For this case, the inverse of the characteristic system can be easily obtained by solving (35) for $\hat{s}(n)$ in terms of $s(n)$ with the result that

$$s(n) = \begin{cases} s(0)\hat{s}(n) + \dfrac{1}{n} \displaystyle\sum_{k=0}^{n-1} k\hat{s}(k) s(n-k) & n \neq 0 \\ e^{\hat{s}(0)} & n = 0. \end{cases} \tag{36}$$

The Complex Cepstrum of Minimum Phase Sequences

For a function $f(t)$ which is zero for $t<0$, the real and imaginary parts of its Fourier transform are related through the Hilbert transform. This relationship is derived by noting that $f(t)$ is uniquely expressible in terms of its even part [8].

For certain classes of functions, the magnitude and phase of the Fourier transform are also related through the Hilbert transform and such functions are generally referred to as minimum phase functions. The relationship between magnitude and phase is derived by treating the log magnitude and phase of the Fourier transform as the real part and imaginary part, respectively, of a new Fourier transform. If the time function associated with this new Fourier transform is zero for $t<0$, then its real and imaginary parts are related. An entirely parallel set of statements can be made for discrete sequences, with the Fourier transform replaced by the z-transform evaluated on the unit circle.

From the above discussion we note that a minimum phase sequence is one for which the complex cepstrum $\hat{s}(n)$ is zero for $n<0$, which is the same condition imposed in deriving (35). Thus we may conclude that for input sequences which are minimum phase, the input and output of the system D are related by the recursion relation of (35) and the input and output of the system D^{-1} are related by the recursion relation of (36). These recursion relations do not necessarily offer a computational advantage. However, they are conceptually important. In particular, they bring to light the fact that for minimum phase inputs, the transformation D is a realizable transformation, i.e., the response $\hat{s}(n)$ for $n=n_0$ is dependent only on samples of the input for $n \leq n_0$. Similarly, the inverse transformation D^{-1} is realizable for input sequences $\hat{s}(n)$ which are zero for $n<0$. From this we may conclude that for minimum phase input sequences the class of homomorphic filters defined by the canonic form of Fig. 10 is realizable in the sense that the output depends only on previous values of the input if the linear filter is also realizable.

An analogous discussion can be carried out for sequences whose complex cepstrum is zero for $n>0$. Such sequences, which have no minimum phase components, could appropriately be called maximum phase sequences. For these cases a relation similar to (35) can be derived, in which values of the complex cepstrum depend only on future rather than past values of the input. It should be remarked that any sequence can always be expressed as the convolution of a minimum phase sequence and a maximum phase sequence, i.e.,

$$s(n) = s_1(n) \otimes s_2(n).$$

The portion of $\hat{s}(n)$ for $n>0$ represents the contribution from the minimum phase component, and the portion for $n<0$ represents the contribution from the maximum phase component.

As a result of these considerations an interesting and perhaps useful result emerges. Consider a time-limited sequence $s(n)$ which contains $(N+1)$ samples. Let us choose the origin and polarity of the waveform so that $S(z)$ can be expressed in the form of (32). Now $s(n)$ can be expressed as the convolution of a minimum phase sequence $s_1(n)$ and a maximum phase sequence $s_2(n)$ where $s_1(n)$ and $s_2(n)$ are time-limited so that

$$s_1(n) \neq 0 \quad 0 \leq n \leq N_1$$
$$= 0 \quad \text{otherwise}$$

and

$$s_2(n) \neq 0 \quad -N_2 \leq n \leq 0$$
$$= 0 \quad \text{otherwise}$$

where

$$N_1 + N_2 = N.$$

The complex cepstrum of $s_1(n)$ is in general not time-limited. However, $\hat{s}_1(n)$ is zero for $n<0$ and $\hat{s}_2(n)$ is zero for $n\geq 0$. Thus, from (34),

$$s_1(n) = \begin{cases} e^{\hat{s}_1(0)} & n = 0 \\ \hat{s}_1(n)s_1(0) + \sum_{k=0}^{n-1} \left(\frac{k}{n}\right)\hat{s}_1(n)s_1(n-k) & n > 0 \end{cases}$$

and

$$s_2(n) = \begin{cases} 1 & n = 0 \\ \hat{s}_2(n) + \sum_{k=n+1}^{0} \left(\frac{k}{n}\right)\hat{s}_2(k)s_2(n-k) & n < 0. \end{cases}$$

Consequently, (N_1+1) values of $\hat{s}_1(n)$ are needed to recover $s_1(n)$ and N_2 values of $\hat{s}_2(n)$ are needed to recover $s_2(n)$, so that (N_1+N_2+1) values of the complex cepstrum are needed to obtain the (N_1+N_2+1) values of $s(n)$.

Sequences with Rational z-Transforms

Thus far, we have restricted the input sequences to be such that $S(z)$ and $\hat{S}(z)$ are analytic and for these cases, the logarithm of the z-transform on the unit circle was defined such that the imaginary part of the logarithm was a continuous, odd, periodic function of ω. It was remarked that this included all sequences with z-transforms of the form of (32). It is reasonable to assume that most input sequences of interest can be represented at least approximately by z-transforms which are rational, of the form

$$S(z) = Kz^r \frac{\prod_{i=1}^{M_0}(1 - a_i z^{-1}) \prod_{i=1}^{M_1}(1 - b_i z)}{\prod_{i=1}^{P_0}(1 - c_i z^{-1}) \prod_{i=1}^{P_1}(1 - d_i z)}. \tag{37}$$

Equation (37) differs from (32) in the inclusion of a term z^r representing a delay or advance of the sequence and removal of the absolute value on the multiplying constant so that $S(z)$ is no longer required to be positive for $z=1$ ($\omega=0$).

While it is possible to generate a formal structure which would include this more general case, it offers no real advantage. Specifically, if we consider the problem at hand, namely, carrying out a separation of convolved signals, we would not expect to be able to determine, and most likely would not be interested in determining how much of the constant K, including its sign, was contributed by each. Similarly we could not expect to be able to determine how much of the net advance or delay r was contributed by each. In summary, we can expect to be generally interested in the shape of the components and not their amplitudes or time origin.

If we are willing to permit this flexibility, then we can measure the algebraic sign of K and the value of r separately and then alter the input (or its transform) so that the z-transform is in the form of (32).

Computation of the Complex Cepstrum

On the basis of the previous discussion, for general input sequences the computation of the complex cepstrum requires a computation of the Fourier transform of the input. Thus practical considerations require that the input $s(n)$ contain only a finite number of points, that is, be time-limited, and that the transform be computed only at discrete frequencies. Thus, in an implementation of the transformation D, we replace the z-transform and its inverse by the discrete Fourier transform pair (DFT) defined as

$$F(k) = \sum_{n=0}^{N-1} f(n)W^{nk}$$

and

$$f(n) = \frac{1}{N}\sum_{k=0}^{N-1} F(k)W^{-nk}$$

where

$$W = e^{-2\pi j/N}.$$

Thus, the complex cepstrum computed by use of the DFT is given by

$$\begin{aligned} \hat{s}_d(n) &= \frac{1}{N}\sum_{k=0}^{N-1} [\log S(k)]W^{-nk} \\ &= \frac{1}{N}\sum_{k=0}^{N-1} [\log |S(k)| + j\theta(k)]W^{-nk} \end{aligned} \tag{38}$$

with

$$S(k) = \sum_{n=0}^{N-1} s(n)W^{nk}.$$

It is straightforward to verify that $\hat{s}_d(n)$ is an aliased version of $\hat{s}(n)$, i.e.,

$$\hat{s}_d(n) = \sum_{a=-\infty}^{+\infty} \hat{s}(aN + n).$$

The effect of the aliasing depends on the value chosen for the rate at which the spectrum is sampled, or equivalently the value of N. In many cases this is not a severe problem since relatively fast and efficient means for computing the discrete Fourier transform for large N have recently been developed [9].

The phase curve $\theta(k)$ can be computed by first computing the phase modulo 2π and then "unwrapping" it to satisfy the requirement that it be continuous and odd. Simple algorithms for doing this are easily generated, provided that the frequency spacing of adjacent points is sufficiently small.

As an alternative to computing the complex cepstrum by means of (38), we may obtain $\hat{s}(n)$ by forming the ratio of the derivative of the spectrum and the spectrum, as suggested by (31). In particular, since samples of the derivative of the spectrum, denoted by $\tilde{S}(k)$, can be obtained by

$$\tilde{S}(k) = -j \sum_{n=0}^{N-1} nf(n)W^{nk}$$

we obtain $\hat{s}_d(n)$ as

$$\hat{s}_d(n) = -\frac{1}{jn}\frac{1}{N}\sum_{k=0}^{N-1}\frac{\tilde{S}(k)}{S(k)}W^{-nk}. \tag{39}$$

The complex cepstrum computed on the basis of (39) differs somewhat from that computed from (38). The difference can be expressed by observing that $n\hat{s}_d(n)$ as represented by (39) is an aliased replica of $n\hat{s}(n)$, i.e.,

$$n\hat{s}_d(n) = \sum_{a=-\infty}^{+\infty}(aN + n)\hat{s}(aN + n)$$

or

$$\hat{s}_d(n) = \frac{1}{n}\sum_{a=-\infty}^{+\infty}(aN + n)\hat{s}(aN + n). \tag{40}$$

We note that in general the effect of the aliasing introduced by the use of (39) is more severe than that introduced by (38). On the other hand, use of (38) requires the explicit computation of the unwrapped phase curve, whereas use of (39) does not.

V. Applications of Homomorphic Filtering

In the preceding paragraphs we have discussed the analytical aspects of homomorphic filtering in general, and multiplicative and convolutional filtering in particular. We now wish to deemphasize the theory and concern ourselves with specific practical applications. The following discussions serve two distinct purposes. The first is to disclose a specific technology which has emerged as a direct result of the theory. The second is to lend to the theory a set of examples which hopefully will serve to clarify concepts and to foster further investigation.

The Multiplicative Processing of Audio Signals

The first application of multiplicative filtering to be discussed involves the processing of audio signals [10]. We are all familiar with the idea of analyzing audio waveforms as sums of harmonic oscillations. However, for the purposes of this discussion we conceive of analyzing audio waveforms as a product instead of a sum. Specifically, we are motivated by the obvious fact that audio signals bear a resemblance to amplitude-modulated waves. They are similar because each grows larger and smaller at a relatively slow rate while dancing around at some other relatively fast rate. In the case of audio the fast motion is irregular and varied. In the case of AM it is neither. In this respect, they are not similar. There are factors in the manufacture of audible signals which are certainly multiplicative in nature. A person modulates his voice both consciously and unconsciously. Musicians play loud and soft passages. Sounds form and die away as their energy is absorbed.

With this motivation let us analyze an audio signal as a product of two components. Let the first be an envelope $e(t)$ which is slowly varying but always positive. Let the second be a carrier or vibration $v(t)$ which is rapidly varying and bipolar. If we call our audio signal $s(t)$, we obtain

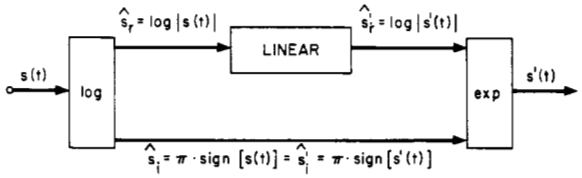


Fig. 11. A multiplicative filter for audio processing.

$$s(t) = e(t)\cdot v(t). \tag{41}$$

Furthermore, let us process this signal with a multiplicative homomorphic filter such that the response $s'(t)$ will be given by

$$s'(t) = e'(t)\cdot v'(t) \tag{42}$$

where $e'(t)$ and $v'(t)$ are the responses for $e(t)$ and $v(t)$ acting separately. Fig. 11 shows this situation in accordance with our previous discussions concerning real bipolar signals and multiplicative filters. We have set $m=1$ since we wish to preserve the sign information embodied in $v(t)$. Notice that since $e(t)$ is always positive

$$\hat{e}_r(t) = \log|e(t)| = \log e(t) \tag{43}$$

and

$$\hat{e}_i(t) = 0 \tag{44}$$

such that

$$\hat{s}_i(t) = \hat{v}_i(t). \tag{45}$$

Furthermore, since

$$\hat{s}'_i(t) = \hat{s}_i(t) \tag{46}$$

it follows that

$$\hat{e}'_i(t) = 0 \tag{47}$$

which implies that $e'(t)$ is always positive as well. More explicitly

$$\hat{s}'_i(t) = \hat{v}'_i(t) \tag{48}$$

and

$$\hat{e}'_r(t) = \log|e'(t)| = \log e'(t). \tag{49}$$

With these equations in mind, we can reconstruct Fig. 11 as in Fig. 12.

If $\log e(t)$ and $\log|v(t)|$ were to possess frequency components occupying separate frequency bands, linear systems could be designed to perform different processing tasks upon each. While it is almost certainly true that any preconceived definition of $e(t)$ and $v(t)$ would not result in this condition, an extremely interesting situation is revealed through the examination of the spectra of $\log|s(t)|$ for typical audio signals. An example of such a spectrum is shown in Fig. 13. This curve is derived from a computer calculation of the periodogram of the log magnitude of seven seconds of speech waveform. We see that above a certain critical fre-

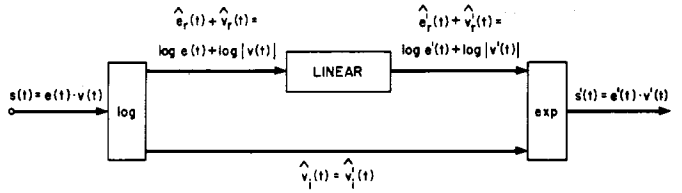


Fig. 12. The filter of Fig. 11 with component signals shown explicitly.

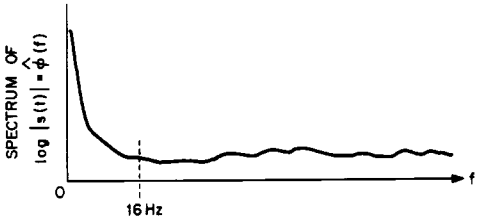


Fig. 13. A typical spectrum of log $|s(t)|$ for audio signals.

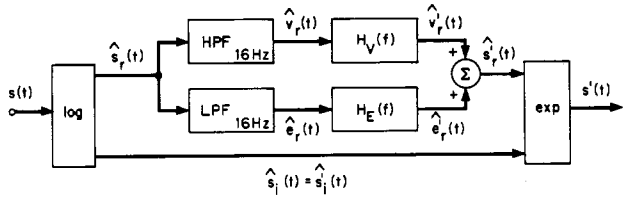


Fig. 14. A multiplicative filter which processes $\hat{E}(f)$ and $\hat{V}(f)$ separately.

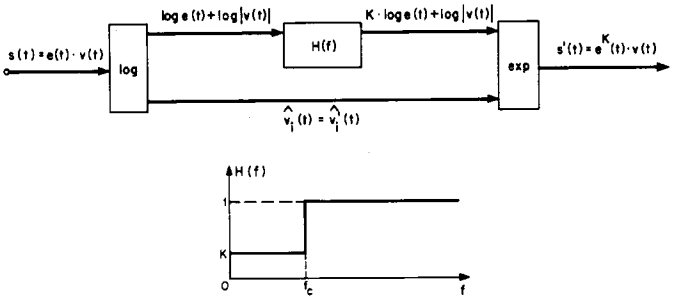


Fig. 15. The system of Fig. 14 with $H_V(f)=1$ and $H_E(f)=K$.

quency near 16 Hz the spectrum is nominally constant, but below that frequency it grows rapidly with decreasing frequency. This behavior suggests that this spectrum can be broken into two components, one more or less constant at all frequencies and the other large at low frequencies but decreasing rapidly with increasing frequency. Although these components overlap each other it is clear that the character of their spectral behavior is sufficiently different to permit effective partial separation by the methods of linear filtering.

For the sake of simplicity, however, let us assume that $\hat{\phi}(f)$ can be broken into two parts occupying distinct frequency bands. Let the first part, called $\hat{E}(f)$, be constituted from all components of $\hat{\phi}(f)$ below 16 Hz. Let the second part, called $\hat{V}(f)$, be constituted from all components of $\hat{\phi}(f)$ above 16 Hz. For any specific audio signal we will associate $\hat{E}(f)$ with the envelope signal and $\hat{V}(f)$ with the vibration signal.

A multiplicative filter which processes each of these components independently is shown in Fig. 14. The high-pass and low-pass filters serve to isolate the component signals of $\hat{s}_r(t)$ and the separate different linear filters $H_V(f)$ and $H_E(f)$ operate upon each independently.

An extremely interesting subclass of filters is generated by considering $H_V(f)=1$. This class leaves $v(t)$ entirely unaffected while operating only on the envelope $e(t)$. A simple specific example is formed by choosing $H_E(f)=K$. For this choice we can reconstitute the filter as shown in Fig. 15 where we find a single frequency response specification for the linear system. Since this system has a gain of unity for vibration signals, they are unaffected by the filter. However, for envelope signals the gain is K, and so for them the system is a power law device with exponent K. This arrangement results in a response $s'(t)$ as given by

$$s'(t) = e'(t)v'(t) = e^K(t)\cdot v(t). \tag{50}$$

If $K<1$, the envelope function is subject to rooting action, thus reducing the dynamic range of the composite signal $s(t)$. If $K>1$, this dynamic range is increased. In this way this multiplicative filter acts as a volume compressor or expander.

If $K=0$ the envelope signal response $e'(t)$ is reduced to unity and $s'(t)=v(t)$. This situation is similar to that obtained with automatic gain control circuits, because the amplitude of response is not dependent upon the amplitude of excitation.

A system based upon the diagram of Fig. 15 for effecting the modification of dynamic range has been simulated on the TX-2 computer at the M.I.T. Lincoln Laboratory and has been constructed for audio signals and employed in practice with remarkable success. Some interesting practical considerations arise in this respect which are worth mentioning here.

The first has to do with the realization of the system function $H(f)$ described in Fig. 15. The ideal filter $H(f)$ can be approximated in practice by a lumped parameter system $L(f)$. Whether employing few or many degrees of freedom, if $|L(0)|=K$ and $|L(\infty)|=1$ the basic operating characteristics discussed above can be realized. This is so because the need for a sharp transition characteristic is not implied by Fig. 13. The reader may ask about the effect of the phase characteristics of $L(f)$ upon this situation. Phase will have negligible effect upon system performance as long as it approaches zero as frequency becomes infinite. This condition assures that there is no delay for the high-frequency components of log $|v(t)|$, which is sufficient for a proper reconstruction of the axis crossing behavior of the component $v(t)$.

Another important practical consideration is that the bandwidth required for transmitting and processing log $|s(t)|$ is considerably wider than that required for $s(t)$ itself. This fact is best appreciated by reference to Fig. 16 which shows a typical $s(t)$ and log $|s(t)|$. Notice that as $s(t)$ passes through zero, log $|s(t)|$ attempts to become negatively infinite. In mathematical terms, when $s(t)$ possesses a zero, log $|s(t)|$ possesses a logarithmic pole. These logarithmic poles must be reproduced fairly well if the zero crossing behavior of $s'(t)$ is to resemble that of $s(t)$. In practice the bandwidth

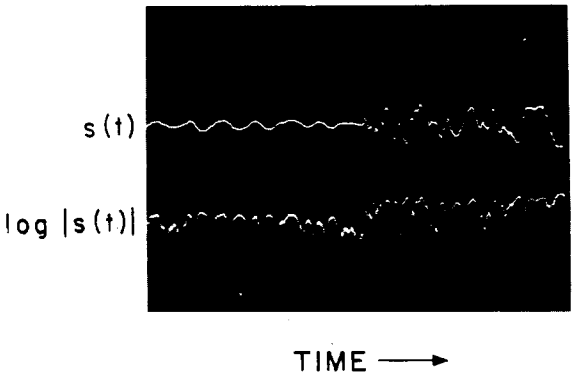


Fig. 16. A typical $s(t)$ and log $|s(t)|$ as measured in the laboratory.

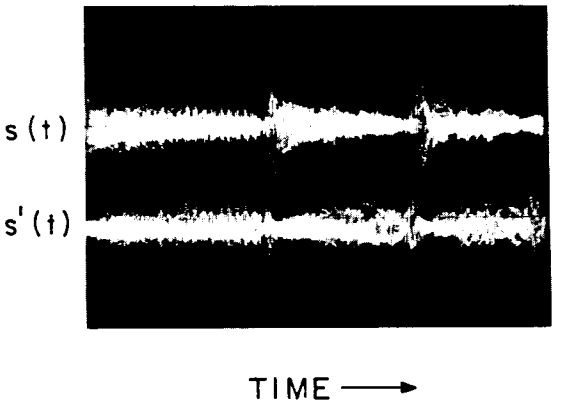


Fig. 17. A typical $s(t)$ and its supercompressed counterpart.

Fig. 18. A compression-expansion system employing a pair of complementary multiplicative filters.

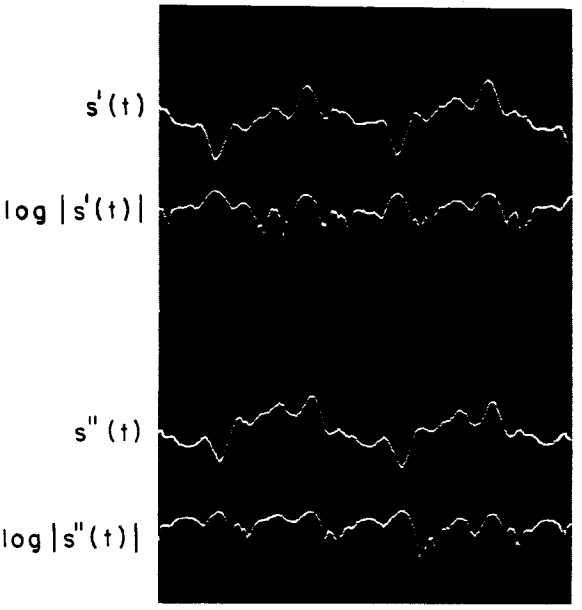


Fig. 19. $s(t)$ and log $|s(t)|$ before and after envelope distortion due to an ac coupled channel.

required for audio is a few kilohertz. One hundred or one thousand times this bandwidth might be required to process log $|s(t)|$ satisfactorily depending upon the degree of precision required. This performance is easily achieved with present technology.

An unusual situation arises if K is made negative. A form of supercompression results in which the role of loud and soft are reversed. For $K=-1$ very loud sounds become barely audible while barely audible sounds become very loud. Fig. 17 shows some typical signals.

It is possible to construct a simple filter $L(f)$ which causes compression and which also has a simple inverse filter $L^{-1}(f)$ which causes exactly complementary expansion. In this way a new type of compression-expansion system can be constructed. Fig. 18 shows such a system in which a noisy channel is to be upgraded for audio use by pre-compression and post-expansion [11]. If the channel is assumed perfect, which of course it is not, then the compressor and expander, being exactly inverse systems, operate as an exactly compensating pair and the received signal $s'''(t) \equiv s(t)$. For only mildly degraded channels this situation is closely approximated. This statement has been demonstrated empirically in the laboratory in applications involving data recording channels. In these situations it was discovered that the channel imperfection which most seriously affects performance is phase shift at low audio frequencies. For dc coupled channels this is never a problem, however. The level of degradation is mild in typical ac coupled situations but provides an occasional serious problem in applications requiring maximum quality. The mechanism of the difficulty is easily described in terms of a hypothetical $s(t)$ composed of two harmonically related sine waves, the amplitude and phase of which are adjusted to yield a waveform which is small for a large percentage of time. The log of this signal will have a relatively small average value as a result of this property. If the two sine waves are shifted in relative phase by 30° or so, $s(t)$ would no longer be small for such a large percentage of time and thus log $|s(t)|$ would have a larger average value. In this way channel phase shift at low frequencies can cause mild envelope distortion in terms of the operation of a multiplicative compressor. See Fig. 19.[2]

The Multiplicative Processing of Images

The second application of multiplicative filtering to be discussed involves the processing of images. This application is motivated very directly, because image formation is predominantly a multiplicative process. This statement applies equally to natural and photographic images [13]. In a natural scene, the illumination and reflectance of objects are combined by multiplication to form observable brightness. The illumination and reflectance in a scene vary independently from object to object and from point to point, thus forming a brightness image. In terms of its projection onto

[2] A channel imperfection which might at first seem troublesome is additive channel noise. It has been verified both theoretically [12] and experimentally that additive noise at moderate levels does not have a major effect on the performance of the type of systems which we are discussing.

the retina this image forms a two-dimensional spatial signal as expressed by

$$I_{x,y} = i_{x,y} \cdot r_{x,y} > 0 \tag{51}$$

where $I_{x,y}$ is the image, $i_{x,y}$ is its illumination component, and $r_{x,y}$ is its reflectance component.[3]

The first step in the production of a photographic scene is usually the manufacture of a negative transparency. The name negative is correct only in a multiplicative sense because it is really an *inverse* image as expressed by

$$N_{x,y} = \frac{1}{I_{x,y}} = I_{x,y}^{-1} = i_{x,y}^{-1} \cdot r_{x,y}^{-1} \tag{52}$$

where $N_{x,y}$ is the negative image. If two such negatives are superimposed by placing the transparencies one on top of the other,[4] a third negative image is formed. That combined image is the *product* of the two components as given by

$${}^{(3)}N_{x,y} = {}^{(1)}N_{x,y} \cdot {}^{(2)}N_{x,y} = \frac{1}{{}^{(1)}I_{x,y} \cdot {}^{(2)}I_{x,y}}. \tag{53}$$

Thus if we wish to process images using a homomorphic system that combines its signals according to the law of image formation, that system must obey superposition multiplicatively.

The image processor thus formulated is depicted in Fig. 20 in accordance with our previous multiplicative discussions concerning real positive signals. It has a response image

$$I'_{x,y} = i'_{x,y} \cdot r'_{x,y} > 0 \tag{54}$$

which is the product of the separately processed illumination and reflectance components. This formulation places in evidence three important properties of multiplicative image processors. Regardless of the specific process invoked by the system, the response $I'_{x,y}$ is always a physically meaningful image in the sense that it cannot contain points of negative brightness. The effect that the process has upon the appearance of objects in a scene is independent of the light falling upon those objects, whether bright, dim, or variable. Similarly, the effect that the process has upon the apparent light falling upon objects is independent of the nature of those objects.

If $\log i_{x,y}$ and $\log r_{x,y}$ were to possess frequency components occupying separate frequency areas,[5] the image processor could be designed to perform different processing tasks upon the illumination and reflectance components of an image. In practice the illumination and reflectance components of typical images behave in a manner similar to the envelope and vibration functions of the audio application. Illumination generally varies slowly, while reflection is sometimes static and sometimes dynamic, because objects vary in texture and size and almost always have well-defined

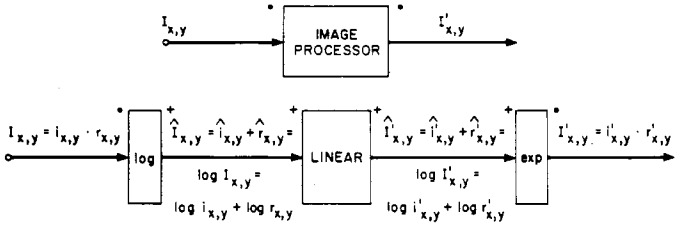


Fig. 20. A multiplicative filter for image processing.

edges. Thus only partially independent processing is possible.

A quantitative measure of the actual spectral content of typical logged images was obtained through computer analysis of the four scenes presented in Fig. 21. The two-dimensional periodograms for the various $\log I_{x,y}$ were computed. They are shown in Fig. 22 where relative brightness represents the magnitude of the periodograms on a decibel scale. In all cases the lowest frequency components are very dominant. To produce a broader view of spectral content the log images were processed by a whitening filter and the periodograms reevaluated. The results of this alternative process are presented in Fig. 23. The two-dimensional frequency characteristics of the whitening filter are shown in Fig. 24. Although the whitening process was only approximate the corresponding periodograms clearly show the high-frequency components of the log images.

The logarithms of all four tested scenes were characterized by an extreme peak in low-frequency energy content and, with minor exceptions, had similar whitened spectra. This two-dimensional data is reminiscent of the one-dimensional spectrum computed for typical log audio and shown in Fig. 13. Again there is an implicit suggestion of two spectral components, one a low-frequency peak, the other a middle- and high-frequency plateau. While it is probably incorrect to associate the peak solely with physical illumination and the plateau entirely with object reflectance, an approximate association of this type has proved most useful in effecting designs involving partially independent processing of the corresponding image components. Before discussing such designs let us explore some of the simpler aspects and uses of the image processor.

If the linear component of the image processor is chosen as a simple amplifier or attenuator with gain γ, the image processor becomes a power law device. The output is given in terms of the input by

$$I'_{x,y} = I_{x,y}^{\gamma}. \tag{55}$$

The parameter gamma is well known to photographers who, by selecting from a variety of photographic materials and using shorter or longer development times for them, control its numerical value. For negative photographic materials $\gamma < 0$ and so they can be thought of as multiplicative inverters.

The general situation calls for the linear component of the image processor to possess a gain which is a function of frequency in two dimensions. Using script letters to represent

[3] We preclude zero image values on practical grounds.

[4] Such superpositions are a prevalent practice in professional photography, expecially color lithography.

[5] Recall that for images the frequency domain is two-dimensional.

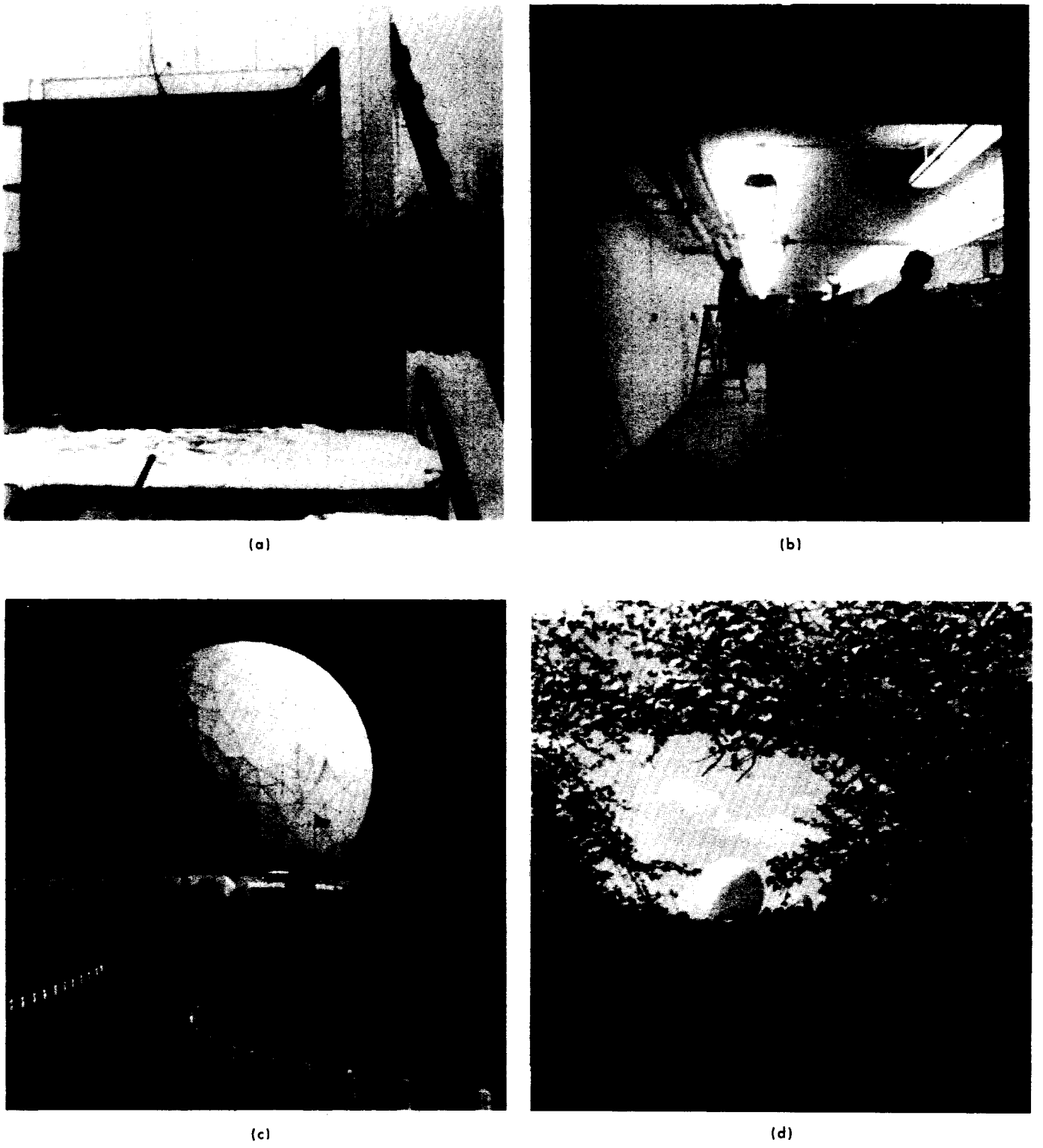


Fig. 21. Four original images.

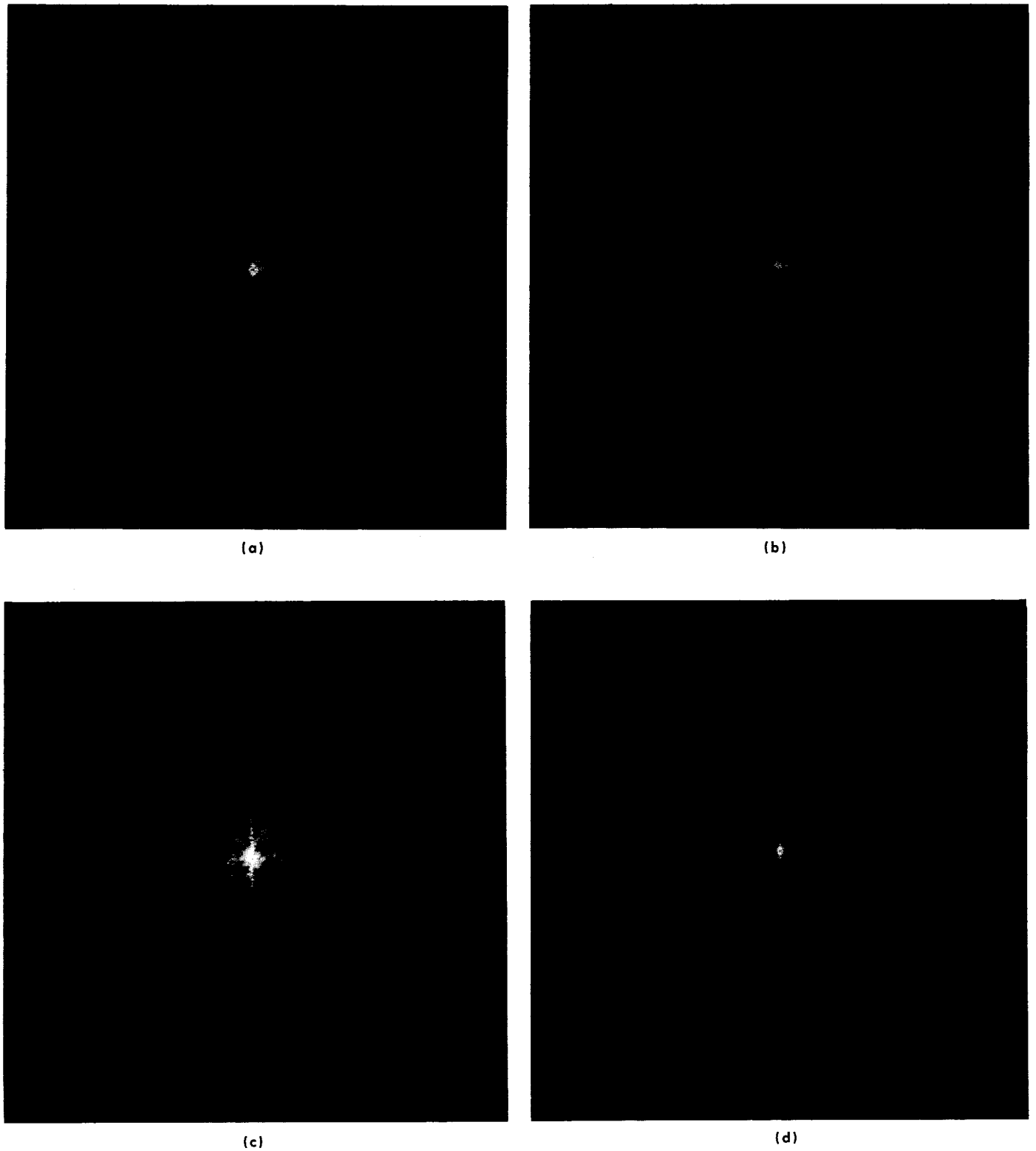


Fig. 22. Log periodograms for the log images corresponding to Fig. 21.

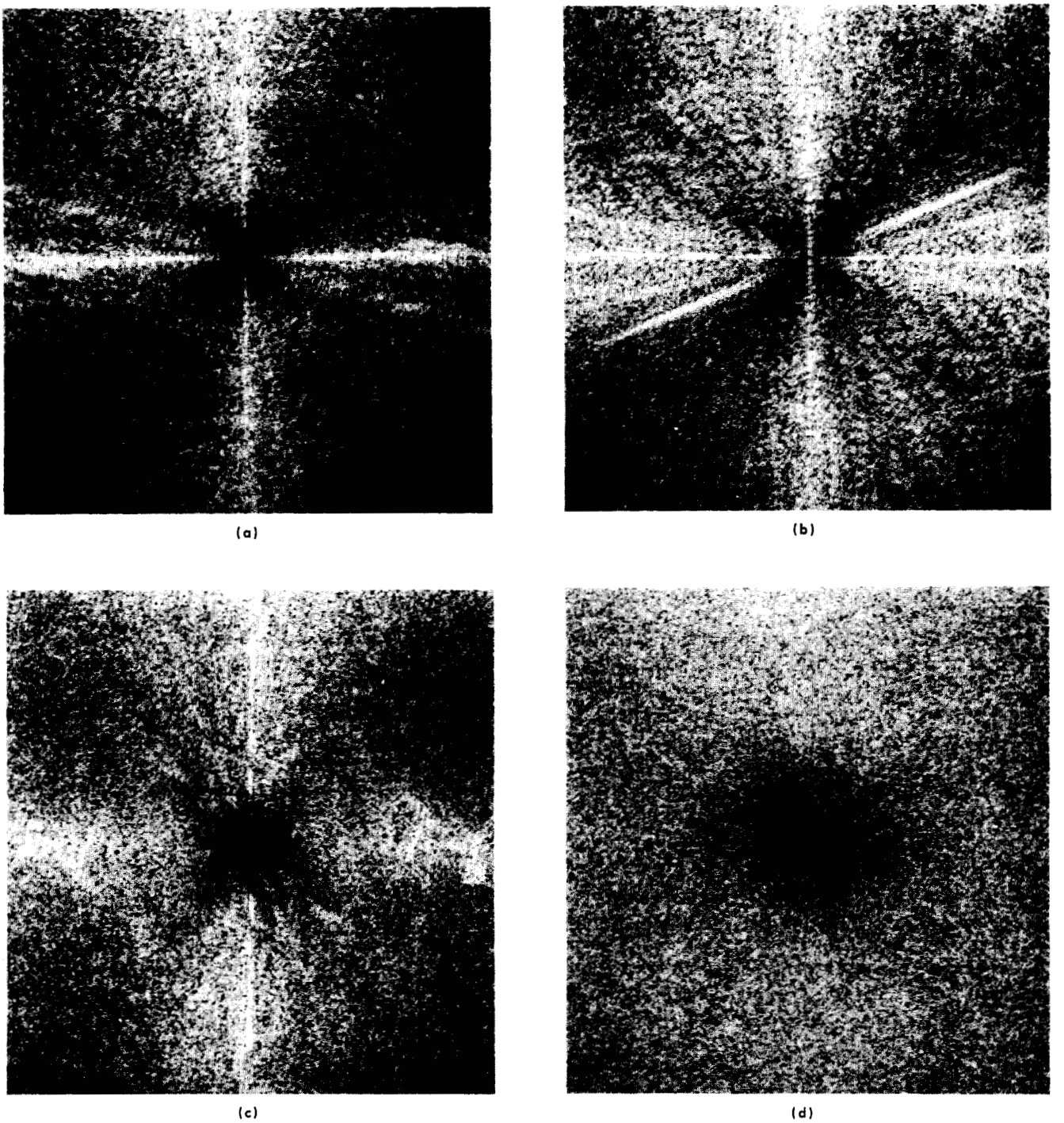


Fig. 23. Log periodograms for the whitened log images corresponding to Fig. 21.

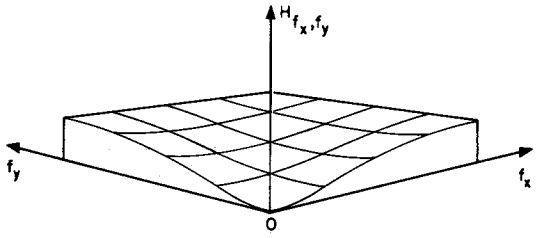


Fig. 24. The whitening filter frequency response. One quadrant shown.

two-dimensional Fourier transforms, and X and Y to represent the frequency variables corresponding to x and y, we can then write

$$\hat{\mathcal{I}}'_{X,Y} = \hat{\mathcal{I}}_{X,Y} \cdot \gamma_{X,Y}.$$

Under these circumstances the image processor can be thought of as having a frequency-sensitive gamma in the sense that it exhibits a different power law behavior for each sinusoidal component of the logarithm of the input image. If we write the input image as a product of components having sinusoidal logarithms, we obtain the following double product:

$$I_{x,y} = \prod_X \prod_Y P_{X,Y;x,y}. \tag{56}$$

The output image is then given by a double product in which each of these components is raised to the appropriate power:

$$I'_{x,y} = \prod_X \prod_Y (P_{X,Y;x,y})^{\gamma_{X,Y}}. \tag{57}$$

A problem common to all forms of image technology is that of excessive dynamic range. Scenes with excessive light-to-dark ratios are usually handled by cramming them to fit the available medium with the result that highlights lose their bright luster and lowlights are without detail. The four scenes of Fig. 21 have been treated in this manner. In an alternate approach gamma may be selected as less than unity so that the image has a reduced ratio and may be reproduced without exceeding the limited dynamic range of practical media. If carried to extremes, this gives the image a muddy or washed-out appearance. Fig. 25(a) shows the original scene of Fig. 21(a) after it has been processed by an image processor with a gamma of one-half. For some applications, gammas greater than one are used in an attempt to give scenes more sharpness about the edges of objects. For example, Fig. 25(b) shows the scene of Fig. 21(a) processed by a gamma of two. A common consequence is that dynamic range capacities are exceeded even more than before.

While the reduction of dynamic range and the enhancement of edge sharpness seem to be conflicting objectives, it is possible to achieve both simultaneously by employing a multiplicative image processor having a linear component with a frequency-dependent gain. To explain this we must return to our previous discussions concerning the partially independent processing of illumination and reflectance components.

The large dynamic range encountered in natural images is contributed to mostly by large variations in illumination $i_{x,y}$, which we recall contains primarily low frequencies in its logarithm. The edges of objects, on the other hand, contribute only to the reflectance component $r_{x,y}$ of a scene, indeed, primarily to the high frequencies of its logarithm. It follows that if one desires to maintain normal contrast for the details of an image, but demands a reduction in dynamic range, the gain of the linear component should be unity for high frequencies and less than unity at low frequencies. This situation is identical to that considered for the audio compressor.

Fig. 25(c) presents the image of Fig. 21(a) processed with the system of Fig. 20 in which the linear system had the frequency response depicted in Fig. 26. This filter was chosen to be spatially isotropic and so a one-dimensional plot of frequency response is sufficient for its unique specification. It also follows that the phase shift of the linear filter was zero at all frequencies. Notice that at low frequencies the gain of the filter was one-half, while at high frequencies it was unity. In Fig. 25(c) the areas which were dark in the original scene have been made far more visible as if illuminated with auxiliary lights, but without disturbing the rendition of the brightly lighted areas.[6] This effect and the effect of sharpening through the use of gammas larger than one are obtained simultaneously in the picture of Fig. 27. In this case, the frequency response was as given in Fig. 28. Again the filter had isotropic properties, so a one-dimensional frequency response curve was sufficient to specify it uniquely, and again the phase was zero. Notice that while the gain at low frequencies was maintained at one-half, the gain at high frequencies was increased to two. Fig. 27 is a kind of blend of Fig. 25(a) and (b) in which the best properties of both have been retained and the worst properties of both greatly reduced. Relying upon our approximate analysis which assigns the lowest-frequency components of an image to the illumination and the higher-frequency components to the reflectance, we can see that according to our definition of partially independent processing the illumination component has been treated by a gamma equal to one-half while the reflectance components have been treated by a gamma equal to two. This situation is summarized as follows:

$$I'_{x,y} = i'_{x,y} \cdot r'_{x,y} \approx i^{0.5}_{x,y} \cdot r^2_{x,y}. \tag{58}$$

Since the large brightness ratios in a subject are usually produced by large variations in illumination, the fact that $I'_{x,y}$ contains the square root of the original illumination explains why the brightness ratio is reduced. On the other hand, the reflectance component has the same [Fig. 25(c)] or greater (Fig. 27) variability than in the original and thus details are preserved or enhanced. As has been stated, (58) is only an approximate equation. In fact, the illumination which formed the scene certainly contains some high-frequency components while the reflectance function certainly

[6] Results similar to this can be obtained by means of the classical photographic technique of unsharp masking [14] or by use of a logetronics printer [15].

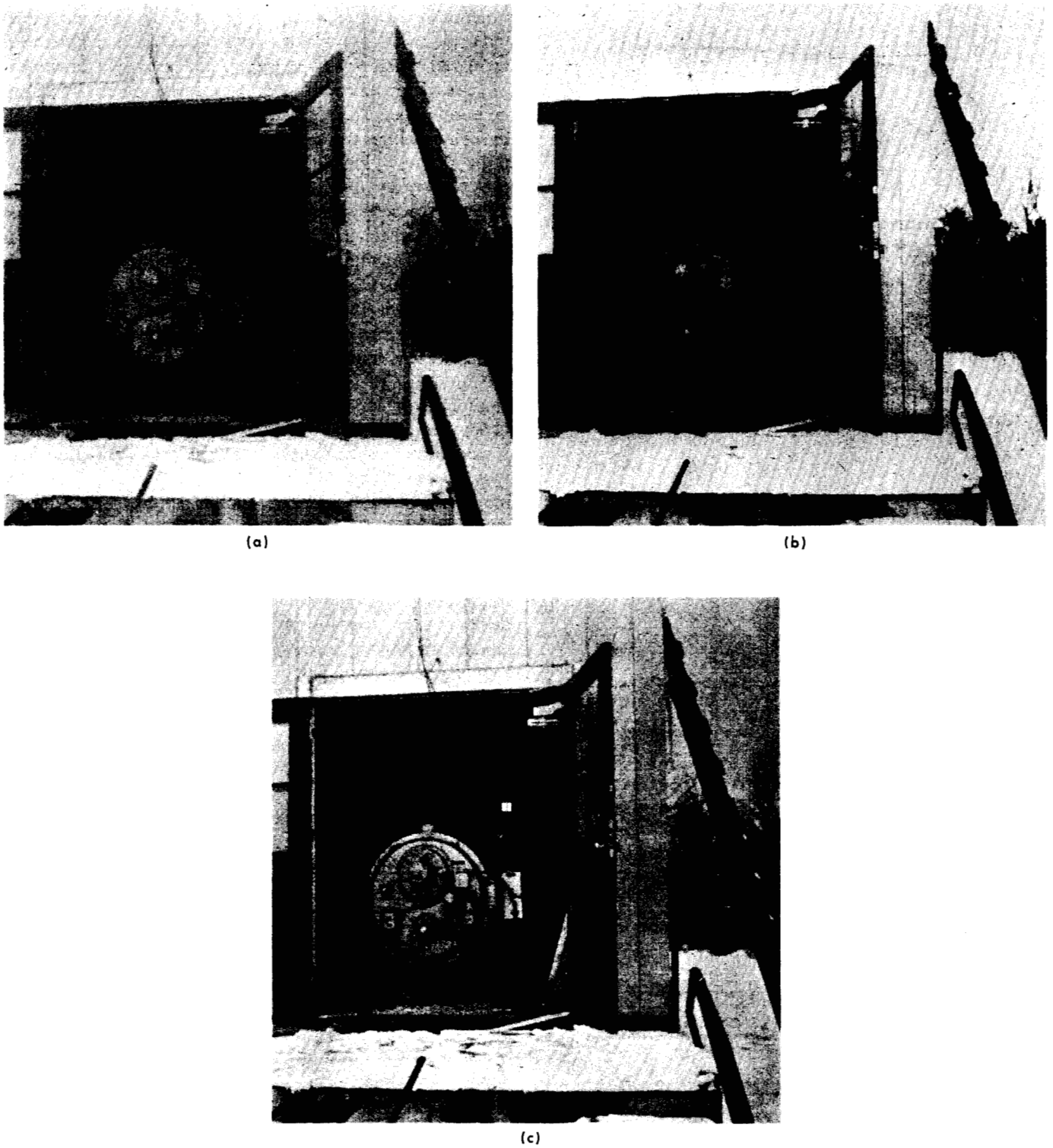


Fig. 25. The image of Fig. 21(a) processed using (a) $\gamma = 1/2$, (b) $\gamma = 2$, (c) a frequency-dependent γ with low-frequency attenuation.

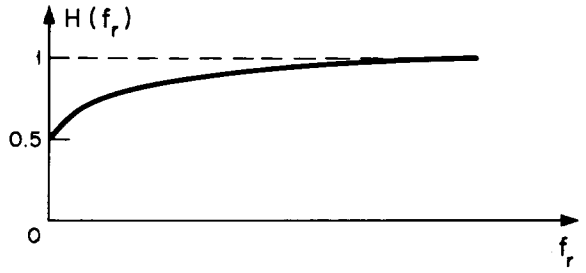


Fig. 26. The radial cross section of the multiplicative frequency response used to produce Fig. 25(c).

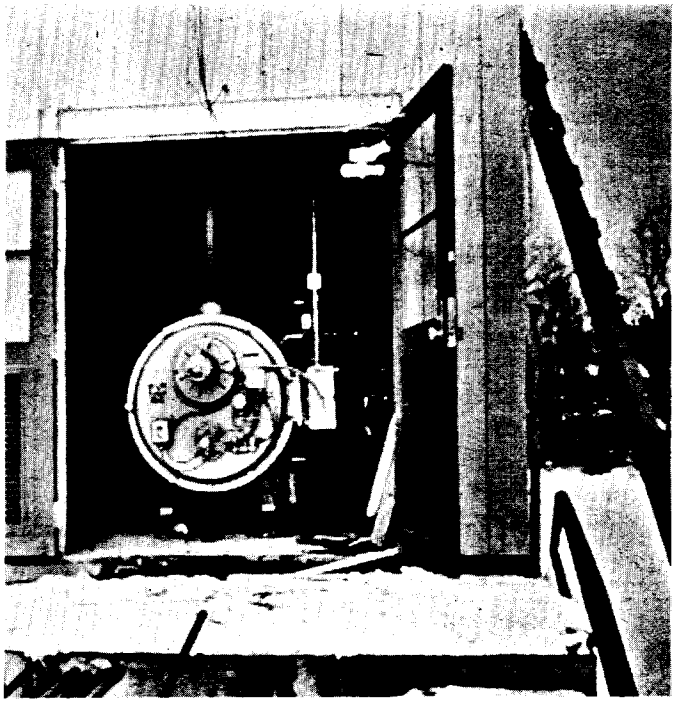


Fig. 27. The image of Fig. 21(a) processed using a frequency-dependent γ with low-frequency attenuation and high-frequency amplification.

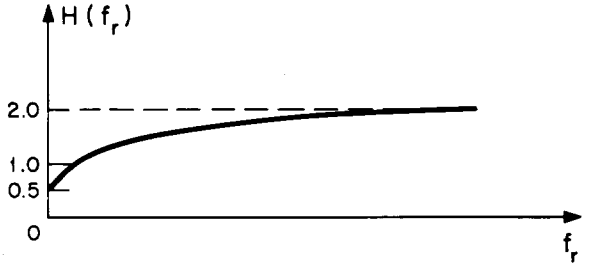


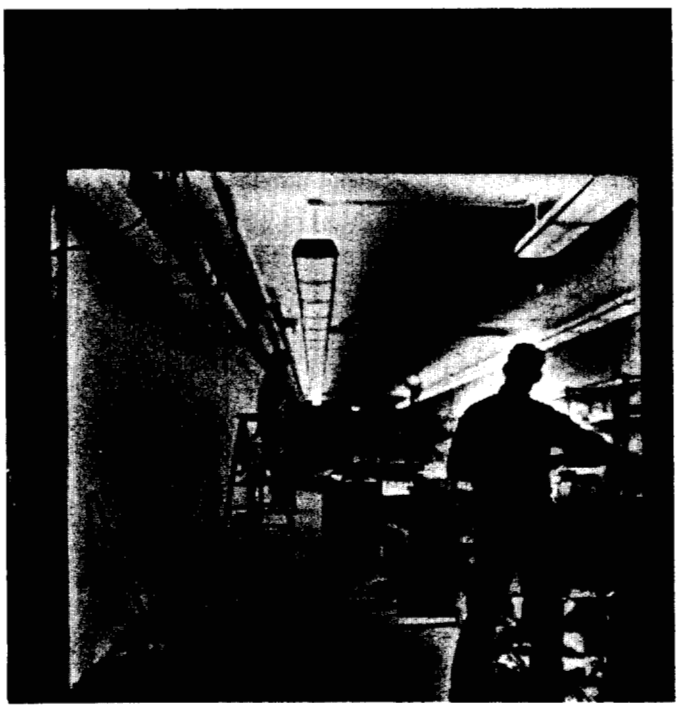
Fig. 28. The radial cross section of the multiplicative frequency response used to produce Fig. 27.

contains some low-frequency components. A communications engineer would say that there is cross talk between these two components of the original image. Thus, in the processing used to obtain the image of Fig. 27, some components of the illumination function have been increased and some components of the reflectance function have been decreased. Subjectively, these are not obtrusive. However, they are there and can be seen most easily, especially if they are pointed out. Specifically, in Fig. 27 the glow around the doorway making the building look whiter than it really is in the vicinity of the blackened room, the intense brightness level of the door dampener, and the white ring around the boiler-shaped object inside the room are typical artifacts of this type. Others can be seen in Fig. 29(a), (b), and (c), which bear the same relationship to Fig. 21(b), (c), and (d), respectively, as Fig. 27 does to Fig. 21(a). In spite of these approximations, the use of these methods in the processing of images has obvious practical implications wherever dynamic range is limited and the preservation of details is important. It has the additional advantage of obeying a law of superposition which facilitates analysis through classical techniques while operating according to the same rules of combination that form the original subject information. The degree to which the artifacts in Fig. 27 and Fig. 29(a), (b), and (c) are visible is a strong function of the exact nature in which the frequency response of Fig. 28 makes its transition from one-half at low frequencies to two at high frequencies. Fig. 30(a) and (b) were obtained from the image of Fig. 21(b) in the same manner as was Fig. 29(a), with the exception of the frequency response shape used. These frequency responses are shown in Fig. 31(a) and (b), respectively. Both frequency responses are characterized by a rather rapid transition between the high-frequency asymptote of two and the low-frequency asymptote of one-half. They differ, however, in that the transition occurs in the case of Fig. 31(a) at a relatively high frequency and in Fig. 31(b) at a relatively low frequency. The halo and flaring effects of Fig. 30(a) and (b) are more objectionable by far than those of Fig. 27. The frequency response of Fig. 28 is a compromise. That of Fig. 31(a) favors flare along large objects at the expense of halo around small ones. That of Fig. 31(b) favors halo around small ones at the expense of flare along large.

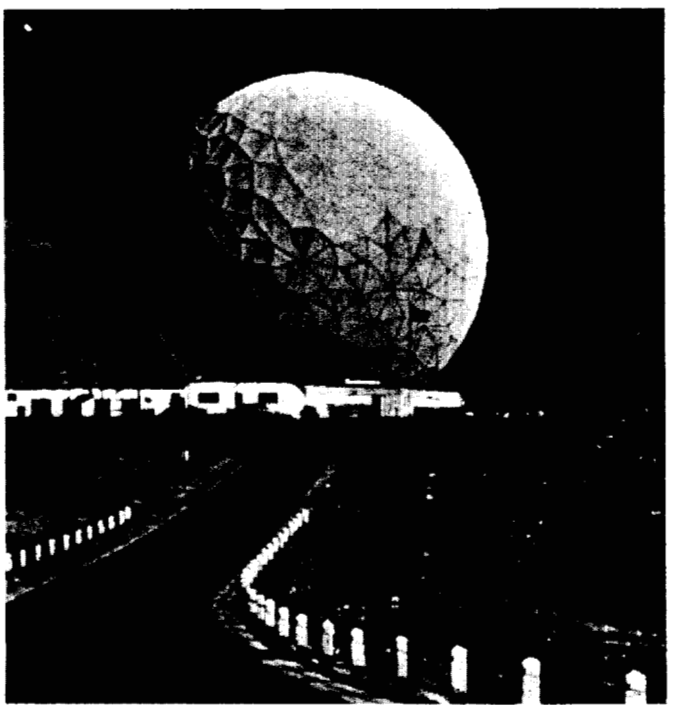
The compromise characteristics of the frequency response of Fig. 28 were arrived at through an attempt to find the frequency response which would treat large and small objects equally or nearly so. Since the two-dimensional Fourier spectrum of an object contracts as the object grows in size and spreads as the object shrinks in size, a frequency response characteristic which is somewhat invariant to changes in frequency scale would meet the objective. A characteristic possessing some invariance is the logarithmic frequency function. This invariance property is described by

$$\log Af = \log A + \log f. \tag{59}$$

If we think of the parameter A as a scale factor on an image of standard size, then we see that, except for an additive constant, the frequency response which an image en-



(a)



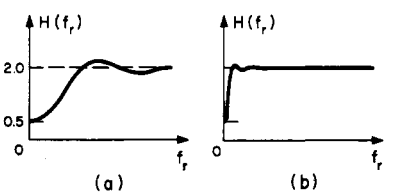
(b)

(c)

Fig. 29. The remaining images of Fig. 21 processed as in Fig. 27.

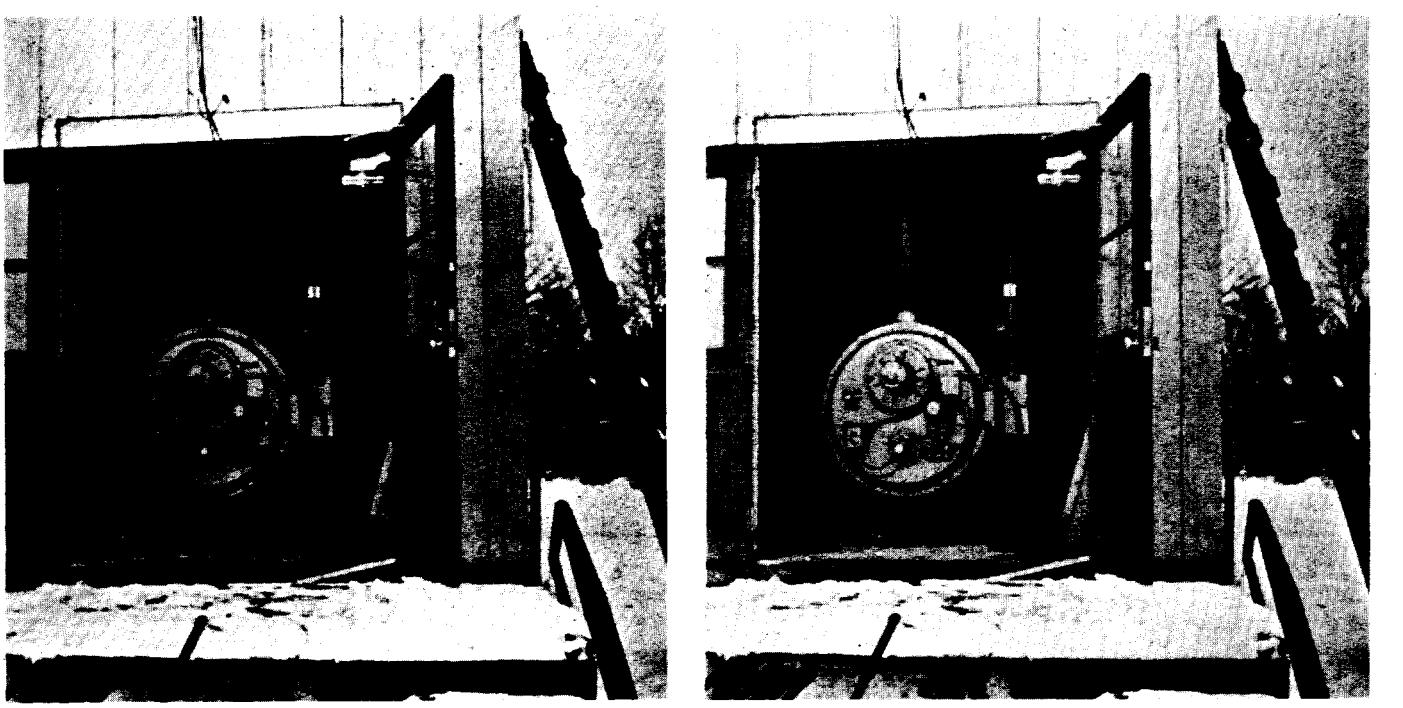
(a) (b)

Fig. 30. Images processed using abruptly changing frequency characteristics.



(a) (b)

Fig. 31. The multiplicative frequency responses used to produce the images of Fig. 30 from the image of Fig. 21(b).



(a) Biased for best appearance. (b) Average value of image restored.

Fig. 32. Two versions of the image of Fig. 21(a) processed as in Fig. 27 but using an ordinary linear filter rather than a multiplicative filter.

counters after magnification or reduction by the scale factor A is the same as that which it encounters at its standard size. The frequency response of Fig. 28 is such a logarithmic characteristic to a first approximation. Only for frequencies extremely close to zero is the logarithmic variation altered to provide an asymptote to the value of one-half rather than minus infinity.

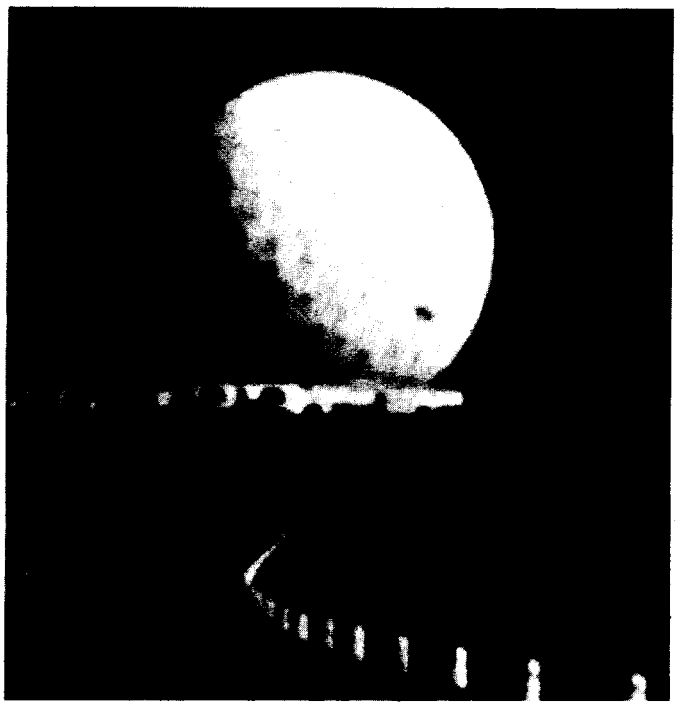
At this point it is quite reasonable to ask what effect ordinary linear filtering would have upon the images presented here and to draw a comparison in effectiveness. To this end we present Fig. 32(a) and (b). Fig. 32(a) represents the original scene of Fig. 21(a) processed using the frequency response of Fig. 28 in an ordinary linear filtering process and biased for best appearance. The results are instantly striking, but careful examination reveals some severe drawbacks. The most serious of these is the black halo around the inside of the doorway in which all detail is lost. While the visibility inside the room has been increased the improvement falls short of that obtained multiplicatively in Fig. 27. Finally, there are many places in the scene that are much darker than they should be. Most of these problems are associated with the fact that the linear processing employed produces negative brightness values in the processed image which due to the half-wave rectification of photographic processing appear as black. The addition of minimum bias sufficient to eliminate negative brightness results in a washed-out image of such poor quality that we do not show it here. Fig. 32(b) is a compromise in this respect in which bias sufficient to restore the average value of the original image brightness is used.

An effective approach to the problem of television bandwidth reduction [16], [17] centers around the idea of separating images into a relatively narrow-bandwidth low-frequency component and the complementary high-frequency component and preserving only a small fraction of the information in the high-frequency component in the form of edge contours. At the receiver, the edge contours are used in an attempt to re-create the high-frequency component. The results are combined with the low-frequency component to produce an approximation of the original image. Standard practice has been to consider the low- and high-frequency components of the image in the additive sense in keeping with the established traditions of signal processing. If these components are taken in the multiplicative sense, the results can be considerably enhanced. There are two reasons for this. When the components are taken in the additive sense, poorly illuminated edges may remain undetected during the process of bandwidth reduction. In the multiplicative case, all edges are represented equally by the high-frequency component since variations in illumination have been more or less separated out with the low-frequency component. Since any bandwidth reduction process will introduce errors in the image and practice indicates that these errors occur more or less uniformly throughout the picture, then in the additive case poorly illuminated portions of the image will be dominated by error and thus rendered totally useless. If the process is carried out multiplicatively, constant errors are made in terms of the logarithm of the picture and thus represent proportional errors in the final image. In this way, brightly illuminated areas and dimly illuminated areas are treated equivalently.

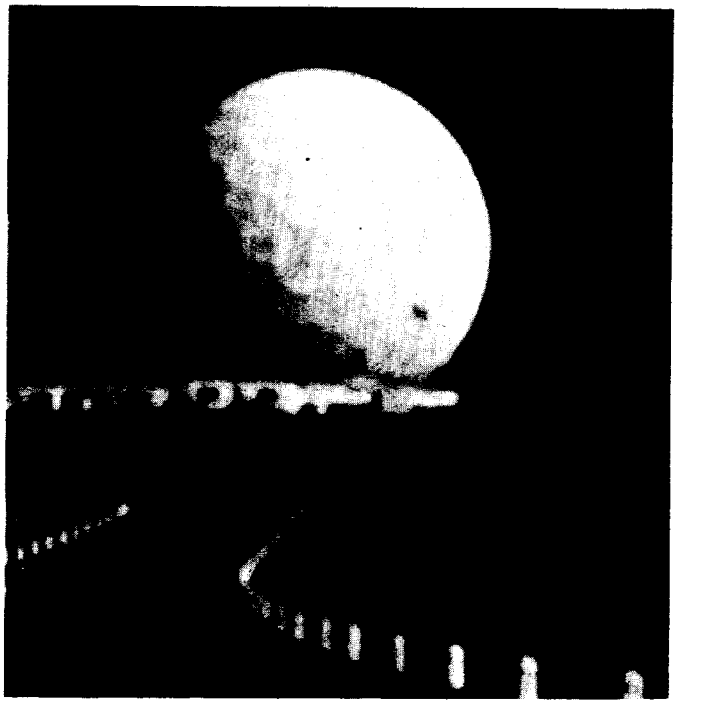
Fig. 33 represents an image as it appears in the various stages of bandwidth reduction when the process is carried out using both the traditional ideas of additive superposition and those of multiplicative superposition. The low-frequency images of Fig. 33(a) and (b) are subjectively nearly indistinguishable and, for all intents and purposes, appear much the same as a defocused photograph. The edge contours of Fig. 33(c) and (d) are markedly different, however. These images were produced by differentiating the original images in two dimensions and clipping the results into three levels as dictated by suitably adjusted thresholds. Notice that for linear processing there is less edge contour information in the regions corresponding to the dark areas of the original. The artificial-highs images of Fig. 33(e) and (f) which were produced directly from the edge contours by restoration filters can be compared similarly. The reduced-bandwidth recreated images of Fig. 33(g) and (h) differ much as would be expected from the statements made above. In the most brightly illuminated areas, the preservation of details is more faithfully carried out by the linear process. In the most dimly illuminated areas, however, the preservation of details is more faithfully carried out by the multiplicative process. Histograms of brightness for typical scenes are heavily skewed towards black. Similar histograms of the logarithm of these scenes reveal more or less rectangular distributions of log brightness placing equal weight on bright and dark areas. This fact further favors the use of the multiplicative scheme. Since the eye is sensitive more nearly to brightness ratios than to absolute brightness, it is not surprising that percentage errors are to be preferred on a subjective basis. A drawback to the multiplicative process is that the bandwidth-reduced edge contour image of Fig. 33(c) contains more information than its counterpart of Fig. 33(d) and thus, all else being equal, the use of the multiplicative scheme can result in smaller bandwidth reductions. This fact is suggested by the first-order entropies of the actual numerical samples of Fig 33(c) and (d) which are 1.077 and 0.945 bits per picture element, respectively.

The images presented here were all processed digitally by the TX-2 computer at the M.I.T. Lincoln Laboratory. The analog signal from a low-noise rotating scanner was fed to a twelve-bit analog-to-digital converter and the resulting numbers stored in the computer memory. Tests have shown that these numbers contain ten bits of significance. Each image was represented by a square array, 340 samples on a side. Before deposition in a permanent image library, the twelve-bit samples were converted to logarithms, the most significant nine bits of which were maintained.

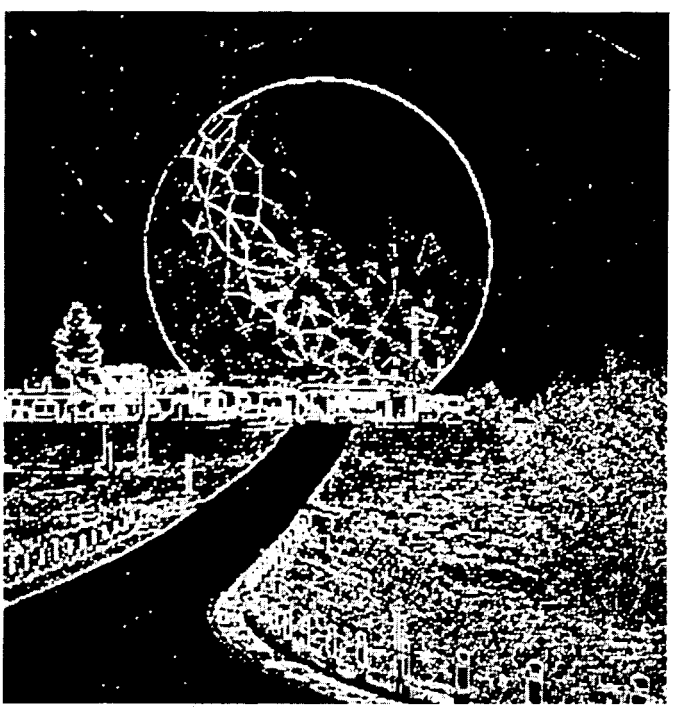
The linear processing was performed through the use of high-speed convolution methods [18] applied in two dimensions. The two-dimensional isotropic convolution kernels were determined through the Hankel transforms [17] of the defining frequency characteristics of Figs. 26, 28, and 31 truncated to possess nonzero values inside circles with diameters of about 80 picture elements. Each convolu-



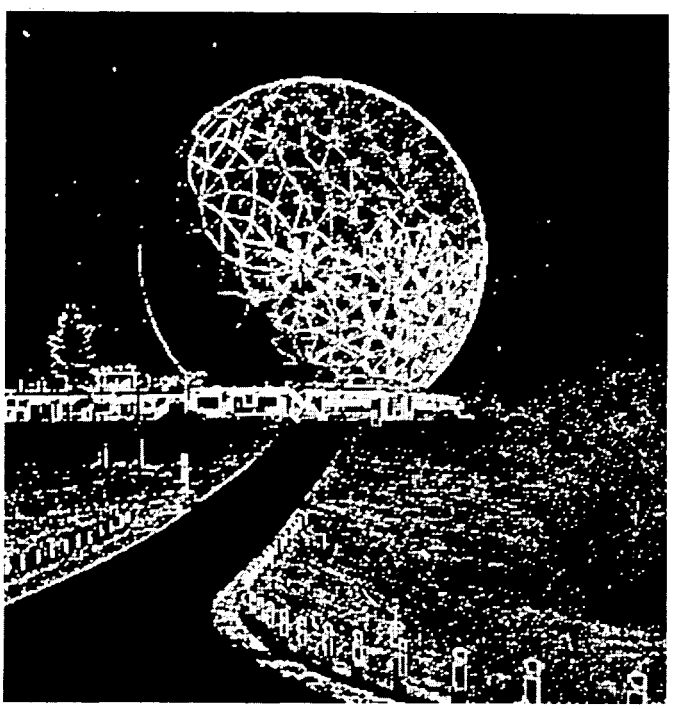
(a) Low-pass, multiplicative.



(b) Low-pass, linear.

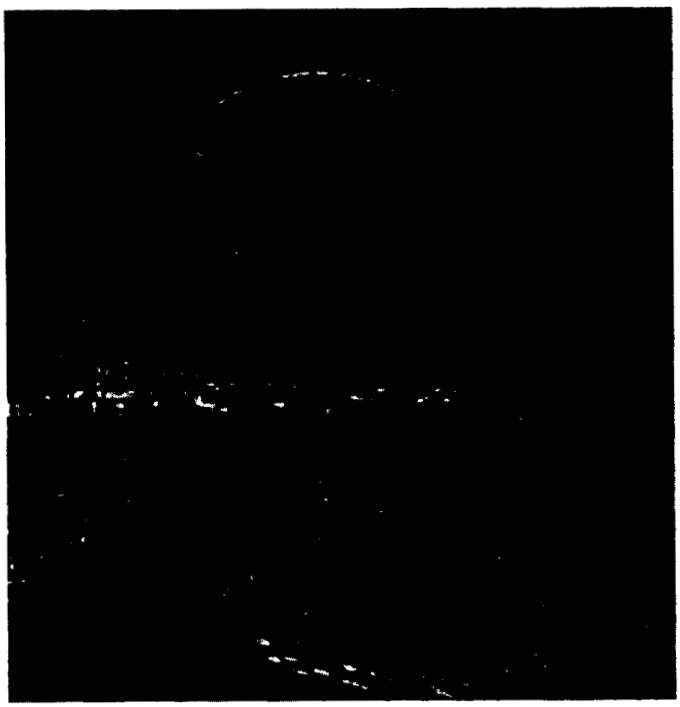


(c) Edge contours, multiplicative.

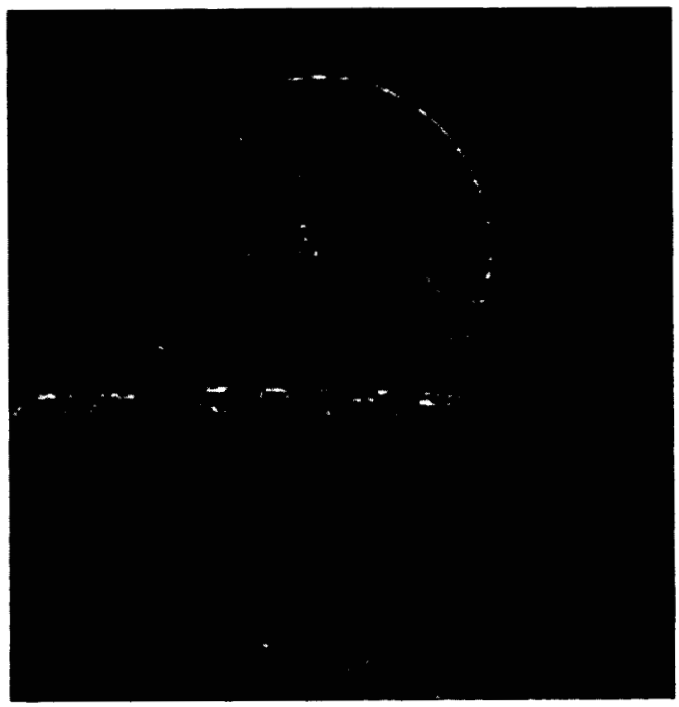


(d) Edge contours, linear.

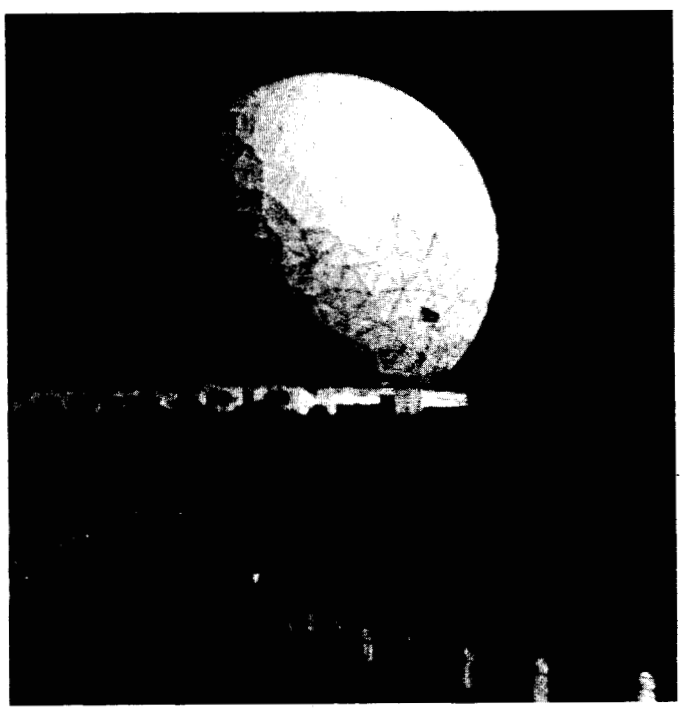
Fig. 33. The image of Fig. 21(c) in various stages of bandwidth compression.



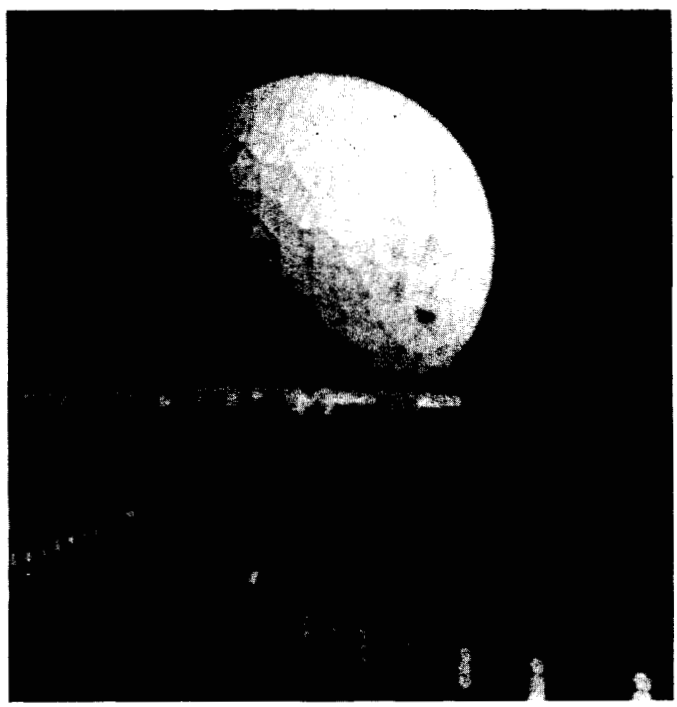
(e) Artificial highs, multiplicative.



(f) Artificial highs, linear.



(g) Recreated, multiplicative.



(h) Recreated, linear.

Fig. 33 (*Cont'd*).

tion required about 13 minutes.

Photographs of processed images were made by exponentiating the filtered image logarithms and controlling a vector-drawing computer graphics display [19] with the resulting eight-bit numerical values. The control was arranged to vary the velocity of scan while holding the cathode ray intensity at a standard constant level, thus avoiding the complications of beam and phosphor nonlinearity. The time to display an output image was between ten and fifteen seconds depending on average scene brightness, thus permitting real-time viewing in a darkened room as well as ordinary photographic recording through time exposure. For the latter, special digital compensation curves were used to straighten the nonlinear photographic characteristics of the films employed, thus resulting in much improved image quality.

Homomorphic Filtering of Echoed Signals [20], [21]

In many areas of application, signals are transmitted or recorded in a reverberant environment, i.e., one which introduces echoes. Reverberation arises, for example, in audio recording, in multipath communication, and in radar and sonar detection. In many cases we wish to remove the distortion represented by the echoes, or to recover the echo structure as a means of probing and characterizing the channel.

A simple model for the distortion introduced by reverberation is a convolution of the original waveform with a train of weighted samples, i.e.,

$$x(n) = s(n) \otimes p(n)$$

$$p(n) = \sum_{k=0}^{\infty} \alpha_k \delta(n - n_k)$$

where $s(n)$ and $x(n)$ are the original and distorted waveforms, respectively. The analysis presented previously suggests that in applying the notion of homomorphic deconvolution to separating the echo pattern and the original waveform, we determine the complex logarithm of the z-transform of $x(n)$, the distorted waveform, and then look for a property of each of the components that permits their separation by means of linear filtering. To help focus the approach let us first consider the case of a simple echo, i.e.,

$$p(n) = \delta(n) + \alpha\delta(n - n_0)$$

so that

$$x(n) = s(n) \otimes [\delta(n) + \alpha\delta(n - n_0)]. \tag{60}$$

The z-transform $X(z)$ evaluated on the unit circle is

$$X(e^{j\omega}) = S(e^{j\omega})[1 + \alpha e^{-j\omega n_0}].$$

We observe that the contribution due to the echo is a periodic function of ω with period $2\pi/n_0$. Furthermore, its repetition rate increases as n_0 increases so that longer echo times are manifested by more rapid fluctuations in the spectrum. Since the logarithm of a periodic waveform remains periodic with the same repetition rate, the echo is represented in the log spectrum as an *additive* periodic component.

The character of the log spectrum of $p(n)$ suggests the possibility that we may separate the contributions of $s(n)$ and $p(n)$ by removing the variations in the log spectrum which occur at repetition rates which are multiples of $2\pi/n_0$. Thus the linear filtering would convolve the complex log spectrum with a kernel designed to remove the periodic components.

Since convolution in frequency corresponds to a multiplication in time, we may view this linear filtering as a multiplication of the complex cepstrum by a fixed weighting. Specifically, we observe that periodic variations in the log spectrum contribute to the complex cepstrum only at values of n which are multiples of the echo time n_0, in precisely the same way that a periodic time function with period T has spectral components at only those frequencies which are multiples of $1/T$. Then the "comb" filtering suggested above will correspond to multiplying the complex cepstrum by a weighting which is unity except in regions that are multiples of the echo time n_0, in which case the weighting is zero. This class of filters is depicted in Fig. 34. Clearly the notion of comb filtering can only be successful if we have approximate information about the echo time. If we do not have this information, and if the complex cepstrum of $s(n)$ is concentrated near $n=0$, then we can replace the idea of comb filtering with that of "low-time" filtering, that is, weighting the complex cepstrum by unity near the origin and zero otherwise. In terms of the log spectrum this filtering corresponds to associating the slow variations with $s(n)$ and the rapid variations with the echo pattern $p(n)$. An alternative is to use an adaptive procedure whereby the parameters of a comb filter are based on a measurement of the echo time. Such a measurement can be based on the fact that a peak is expected to occur in the complex cepstrum at multiples of the echo time. If we return to the more general case of multiple echoes, we recognize that as long as the echoes are equally spaced so that

$$p(n) = \sum_{k=0}^{\infty} \alpha_k \delta(n - kn_0)$$

the approach is essentially identical to the case of a single echo. If the echoes are not equally spaced the situation becomes more complex since we can no longer localize the effect of the echo pattern in the complex cepstrum.

In principle the approach taken to echo removal by means of homomorphic filtering is based on the reasoning presented above. However, the above discussion assumes that we have available the entire waveform $x(n)$ and are able to compute its Fourier transform. The more typical situation in practice is that the waveform $s(n)$ is indefinite in duration and consequently it is impractical to compute the Fourier transform of the entire waveform. In addition, there are situations in which the reverberation times vary slowly. Thus we are led to considering echo removal in which a short-time analysis of the waveform is more appropriate. In this case the waveform is processed in pieces and the results fitted together to obtain the output. To illustrate how this can be done, let us again consider a simple echo as in (60) and for which $s(n)$ is a waveform of indefinite dura-

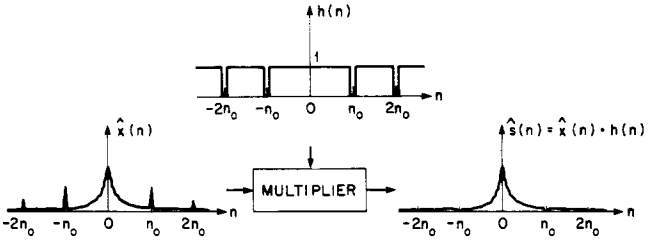


Fig. 34. A linear time-varying "comb" filter for removing the component in the complex cepstrum due to a simple echo.

tion. The situation is depicted in Fig. 35(a) where $s(n)$ is represented by the solid curve and $\alpha s(n-n_0)$ is represented by the dotted curve. Let us consider segments of $x(n)$ consisting of L samples. To facilitate our discussion, we define for $\xi = 0, L, 2L, \cdots$,

$$\begin{aligned} x(\xi, n) &= x(\xi + n) \quad 0 \le n < L \\ &= 0 \qquad\qquad \text{otherwise.} \end{aligned}$$

From Fig. 35(a) we see that, in general, a particular segment of the input can be expressed in the form

$$x(\xi, n) = s(\xi, n) + \alpha s(\xi, n - n_0) + e(\xi, n)$$

where

$$\begin{aligned} s(\xi, n) &\neq 0 \qquad\qquad 0 \le n < L \\ &= 0 \qquad\qquad \text{otherwise} \end{aligned}$$

and

$$\begin{aligned} e(\xi, n) &\neq 0 \quad 0 \le n < n_0 \quad \text{and} \quad L \le n < L + n_0 \\ &= 0 \qquad\qquad \text{otherwise.} \end{aligned}$$

The term $e(\xi, n)$ accounts for the overlap of the echo from the previous segment at the beginning of the segment and the overlap into the next segment at the end of the segment. Since it is the amount by which $s(\xi, n)$ fails to have the desired form, $e(\xi, n)$ is referred to as the error in the segment.

The nature of the errors is depicted in Fig. 35(b) for three consecutive segments of the input of Fig. 35(a). It can be seen that the error at the end of a segment is the negative of the error at the beginning of the next segment.

If we take the z-transform of a segment of the input, we obtain

$$\begin{aligned} X(\xi, z) &= \sum_{n=0}^{L-1} x(\xi, n) z^{-n} \\ &= S(\xi, z)(1 + \alpha z^{-n_0}) + E(\xi, z) \qquad (61) \\ &= \left[S(\xi, z) + \frac{E(\xi, z)}{1 + \alpha z^{-n_0}} \right] (1 + \alpha z^{-n_0}) \end{aligned}$$

where $S(\xi, z)$ is given by

$$S(\xi, z) = \sum_{n=0}^{L-1} s(\xi, n) z^{-n}.$$

We note that $X(\xi, z)$ is not simply a product as it would be if the entire waveforms were transformed. It is true, however, that if

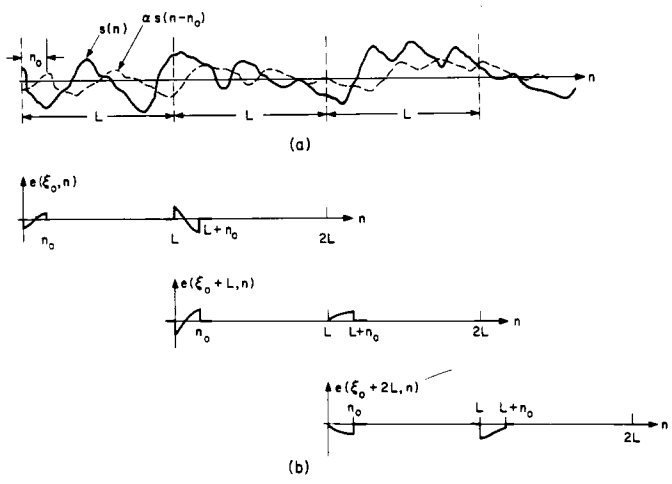


Fig. 35. (a) A signal $s(n)$ (solid line) and delayed scaled replica (dashed line). Each of the depicted segments L is to be represented separately as the sum of a component signal, an echo of this signal, and an error. (b) The error term in this representation depicted for each segment.

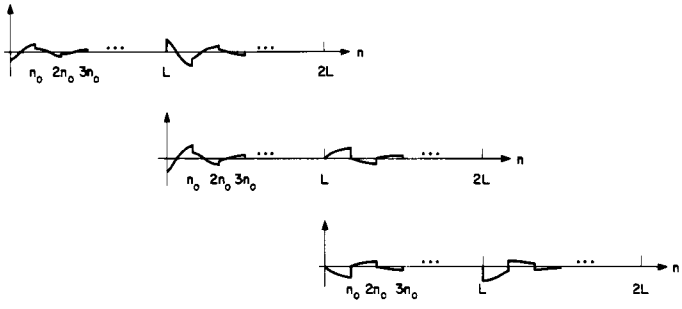


Fig. 36. The output due to the error terms of Fig. 35(b). Successive lines represent the response due to the error term for adjacent segments of length L in Fig. 35(a). Note that the error at the end of one segment is the negative of the error at the beginning of the next segment.

$$L \gg n_0$$

then there will still be impulses in the complex cepstrum at $n_0, 2n_0, \cdots$. Removing these impulses is equivalent to removing the factor $(1+\alpha z^{-n_0})$ in (61) so that operating with the inverse characteristic system on the complex cepstrum with the impulses removed yields an output whose z-transform is

$$Y(\xi, z) = S(\xi, z) + \frac{E(\xi, z)}{1 + \alpha z^{-n_0}}.$$

If $|\alpha| < 1$, then the corresponding output sequence is

$$y(\xi, n) = s(\xi, n) + \sum_{k=0}^{\infty} (-\alpha)^n e(\xi, n - kn_0).$$

Thus the output consists of the desired output segment $s(\xi, n)$ plus an error term. This error term is effectively the error in the input segment passed through a linear system whose system function is the reciprocal of the z-transform of the impulse train which represents the echoes.

The error in the output for the three consecutive segments of Fig. 35 is depicted in Fig. 36. The figure suggests

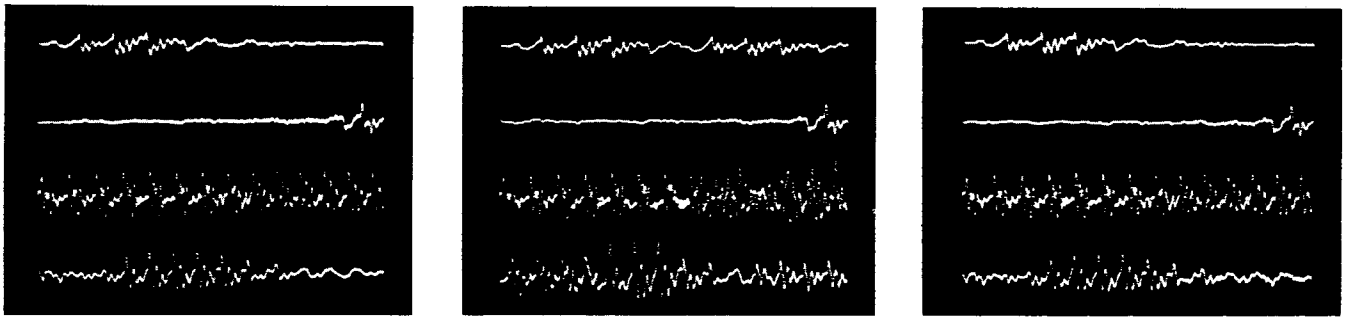


Fig. 37. An example of homomorphic echo removal. (a) 410 ms of speech sampled at 10 kHz with the four traces from top to bottom representing contiguous segments of 102.5 ms. (b) The speech sample of (a) with a 50-ms echo. (c) The speech sample of (b) processed to remove the echo.

how the output segments can be put together. If we simply add the appropriately delayed output segments, the result will be the desired output signal $s(n)$. An alternate procedure is suggested by the fact that if $L \gg n_0$, the error in the output decays as n gets large. Thus there may be large portions of each output segment which are relatively free of error. In this case we can process overlapping segments of length L, and save only part of each output segment. If we choose segments such that the portions saved are contiguous parts of the waveform, these pieces can simply be placed end-to-end to produce the output.

This approach to echo removal has been carried out on the TX-2 computer at the M.I.T. Lincoln Laboratory using speech as the original waveform with echoes artificially introduced. Informal listening tests indicate that echoes can be removed to the extent that they are inaudible with only minor degradation of the speech. An illustration of typical waveforms obtained is shown in Fig. 37. Fig. 37(a) represents 410 ms of speech sampled at 10 kHz with the four traces from top to bottom representing contiguous segments of 102.5 ms. This waveform with a 50-ms echo is shown in Fig. 37(b). The result of carrying out the processing which has been described above is shown in Fig. 37(c), indicating that the echo has, to a large extent, been removed.

Deconvolution of Speech [22], [23]

During voiced sounds such as vowels, the speech waveform may be considered as the result of periodic puffs of air released by the vocal cords exciting an acoustic cavity, the vocal tract [4]. Thus a simple and often useful model of the speech waveform consists of the convolution of three components, representing pitch, the shape of the vocal cord or glottal excitation, and the configuration of the vocal tract. Many systems for compressing the bandwidth of speech and for carrying out automatic speech recognition have as the basic strategy the separate isolation and characterization of the vocal tract excitation and the vocal tract impulse response. Thus many speech processing systems are directed in part toward carrying out a deconvolution of the speech waveform.

As in the previous example of echo removal, the speech waveform is a continuing signal and therefore must be processed on a short-time basis. Thus we consider a portion $s(t)$ of the speech waveform as viewed through a time-limited window $w(t)$. Although the vocal tract configuration changes with time we will choose the duration of the window to be sufficiently short so that we can assume that, over this duration, the shapes of the vocal tract impulse response and the glottal pulse are constant. Then, if we denote by $p(t)$ a train of ideal impulses whose timing corresponds to the occurrence of the pulses released by the vocal cords, by $g(t)$ the shape of the glottal pulse, and by $v(t)$ the impulse response of the vocal cavity, we express $s(t)$ approximately as

$$s(t) = [p(t) \otimes g(t) \otimes v(t)]w(t). \tag{62}$$

Furthermore, if $w(t)$ is smooth over the effective duration of the glottal pulse and the vocal tract impulse response, then we can approximate (62) as

$$s(t) = [p(t)w(t)] \otimes g(t) \otimes v(t). \tag{63}$$

Thus, if a smooth window is used to weight the speech waveform we can consider the weighted aperiodic function as a convolution of weighted pitch, glottal pulse, and vocal tract impulse response.

In keeping with the previous discussion we wish to phrase our remarks in terms of samples of $s(t)$, which we denote by $s(n)$. Assuming that we can replace the continuous convolution of (63) by a discrete convolution of samples of each of the component terms, we write that

$$s(n) = [p(n)w(n)] \otimes g(n) \otimes v(n)$$

or

$$s(n) = p_1(n) \otimes g(n) \otimes v(n) \tag{64}$$

where $w(n)$, $g(n)$, and $v(n)$ are samples of $w(t)$, $g(t)$, and $v(t)$, respectively, and $p_1(n)$ is a train of unit samples weighted with the window $w(n)$.

The vocal tract impulse response $v(n)$ is often modeled as the response of a cascade of damped resonators so that its z-transform is

$$V(z) = \frac{K}{\prod_{i=1}^{M}(1 - a_i z^{-1})(1 - a_i^* z^{-1})} \qquad |a_i| < 1.$$

In this case, $v(n)$ is minimum phase, and it follows from (31) that $\hat{v}(n)$, the complex cepstrum of $v(n)$, is of the form

$$\hat{v}(n) = \begin{cases} \sum_{i=1}^{M} \frac{|a_i|^n}{n} \cos \omega_i n & n > 0 \\ 0 & n < 0 \end{cases}$$

where

$$a_i = |a_i| e^{j\omega_i}.$$

Thus, $\hat{v}(n)$ decays as $1/n$ and therefore tends to have its major contribution near the origin for $n>0$.

An accurate analytical representation of the glottal pulse $g(n)$ is not known and consequently it is difficult to make any specific statements regarding the characteristics of its complex cepstrum $\hat{g}(n)$. However, we can expect in general that $g(n)$ is nonminimum phase [24]. Expressing $g(n)$ as the convolution of a minimum phase sequence $g_1(n)$ and a maximum phase sequence $g_2(n)$, we will assume that $\hat{g}_1(n)$, which is zero for $n < 0$, and $\hat{g}_2(n)$, which is zero for $n > 0$, both have an effective duration which is less than a pitch period.

The complex cepstrum of the train of weighted unit samples representing pitch is, as we have seen previously, a train of weighted unit samples with the same spacing. Thus, we can diagrammatically represent the complex cepstrum as in Fig. 38.

The components of $s(n)$ due to pitch and to the combined effects of vocal tract and glottal pulse tend to provide their primary contributions in non-overlapping time intervals. The degree of separation will of course depend to some extent on the pitch, with more separation for low-pitched male voices than for high-pitched female voices. Experience has indicated, however, that except in cases of very high pitch a good separation of these components occurs. To illustrate, consider the example of Fig. 39. Fig. 39(a) shows a portion of the vowel "ah" as in "father," with a male speaker, and Fig. 39(b) shows the complex cepstrum. Based on the previous discussion, we can recover the term $p_1(n)$ of (64) by multiplying the complex cepstrum by zero in the vicinity of the origin (with a time-width of, say, 8 ms) and by unity elsewhere. Alternatively, to recover $v(n) \otimes g(n)$ we would multiply the complex cepstrum by unity in the vicinity of the origin and by zero elsewhere. After this weighting, the result is transformed by means of the system D^{-1}. In Fig. 39(c) is shown the result of attempting to recover the weighted train of pitch pulses $p_1(n)$. Pulses with the correct spacing are clearly evident.[7] In Fig. 39(d) the result of retaining only the low-time portion of $\hat{s}(n)$, corresponding to attempting to recover $[v(n) \otimes g(n)]$, is shown. To verify that the pulse of Fig. 39(d) can be considered as a convolutional speech component the speech was resynthesized by convolving this pulse with a train of unit samples whose spacing was chosen to be a pitch period as measured from the waveform of Fig. 39(c). The resynthesized speech is shown in Fig. 39(e) and should be compared with the original speech of Fig. 39(a).

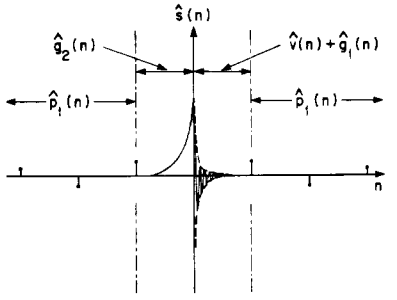


Fig. 38. An illustration of the characteristics of the complex cepstrum for speech.

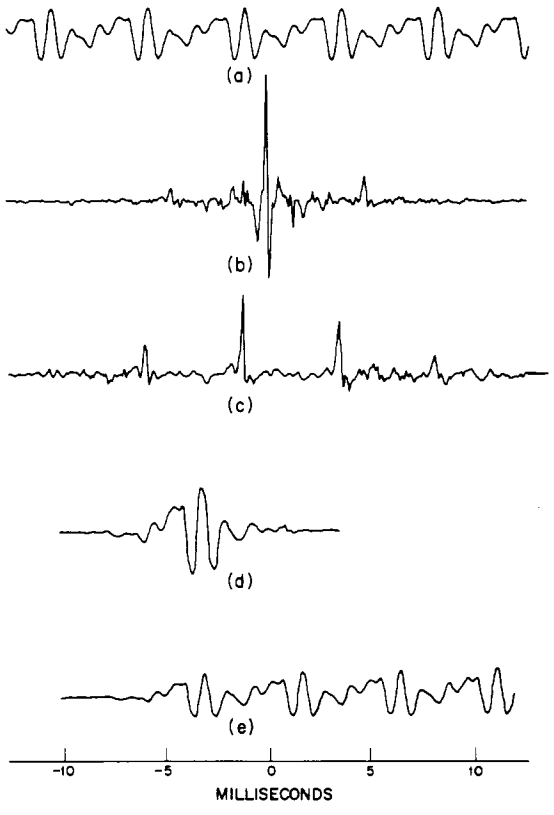


Fig. 39. An example of the deconvolution of speech. (a) Original sample of the vowel "ah" for a male speaker. (b) Complex cepstrum of the sample of (a). (c) Recovered pitch pulses $p_1(n)$. The Hanning weighting applied to the original speech should be reflected in this output. (d) Recovered impulse response function reflecting the combined effects of glottal pulse and vocal tract impulse response. (e) Resynthesized speech using the impulse response of (d) and pitch as measured from (c).

From the diagram of Fig. 38 it is clear that we could not expect to separate the glottal pulse $g(n)$ and vocal tract impulse response $v(n)$ by simple weighting of the complex cepstrum, although we might expect to recover the maximum phase part of the glottal pulse. This has been tried in a few cases to verify the idea. An example of the type of pulse obtained is shown in Fig. 40. However, the value of recovering only the maximum phase portion of the glottal pulse is not clear.

[7] Pitch detection based directly on a measurement of the location of a peak in the cepstrum (as defined by Bogert *et al.* [6]) has been successfully demonstrated by Noll [25].

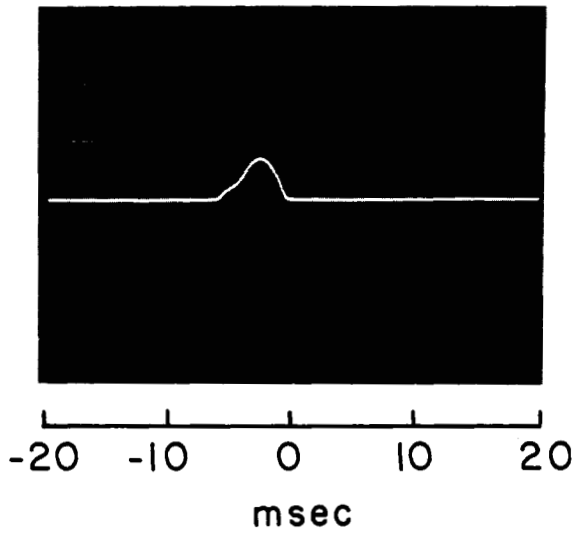


Fig. 40. Pulse obtained by retaining only those values in the complex cepstrum near the origin for $n<0$.

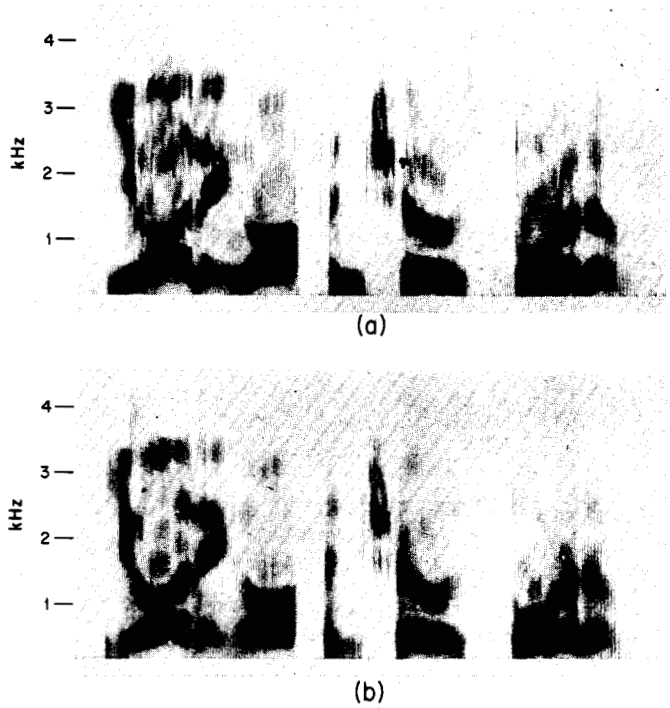


Fig. 41. (a) Spectrogram of original speech. The sentence is "yawning often shows boredom." (b) Spectrogram of synthesized speech.

To explore the feasibility of these ideas, a speech analysis-synthesis system based on homomorphic processing has been under investigation. In the analysis the cepstrum is obtained by weighting the input speech with a Hanning window 40 ms in duration. The cepstrum is separated into its high- and low-time parts with the cutoff time presently taken to be 3.2 ms. A decision as to whether the 40-ms sample is voiced or unvoiced and a measurement of the pitch frequency if voiced is made from the high-time portion. Thus, the synthesizer receives the low-time cepstral values, a voiced-unvoiced decision, and a pitch frequency measurement updated at 10-ms intervals. In the synthesizer, the pitch and voiced-unvoiced information is converted to an excitation function consisting of impulses during voicing and noise during unvoicing. The low-time cepstral values are converted to an impulse response function which is then convolved with the excitation function to form the synthetic speech. Informal listening tests indicate that the synthetic speech is of high quality and natural sounding. In Fig. 41 are shown spectrograms of a sentence before and after processing.[8]

VI. Conclusions

The applications which we have presented represent an attempt to apply the point of view provided by the theory of homomorphic systems to problems of general practical interest. The audio compression-expansion system has been realized economically in analog hardware and its success has led to its use in equipment in which dynamic range compression was required. As we have already stated, the other applications at present have been simulated on general-purpose digital computers.

The deconvolution of speech is being pursued as an approach to obtaining a high-quality speech bandwidth compression system which can be implemented in digital hardware with present technology. The homomorphic processing of images has immediate practical possibilities for non-real-time applications in which a small computer with a graphics facility is available. The removal of echoes by homomorphic deconvolution appears applicable to problems in which processing can be carried out on a large digital computer.

The full potential of these and other applications relies to a large extent on the advances which we can expect in system implementation. In particular, large-scale integration (both analog and digital) most certainly will play a vital role in providing an efficient and inexpensive realization of real-time processing of the kind which we have been discussing.

It is natural to ask if there are other areas of application for homomorphic filtering. In this respect we offer the following topics which we feel deserve consideration.

Applications of multiplicative filtering which may have potential are compensators for channel fading, systems for simultaneous amplitude and phase modulation and detection, automatic gain controls for other than audio application, ac and dc power regulators, and radar signal processing.

Some problems in convolutional filtering which may be promising involve the restoration of images blurred in an unknown manner [26], the suppression of multipath distortion, the enhancement of sonar signals, seismographic exploration, the sharpening of bioelectric signals blurred by propagation through tissue, the analysis of probability density functions, the measurement of auditorium acoustics,

[8] The speech analysis-synthesis system was simulated on the M.I.T. Lincoln Laboratory 1219 speech facility. Other parts of the work described in this section were carried out on the M.I.T. Lincoln Laboratory TX-2 computer and the PDP-1 computer facility operated by the Department of Electrical Engineering and the Research Laboratory of Electronics at M.I.T.

and the separation of antenna pattern and target impulse response in radar and sonar detection.

During the course of the research which resulted in this paper, the thinking of the authors was consistently influenced by some speculative ideas involving vision and hearing. While these ideas are presently being studied and no specific conclusions have been reached, we feel that the paper would be incomplete without mentioning them.

The presence of a logarithmic response in vision and hearing has been accepted for some time. Even more readily evident, and mechanized through the process of neural interaction, is the means for linear filtering [4], [27], [28]. This combination is so suggestive of both forms of homomorphic filtering which we have been discussing that questions concerning a possible relationship arose early during the research. Specifically, it seems reasonable to inquire whether to some approximation the processes of vision and hearing can be modeled as homomorphic systems directed toward a separation of multiplied components in the case of vision and a separation of convolved components in the case of hearing.[9] While the question as to whether a variety of psychophysical data can be modeled in these terms is purely speculative at present, some preliminary investigations have been encouraging. If it can in fact be verified that such models are reasonable, the resulting point of view may have a bearing on the design of communication systems for which the final receiver is the human eye or the human ear.

References

[1] A. V. Oppenheim, "Superposition in a class of non-linear systems," Research Lab. of Electronics, M.I.T., Cambridge, Mass., Tech. Rept. 432, March 31, 1965.

[2] ——, "Optimum homomorphic filters," Research Lab. of Electronics, M.I.T., Cambridge, Mass., Quart. Progr. Rept. 77, pp. 248–260, April 15, 1965.

[3] References [1] and [2] are summarized in A. V. Oppenheim, "Generalized superposition," *Information and Control*, vol. 11, pp. 528–536, November 1967.

[4] J. L. Flanagan, *Speech Analysis, Synthesis and Perception*. New York: Academic Press, 1965.

[5] R. L. Miller, "Nature of the vocal cord wave," *J. Acoust. Soc. Am.*, vol. 31, pp. 667–677, June 1959.

[6] B. Bogert, M. Healy, and J. Tukey, "The quefrency alanysis of time series for echoes," in *Proc. Symp. on Time Series Analysis*, M. Rosenblatt, Ed. New York, Wiley, 1963, ch. 15, pp. 209–243.

[7] A. V. Oppenheim, "Non-linear filtering of convolved signals," Research Lab. of Electronics, M.I.T., Cambridge, Mass., Quart. Progr. Rept. 80, pp. 168–175, January 15, 1966.

[8] E. A. Guillemin, *Theory of Linear Physical Systems*. New York: Wiley, 1963, ch. 18.

[9] J. W. Cooley and J. W. Tukey, "An algorithm for the machine calculation of complex Fourier series," *Math. of Comput.*, vol. 19, pp. 297–301, April 1965.

[10] T. G. Stockham, Jr., "The application of generalized linearity to automatic gain control," *IEEE Trans. Audio and Electroacoustics*, vol. AU-16, pp. 267–270, June 1968.

[11] C. W. Carter, Jr., A. C. Dickieson, and D. Mitchell, "Application of companders to telephone circuits," *Trans. AIEE*, vol. 65, pp. 1079–1086, December 1946.

[12] M. Medress, "Noise analysis of a homomorphic automatic volume control system," S.M. thesis, Dept. of Elec. Engrg., M.I.T., Cambridge, Mass., January 1968.

[13] C. B. Neblette, *Photography—Its Materials and Processes*. New York: Van Nostrand, 1962, chs. 20, 21, and 22.

[14] *Masking Color Transparencies*, Kodak Graphic Arts Data Book. Rochester, N. Y.: Eastman Kodak Co., 1960.

[15] E. G. St. John and D. R. Craig, "Logetronography," *Am. J. Roentgenol., Radium Therapy Nucl. Med.*, vol. 75, July 1957.

[16] W. F. Schreiber, "The mathematical foundation of the synthetic highs system," Research Lab. of Electronics, M.I.T., Cambridge, Mass., Quart. Progr. Rept. 68, p. 140, January 1963.

[17] D. N. Graham, "Image transmission by two-dimensional contour coding," *Proc. IEEE*, vol. 55, pp. 336–346, March 1967.

[18] T. G. Stockham, Jr., "High-speed convolution and correlation," *Spring Joint Computer Conf., AFIPS Proc.*, vol. 28, pp. 229–233, 1966.

[19] L. G. Roberts, "Conic display generator using multiplying digital-analog converters," *IEEE Trans. Electronic Computers (Short Notes)*, vol. EC-16, pp. 369–370, June 1967.

[20] R. W. Schafer, "Echo removal by generalized linear filtering," *NEREM Record*, pp. 118–119, 1967.

[21] ——, "Echo removal by discrete generalized linear filtering," Ph.D. thesis, Dept. of Elec. Engrg., M.I.T., Cambridge, Mass., February 1968.

[22] A. V. Oppenheim, "Deconvolution of speech" (abstract), *J. Acoust. Soc. Am.*, vol. 41, p. 1595, 1967.

[23] A. V. Oppenheim and R. W. Schafer, "Homomorphic analysis of speech," *IEEE Trans. Audio and Electroacoustics*, vol. AU-16, pp. 221–226, June 1968.

[24] M. V. Mathews, J. E. Miller, and E. E. David., Jr., "Pitch synchronous analysis of voiced sounds," *J. Acoust. Soc. Am.*, vol. 33, pp. 179–186, February 1961.

[25] A. M. Noll, "Cepstrum pitch determination," *J. Acoust. Soc. Am.*, vol. 41, pp. 293–309, February 1967.

[26] C. M. Rader (private communication).

[27] S. S. Stevens, *Handbook of Experimental Psychology*. New York: Wiley, 1951, chs. 23–27.

[28] F. Ratcliff, *Mach Bands: Quantitative Studies on Neural Networks in the Retina*. San Francisco: Holden-Day, 1965.

[29] R. B. Marimont, "Linearity and the Mach phenomenon," *J. Opt. Soc. Am.*, vol. 53, pp. 400–401, March 1963.

[30] ——, "Model for visual response to contrast," *J. Opt. Soc. Am.*, vol. 52, pp. 800–806, July 1962.

[31] W. H. Huggins, "A phase principle for complex-frequency analysis and its implications in auditory theory," *J. Acoust. Soc. Am.*, vol. 24, pp. 582–589, November 1952.

[9] Although from a different point of view and with a different motivation, ideas suggestive of this have appeared before. Marimont [29], [30] suggests a model for vision similar to the canonic form of Fig. 2. With regard to hearing, Huggins [31] discusses the notion that hearing is a process of deconvolution, although the mechanism which he proposes is different from that discussed here.

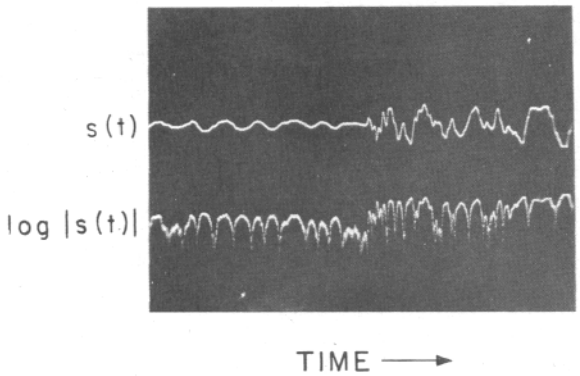


Fig. 16. A typical $s(t)$ and log $|s(t)|$ as measured in the laboratory.

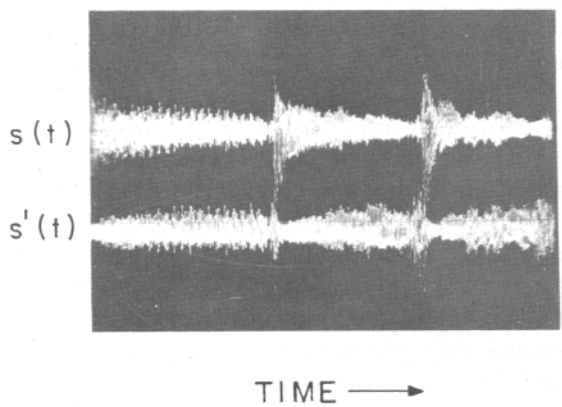


Fig. 17. A typical $s(t)$ and its supercompressed counterpart.

Fig. 18. A compression-expansion system employing a pair of complementary multiplicative filters.

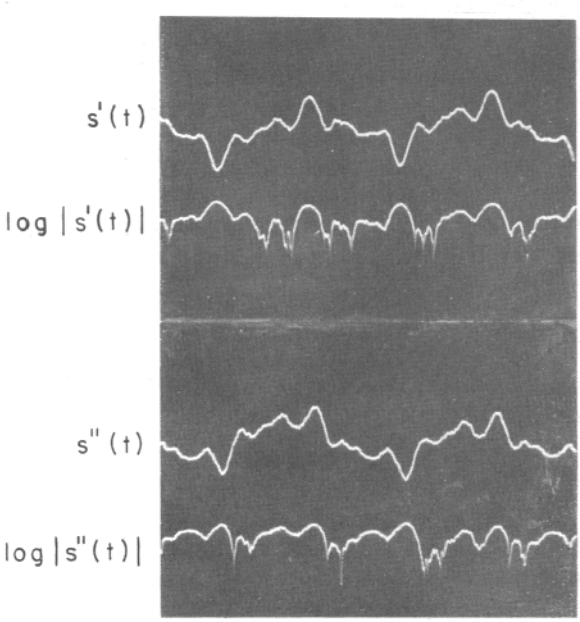
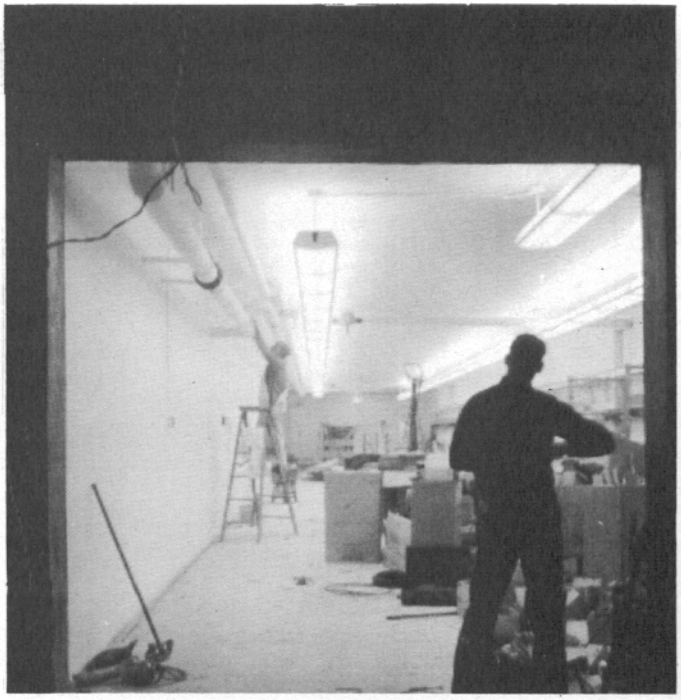
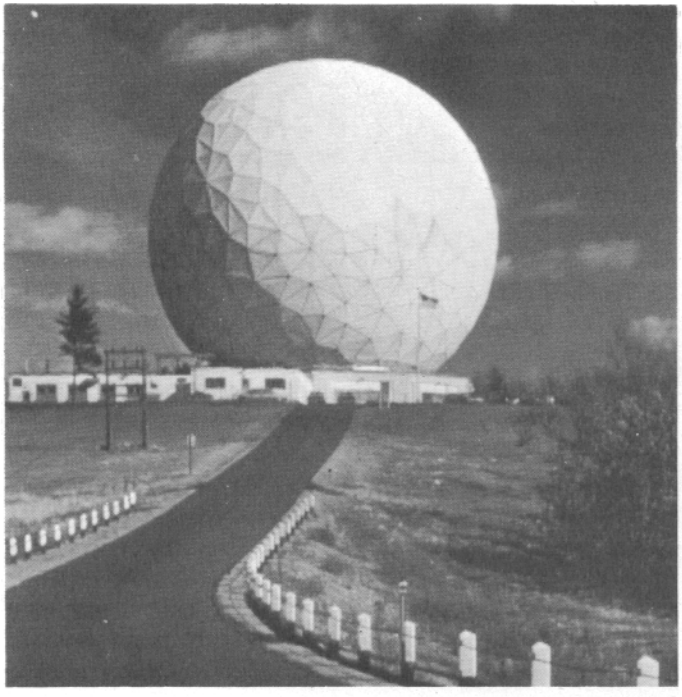


Fig. 19. $s(t)$ and log $|s(t)|$ before and after envelope distortion due to an ac coupled channel.

(a)



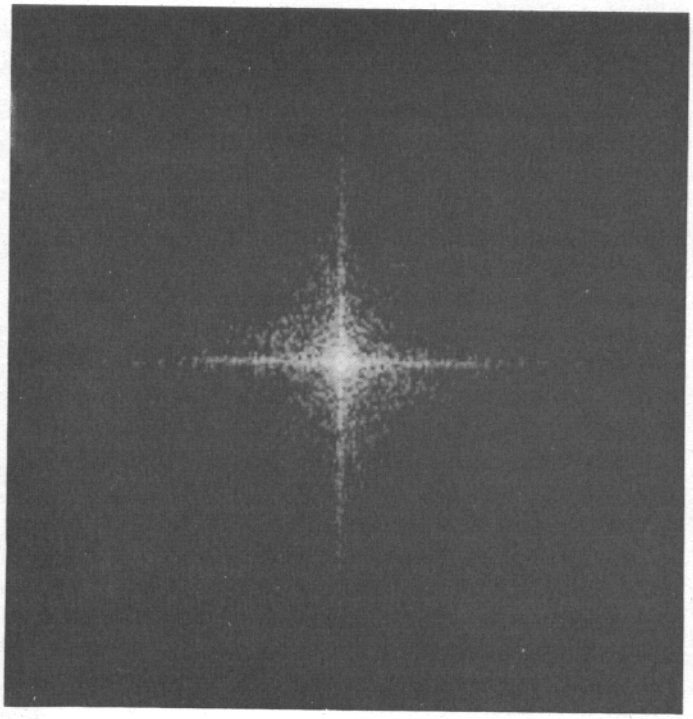
(b)



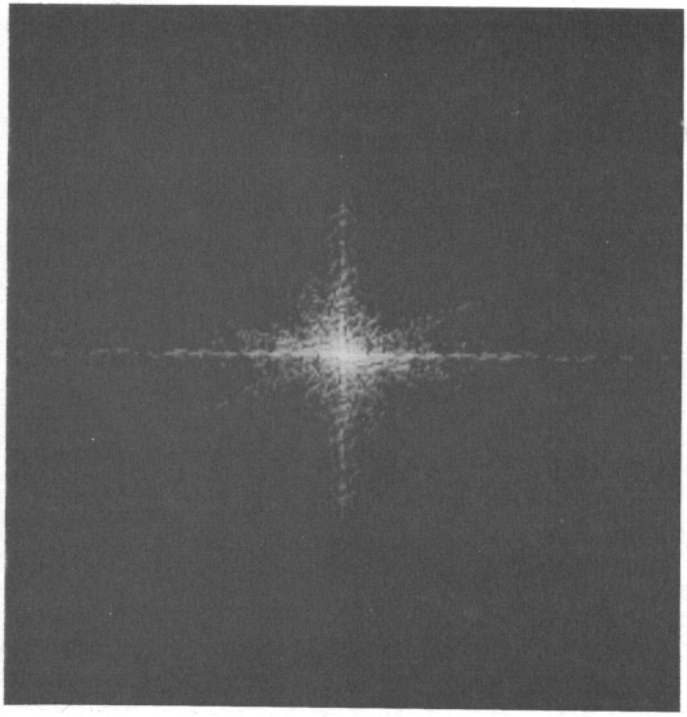
(c)

(d)

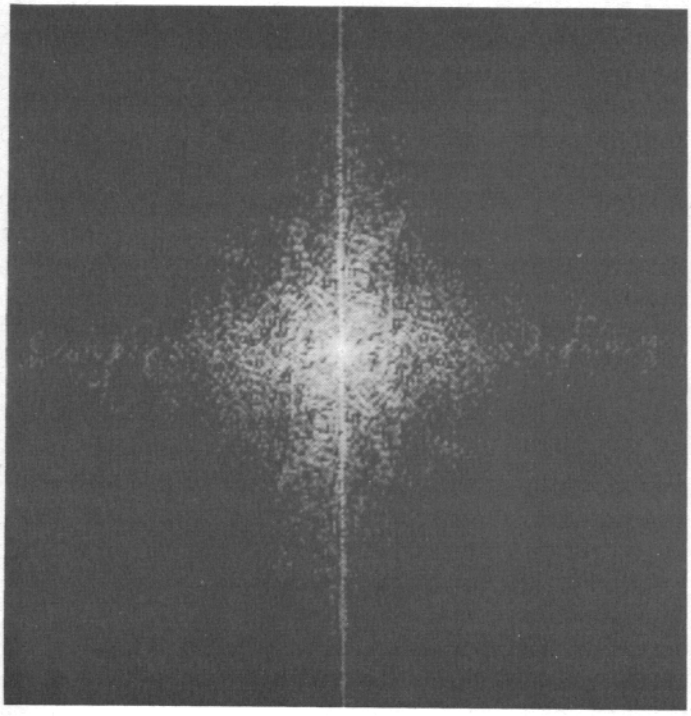
Fig. 21. Four original images.



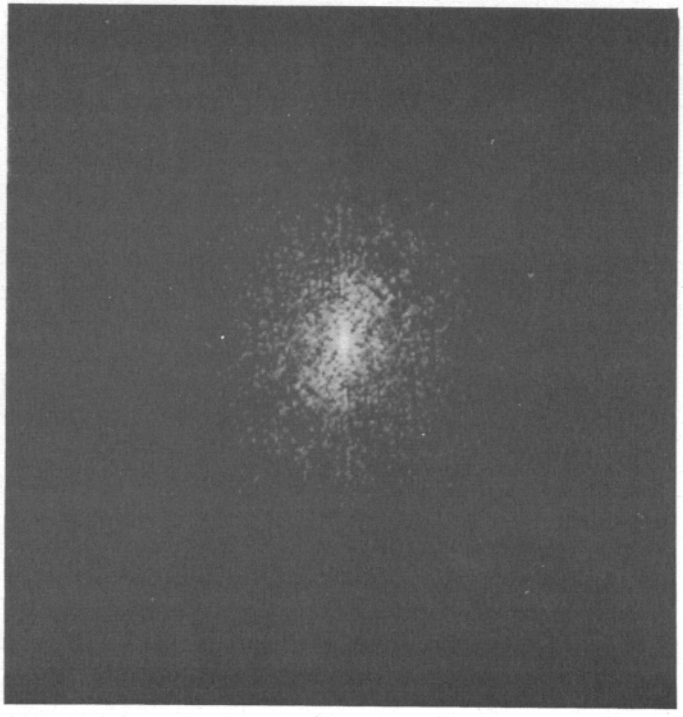
(a)



(b)

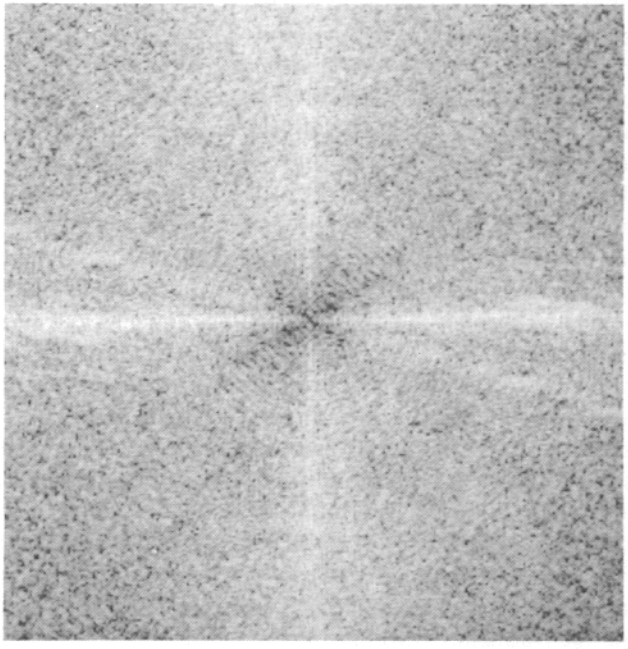


(c)

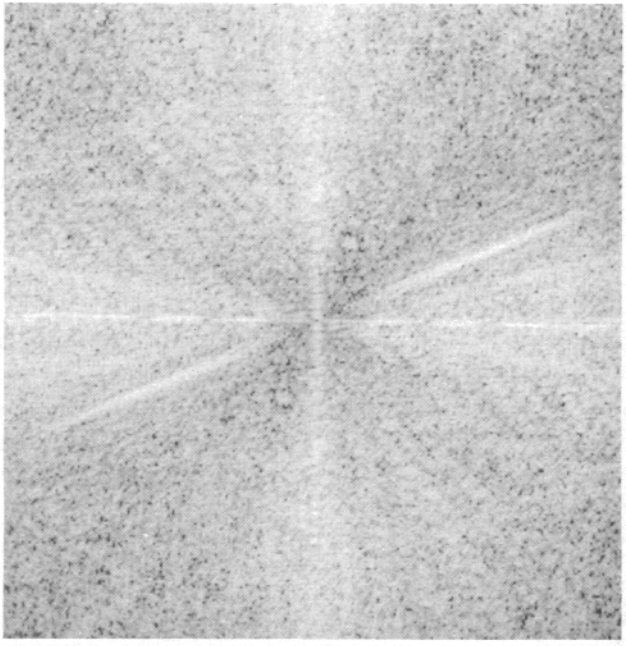


(d)

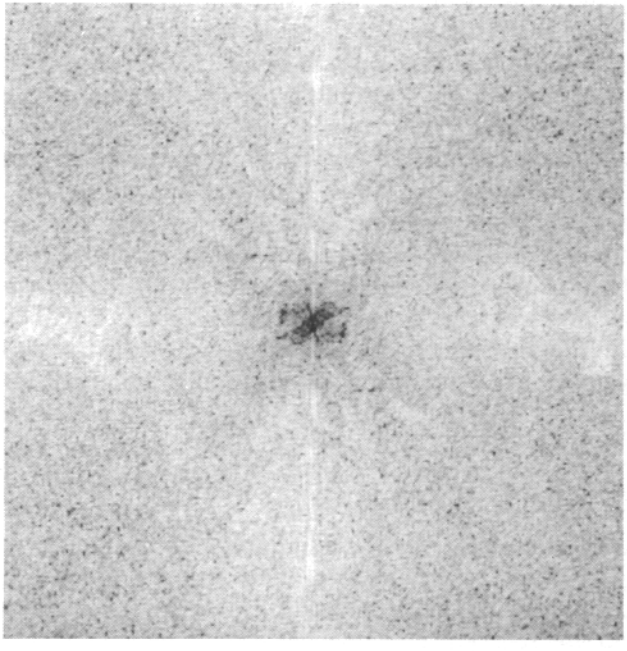
Fig. 22. Log periodograms for the log images corresponding to Fig. 21.



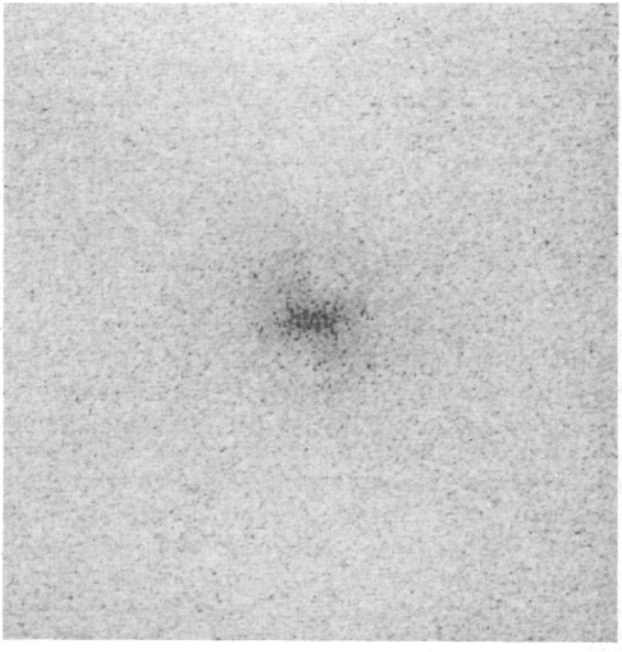
(a)



(b)

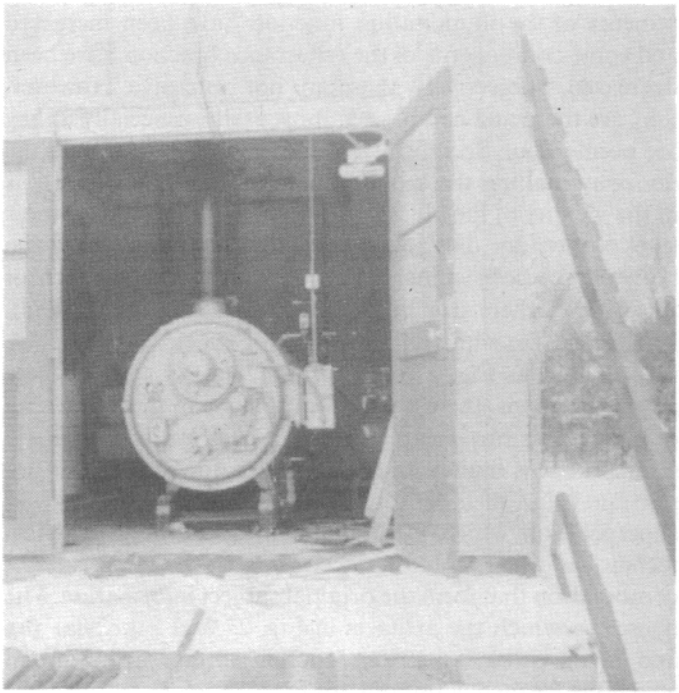


(c)

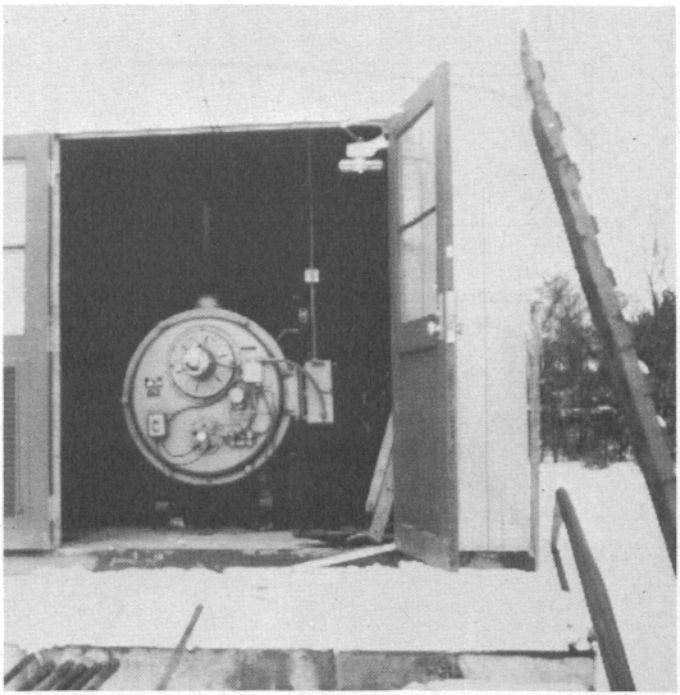


(d)

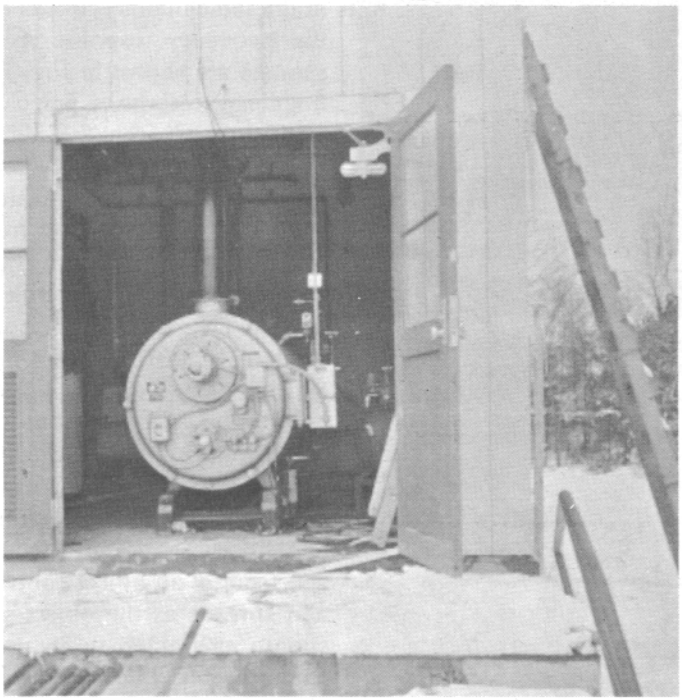
Fig. 23. Log periodograms for the whitened log images corresponding to Fig. 21.



(a)



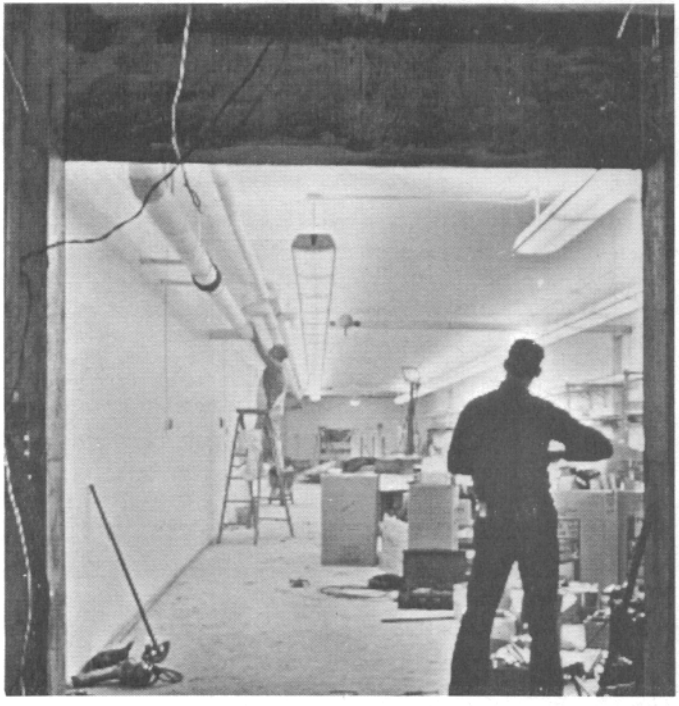
(b)



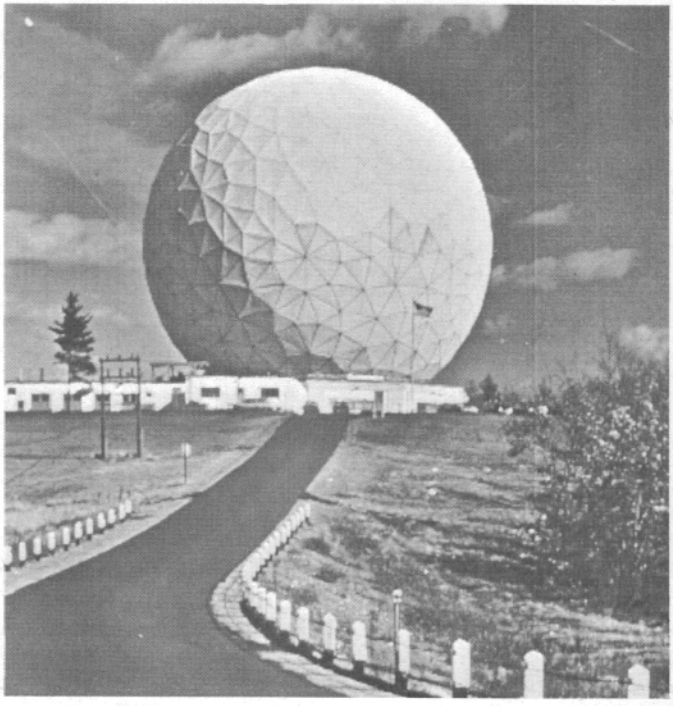
(c)

Fig. 25. The image of Fig. 21(a) processed using (a) $\gamma = 1/2$, (b) $\gamma = 2$, (c) a frequency-dependent γ with low-frequency attenuation.

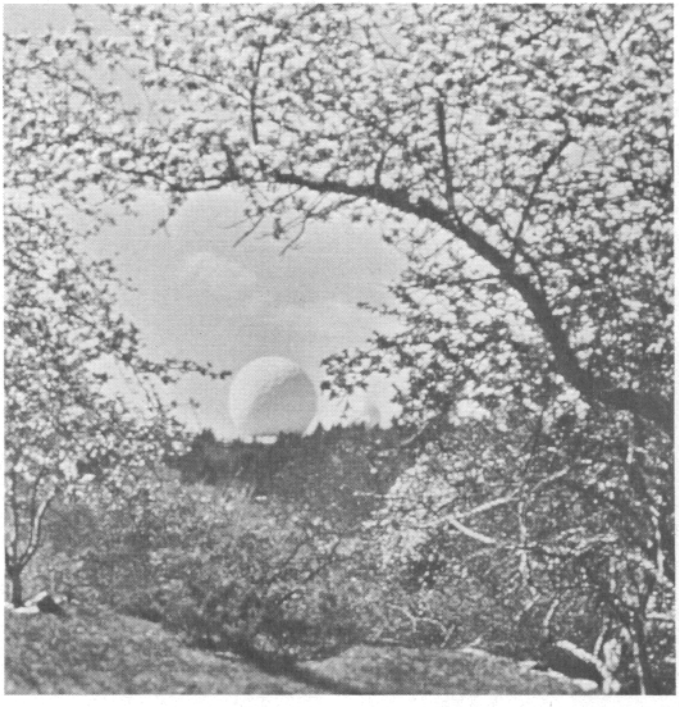
Fig. 27. The image of Fig. 21(a) processed using a frequency-dependent γ with low-frequency attenuation and high-frequency amplification.



(a)



(b)

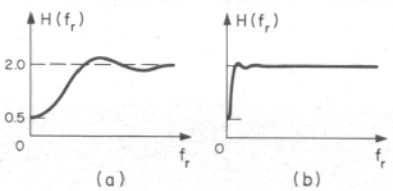


(c)

Fig. 29. The remaining images of Fig. 21 processed as in Fig. 27.

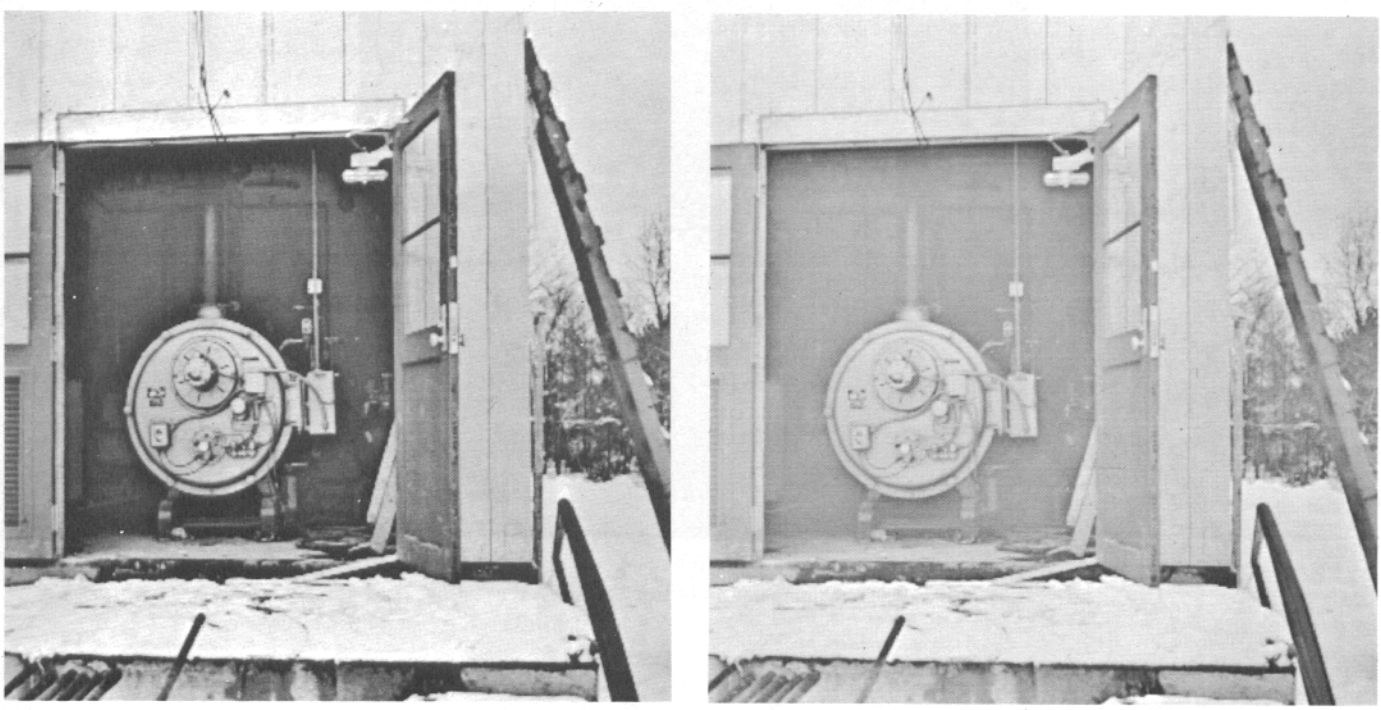
(a) (b)

Fig. 30. Images processed using abruptly changing frequency characteristics.



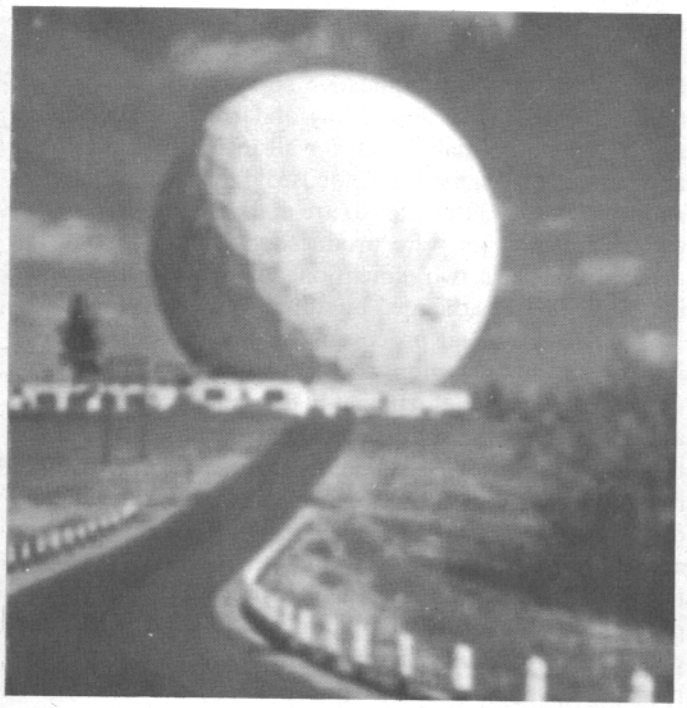
(a) (b)

Fig. 31. The multiplicative frequency responses used to produce the images of Fig. 30 from the image of Fig. 21(b).

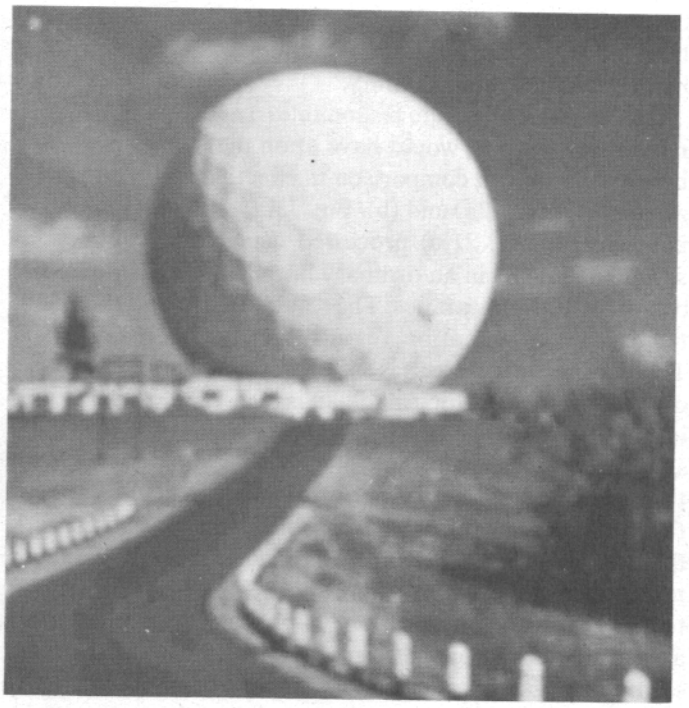


(a) Biased for best appearance. (b) Average value of image restored.

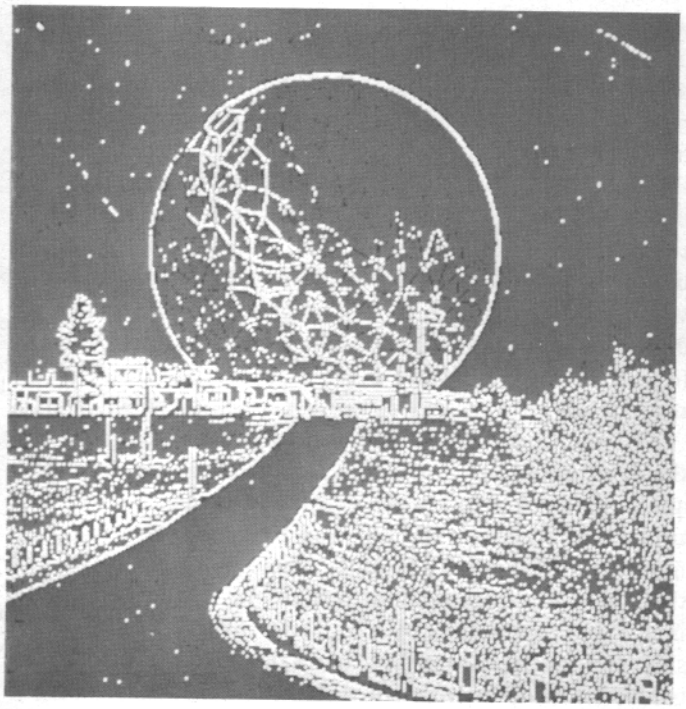
Fig. 32. Two versions of the image of Fig. 21(a) processed as in Fig. 27 but using an ordinary linear filter rather than a multiplicative filter.



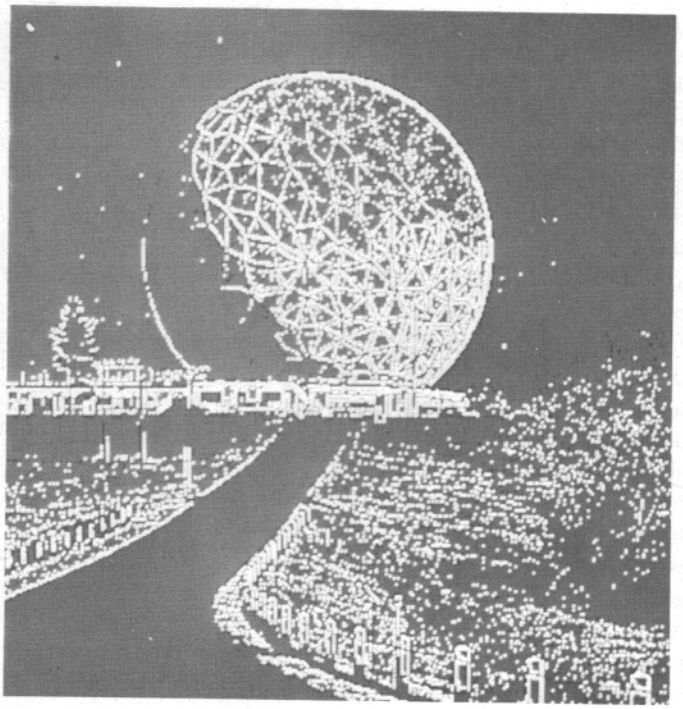
(a) Low-pass, multiplicative.



(b) Low-pass, linear.

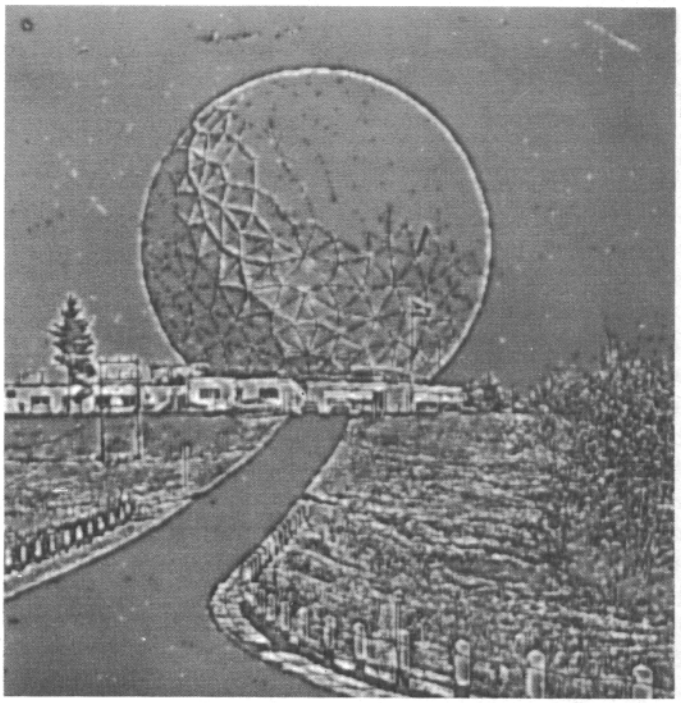


(c) Edge contours, multiplicative.

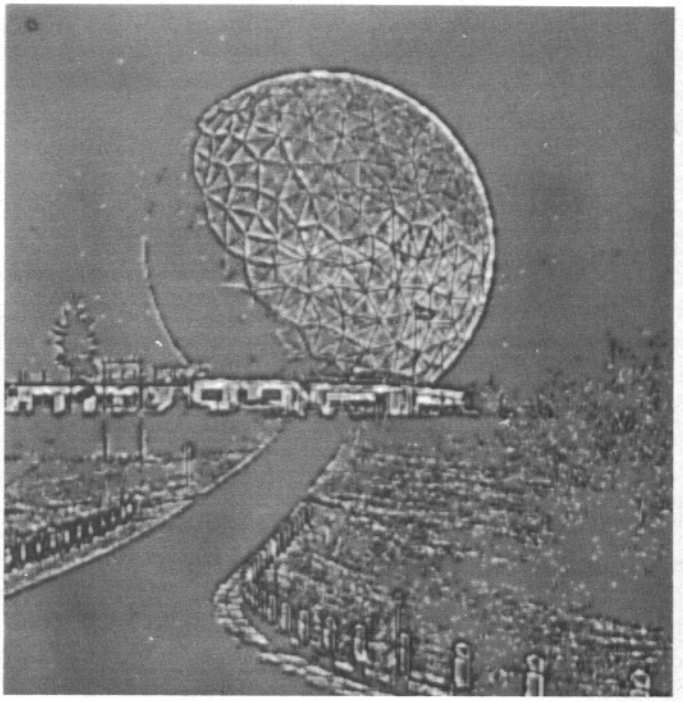


(d) Edge contours, linear.

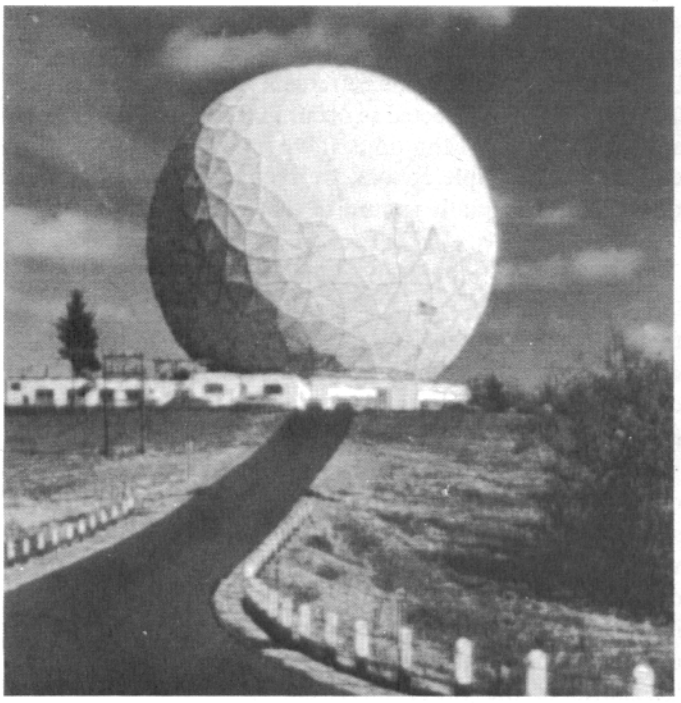
Fig. 33. The image of Fig. 21(c) in various stages of bandwidth compression.



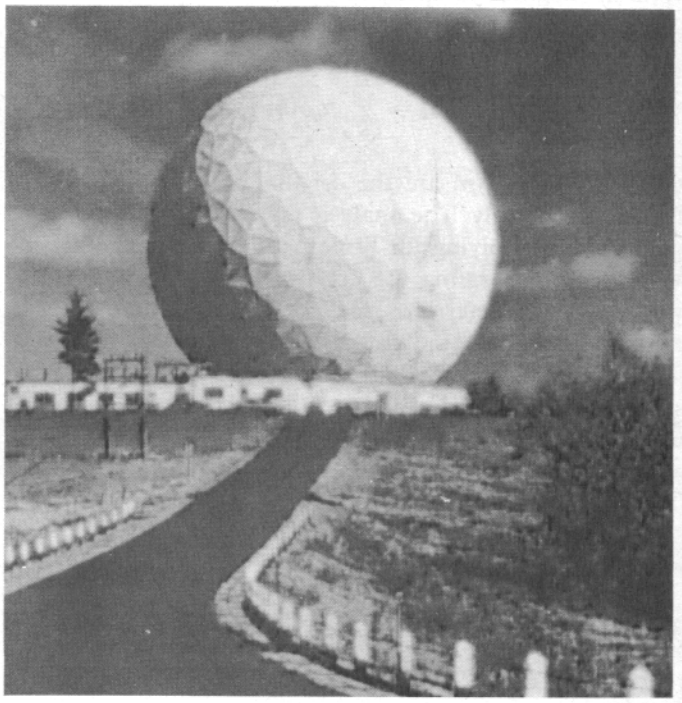
(e) Artificial highs, multiplicative.



(f) Artificial highs, linear.



(g) Recreated, multiplicative.



(h) Recreated, linear.

Fig. 33 (*Cont'd*).

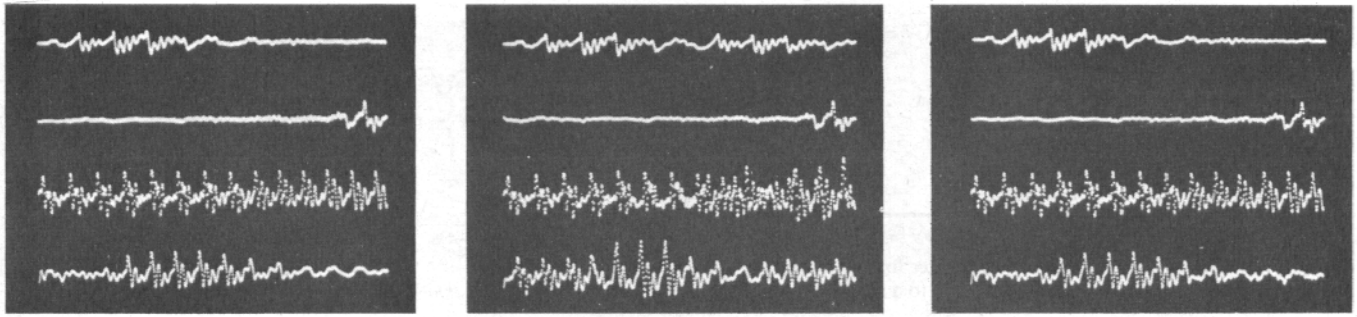


Fig. 37. An example of homomorphic echo removal. (a) 410 ms of speech sampled at 10 kHz with the four traces from top to bottom representing contiguous segments of 102.5 ms. (b) The speech sample of (a) with a 50-ms echo. (c) The speech sample of (b) processed to remove the echo.

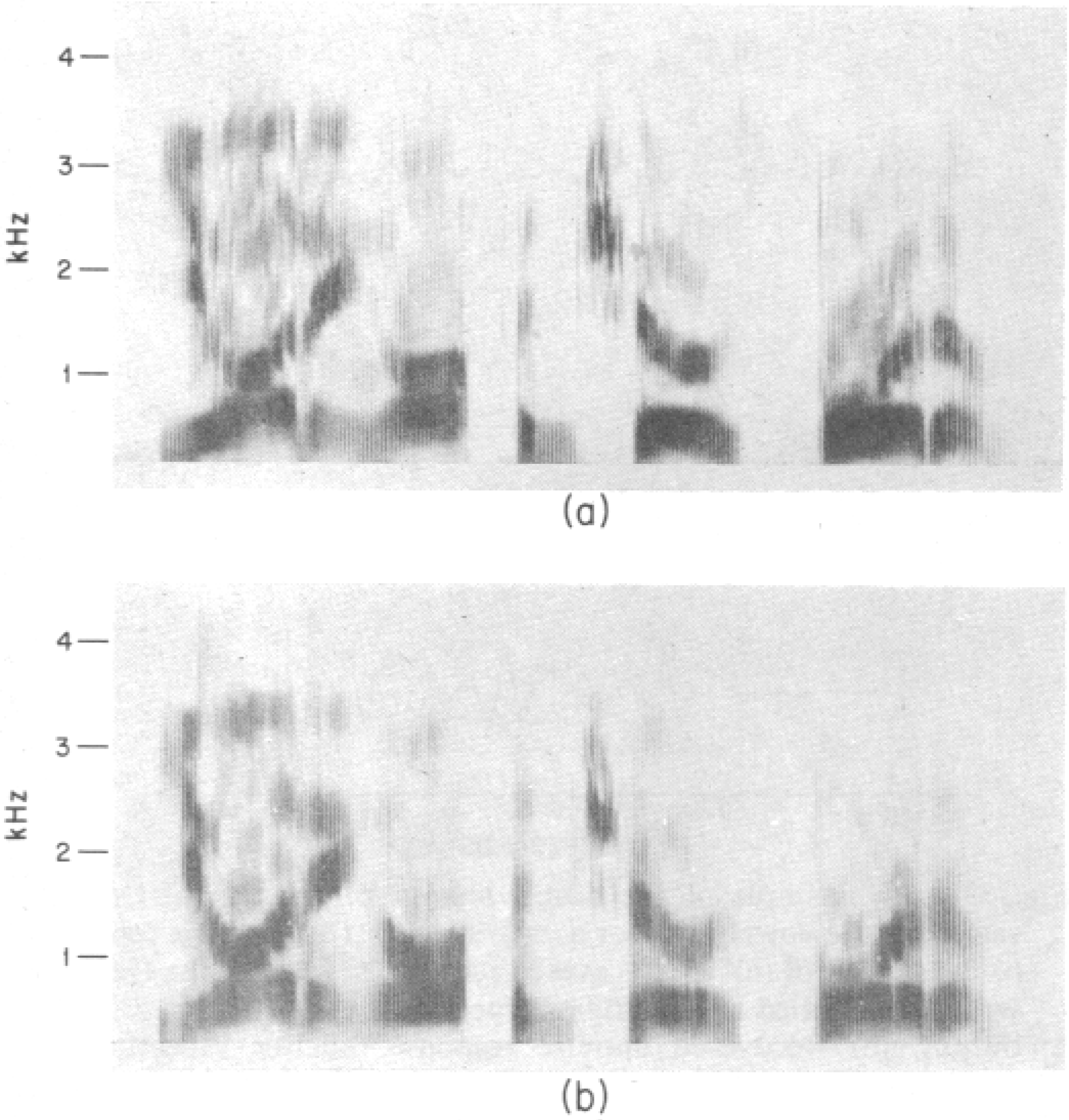


Fig. 41. (a) Spectrogram of original speech. The sentence is "yawning often shows boredom." (b) Spectrogram of synthesized speech.