AD688195

ECHO REMOVAL BY DISCRETE GENERALIZED LINEAR FILTERING

RONALD W. SCHAFER

TECHNICAL REPORT 466

FEBRUARY 28, 1969

D D C
JUN 3 1969
A

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
RESEARCH LABORATORY OF ELECTRONICS
CAMBRIDGE, MASSACHUSETTS 02139

Reproduced by the
CLEARINGHOUSE
for Federal Scientific & Technical
Information Springfield Va. 22151

133

The Research Laboratory of Electronics is an interdepartmental laboratory in which faculty members and graduate students from numerous academic departments conduct research.

The research reported in this document was made possible in part by support extended the Massachusetts Institute of Technology, Research Laboratory of Electronics, by the JOINT SERVICES ELECTRONICS PROGRAMS (U.S. Army, U.S. Navy, and U.S. Air Force) under Contract No. DA 28-043-AMC-02536(E).

Reproduction in whole or in part is permitted for any purpose of the United States Government.

Qualified requesters may obtain copies of this report from DDC.

ACCESSION for
CFSTI WHITE SECTION
DDC BUFF SECTION
UNANNOUNCED
JUSTIFICATION
BY
DISTRIBUTION/AVAILABILITY CODES
DIST. AVAIL. and/or SPECIAL

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

RESEARCH LABORATORY OF ELECTRONICS

Technical Report 466 February 28, 1969

ECHO REMOVAL BY DISCRETE GENERALIZED LINEAR FILTERING

Ronald W. Schafer

This report is based on a thesis submitted to the Department of Electrical Engineering, M. I. T., January 19, 1968, in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

(Revised manuscript received November 18, 1968)

Abstract

A new approach to separating convolved signals, referred to as homomorphic deconvolution, is presented. The class of systems considered in this report is a member of a larger class called homomorphic systems, which are characterized by a generalized principle of superposition that is analogous to the principle of superposition for linear systems.

A detailed analysis based on the z-transform is given for discrete-time systems of this class. The realization of such systems using a digital computer is also discussed in detail. Such computational realizations are made possible through the application of high-speed Fourier analysis techniques.

As a particular example, the method is applied to the separation of the components of a convolution in which one of the components is an impulse train. This class of signals is representative of many interesting signal-analysis and signal-processing problems such as speech analysis and echo removal and detection. It is shown that homomorphic deconvolution is a useful approach to either removal or detection of echoes.

TABLE OF CONTENTS

I. INTRODUCTION 1

1.1 Generalized Superposition 2

1.2 The Cepstrum 6

II. ANALYSIS OF DISCRETE-TIME HOMOMORPHIC DECONVOLUTION 7

2.1 Complex Logarithm 8

2.2 Realizations for the Systems D and D^{-1} 12

2.3 Integral Relations for the Complex Cepstrum 13

2.4 "Time-Domain" Expressions for the Complex Cepstrum 15

2.5 Complex Cepstrum for Sequences with Rational z-transforms 17

2.6 Minimum-Phase and Maximum-Phase Sequences 24

2.7 Exponential Weighting of Sequences 31

2.8 More General Rational z-Transforms 33

2.9 Examples of Complex Cepstra 36

2.10 Linear System in the Canonic Representation 42

III. COMPUTATIONAL CONSIDERATIONS IN HOMOMORPHIC DECONVOLUTION 45

3.1 Sampled z-Transform 45

3.2 Fast Fourier Transform 51

3.3 Properties of the Sampled-Phase Curves 52

3.4 An Algorithm for Computing arg [X(k)] 54

3.5 Other Computational Considerations 59

3.6 Computation Time Requirements 61

3.7 Minimum-Phase Computations 62

3.8 Sampling of Continuous-Time Signals 65

IV. ANALYSIS OF SPEECH WAVEFORMS 68

4.1 Speech Production and the Speech Waveform 68

4.2 Short-Time Transform 70

4.3 Short-Time Complex Cepstrum of Speech 71

4.4 Examples 73

V. APPLICATIONS TO ECHO REMOVAL AND DETECTION 77

5.1 A Simple Example 77

5.2 Complex Cepstrum of an Impulse Train 83

5.3 Distorted Echoes 91

5.4 Linear Systems for Echo Removal and Detection 95

5.5 Short-Time Echo Removal 101

5.6 Removal of Echoes from Speech Signals 110

5.7 Effect of Additive Noise 114

5.8 Detection of Echoes 118

CONTENTS

VI. CONCLUSION 120
6.1 Summary 120
6.2 Suggestions for Future Research 120

Appendix Vector Space for Convolution 121

Acknowledgment 124

References 125

I. INTRODUCTION

In many physical situations, we encounter signals or waveforms that may be represented as the convolution of two or more components. One class of these problems arises when a signal is distorted by transmission through a linear system. For example, the effects of multipath and reverberation may be modeled in terms of a signal that is passed through a linear system whose impulse response is an impulse train. In this case we may be interested either in recovering the undistorted signal or in determining the parameters of the impulse response. A similar class of problems arises when we are given a waveform that can be represented as a convolution of two or more component signals, and we may wish to determine these components so as to characterize the waveform or the physical process from which it originated. For example, certain segments of speech waveforms may be represented as the convolution of several components. Most speech bandwidth-compression schemes are based on the determination of the parameters of these component waveforms.

The process of separating the components of a convolution is termed deconvolution. In performing deconvolution of a waveform we must determine an appropriate transformation of the waveform into the desired component waveform. A common method of deconvolution is called inverse filtering. In this method, the signal is transformed by a linear time-invariant system whose system function is the reciprocal of the Fourier transform of the components to be removed. Although inverse filtering has been successfully applied in processing many different types of signals,[5,6] it is limited by the necessity of knowing the signal to be removed, as well as having a sensitivity to additive noise. Another deconvolution technique is based on the Wiener theory of linear filtering. This technique has been extensively applied in processing seismic waveforms.[6] In detection of echoes, maximum-likelihood methods[8] and correlation have been used. Various other techniques have been developed for special situations.[4,7] It is difficult to compare the various methods of deconvolution because generally each method requires different information about the signals and the objectives of each method are not precisely the same. Nevertheless, it is clear that there is not a single best method that can be applied to all deconvolution problems. Given the importance of the problem of deconvolution, it seems that even though a variety of methods are available, at present, it is cogent to investigate other approaches. The detailed consideration of a new approach to deconvolution is therefore the subject of this report.

The approach to deconvolution presented here was originally proposed by Professor Alan V. Oppenheim as an application of the theory of generalized superposition.[1,3] The parallel development of the applications of this technique to speech analysis[19,20] by Oppenheim, and to echo removal[9] by the author led to the theoretical formulation of the technique presented in this report.

Our purpose is to give a detailed discussion of the characteristics of this new approach to separating convolved signals. Since it appears that digital realizations of

this signal-processing method are most promising, our analysis will be confined to discrete-time signals and will be based on the z-transform. We shall also investigate carefully the actual realization of technique in the form of algorithms for a digital computer. As an example of the use of this technique, we have considered the problem of deconvolution for the class of signals that are represented as the convolution of one or more waveforms with an impulse train. This kind of representation is characteristic of the waveforms of speech and music and many other acoustic disturbances. Also, seismic signals, sonar signals, and many biological signals are in this class. In fact, any signal that is quasi-periodic by nature, or any signal that has been transmitted through a reverberant environment will have such a representation.

We shall now review the theory of generalized superposition, its relation to "cepstral" analysis,[10-13] and its application to deconvolution. In Section II a detailed analysis of the technique will be presented, and in Section III we shall focus on computational considerations. In the rest of the report we shall discuss applications to speech processing and to echo removal and detection.

1.1 GENERALIZED SUPERPOSITION

A system is often defined abstractly as a unique transformation of an input signal or waveform x into an output signal y. The signals are represented by functions of time, and the system corresponds to the mathematical concept of an operator. Such transformations are denoted by

$$y = T[x].$$

In order to characterize and classify systems, we place restrictions on the form of the operator T[ ]. For example, the class of linear systems is characterized by the property

$$T[ax_1+bx_2] = aT[x_1] + bT[x_2]. \tag{1}$$

Similarly, the class of time-invariant systems is characterized by the property that if

$$T[x(t)] = y(t),$$

then

$$T[x(t+t_0)] = y(t+t_0). \tag{2}$$

The class of linear time-invariant (LTI) systems has both of the properties expressed by Eqs. 1 and 2. As a direct consequence of these properties, it can be shown[27] that all LTI systems are described by the convolution integral

$$y(t) = \int_{-\infty}^{\infty} x(\tau)\, h(t-\tau)\, d\tau = \int_{-\infty}^{\infty} h(\tau)\, x(t-\tau)\, d\tau, \tag{3}$$

where y(t) is the output, x(t) is the input, and h(t) is the response of the system

to a unit impulse. The class of LTI systems is very important for three basic reasons.

1. Linear time-invariant systems are rather easy to analyze and characterize.
2. It is possible to design linear systems to perform a large variety of useful functions.
3. Many naturally occurring phenomena are accurately modeled using linear system theory.

The first of these comments is primarily a consequence of the principle of superposition (Eq. 1) which characterizes linear systems. In particular, when the input is a sum of component signals, a linear system is very convenient for separating one component from the other. As we shall see, our approach to deconvolution is motivated by similar considerations.

Classes of systems are defined by placing restrictions on the transformation that represents the system. To state that a system is nonlinear does nothing to characterize the properties of that system. An approach to characterizing nonlinear systems which is based on linear algebra has been presented by Oppenheim.[1] In this approach it is recognized that vector spaces of time functions at the input and output of a system can be constructed with a variety of definitions of vector addition and scalar multiplication. Thus many nonlinear systems can be represented as linear transformations between vector spaces and can thus be said to obey a generalized principle of superposition. Nonlinear systems of this type have been called homomorphic systems to emphasize the fact that they are represented by algebraically linear transformations. If we take the operations of vector addition to be the same in the input and output spaces, then a generalization of the linear filtering problem follows.[2] This approach applied to the separation of convolved signals is appropriately termed homomorphic deconvolution.

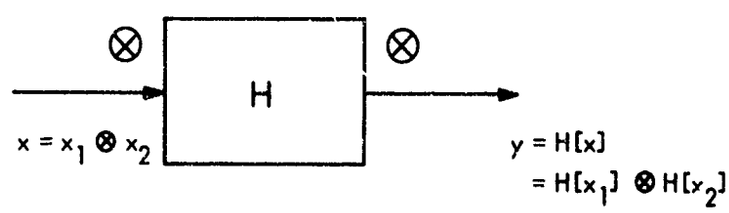


Fig. 1. Representation of a homomorphic system that obeys a generalized principle of superposition for convolution.

The class of homomorphic systems of interest for deconvolution is one in which vector addition is defined as convolution. A system of this class is shown in Fig. 1. The system H is characterized by the fact that if

$$H[x_1] = y_1 \quad \text{and} \quad H[x_2] = y_2,$$

then

$$H\left[{}^{(a)}x_1 \otimes {}^{(b)}x_2\right] = {}^{(a)}H[x_1] \otimes {}^{(b)}H(x_2) = {}^{(a)}y_1 \otimes {}^{(b)}y_2, \tag{4}$$

where $\otimes$ denotes convolution, and ${}^{(a)}$ denotes scalar multiplication. (The meaning of scalar multiplication is discussed in the Appendix.) Comparison of Eqs. 1 and 4 should suffice to show why we use the term "generalized superposition." It has been shown[1] that all homomorphic systems have a canonic representation as the cascade of a nonlinear system followed by a linear system and then another nonlinear system. For convolutional input and output spaces, this canonic form is shown in Fig. 2. The system D is a homomorphic transformation from a convolutional space to an additive space so that if $D[x_1] = \hat{x}_1$ and $D[x_2] = \hat{x}_2$, then

$$D\left[{}^{(a)}x_1 \otimes {}^{(b)}x_2\right] = aD[x_1] + bD[x_2] = a\hat{x}_1 + b\hat{x}_2.$$

The system L is a linear system in the conventional sense so that if

$$L[\hat{x}_1] = \hat{y}_1 \quad \text{and} \quad L[\hat{x}_2] = \hat{y}_2,$$

then

$$L[a\hat{x}_1+b\hat{x}_2] = aL[\hat{x}_1] + bL[\hat{x}_2] = a\hat{y}_1 + b\hat{y}_2.$$

The system D^{-1} is the inverse of the system D and it serves to transform from the additive space of L back to the convolutional space.

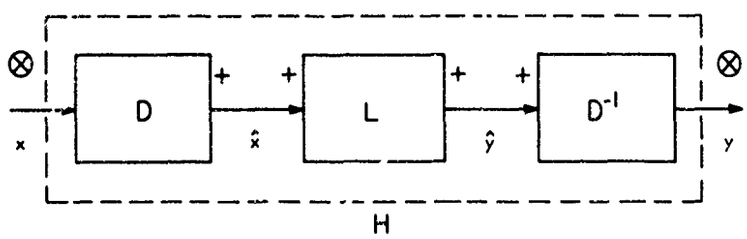


Fig. 2. Canonic form for homomorphic deconvolution.

The canonic representation is extremely important. All homomorphic systems with convolution for both input and output operations have the same form and differ only in the linear part, L. This is the reason for referring to Fig. 2 as a canonic representation. It should be clear that such a representation allows us to study such systems by first focusing our attention on the system D, and then applying the well-developed techniques of linear system theory to aid in understanding a particular over-all system H. For example, if we are interested in designing a homomorphic system for recovering signal x_1 from the convolution $x = x_1 \otimes x_2$, we need to choose the system L so that $\hat{x}_2$ is removed from the additive combination existing at the output of D.

The system D depends entirely on the specific operation for combining signals at

the input and thus is the same for all homomorphic systems for deconvolution. For this reason, the system D is called the characteristic system for homomorphic deconvolution.

The nature of the transformation D is suggested by considering the Fourier transforms of x and $\hat{x}$. Suppose

$$x = x_1 \otimes x_2,$$

so that the Fourier transform of x is

$$X = X_1 \cdot X_2, \tag{5}$$

where X, X_1, and X_2 are the Fourier transforms of x, x_1, and x_2. We see also that the Fourier transform of $\hat{x}$ must be of the form

$$\hat{X} = \hat{X}_1 + \hat{X}_2. \tag{6}$$

Equations 5 and 6 suggest that under an appropriate definition of the logarithm, we might define the system D to be the system whose output Fourier transform is the complex logarithm of the transform of the input; that is,

$$\hat{X} = \log [X].$$

Furthermore, this suggests the method of realizing the transformation D shown in Fig. 3.

Thus homomorphic deconvolution is based on transforming a convolution into a sum and then using a linear system to separate the additive components. The result is then transformed back to the original input space.

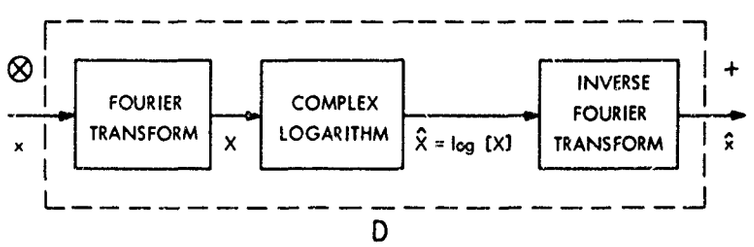


Fig. 3. Formal realization of the characteristic system for homomorphic deconvolution.

We have chosen for investigation, as examples of the application of homomorphic deconvolution, the class of signals that can be represented as a convolution in which one of the components is an impulse train. As an example of this class consider

$$x(t) = s(t) + \alpha s(t-t_o) = [u_o(t)+\alpha u_o(t-t_o)] \otimes s(t).$$

The Fourier transform of this equation is

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

The complex logarithm is formally

$$\hat{X}(\omega) = \log\left[S(\omega)\right] + \log\left(1 + \alpha e^{-j\omega t_o}\right).$$

We note that the second term in this expression is periodic in ω with a repetition rate proportional to t_o.

Suppose we view $\log\left[X(\omega)\right]$ as a waveform to be filtered with a linear system. We note that if the spectra of $\log\left[S(\omega)\right]$ and $\log\left(1 + e^{-j\omega t_o}\right)$ do not overlap, the separation of the two components is relatively easy. Alternatively, we require that the term $\log\left(1 + \alpha e^{-j\omega t_o}\right)$ vary rapidly, compared with the variations in $\log\left[S(\omega)\right]$. Thus we see that the transformation D allows us to transform a convolution of waveforms into a sum that, under appropriate conditions, can be separated by a linear system. This allows one who is familiar with linear system theory to apply all of his experience and intuition to this technique of deconvolution simply by focusing his attention on the log of the Fourier transform and interchanging the roles of time and frequency.

1.2 THE CEPSTRUM

Independently of Oppenheim's formulation of the theory of homomorphic systems and our subsequent work, Bogert, Healy, and Tukey[10] recognized that the logarithm of the power spectrum (the Fourier transform of the autocorrelation function) for a signal containing an echo should have a periodic component whose repetition rate is related to the echo delay. Thus the power spectrum of the logarithm of the power spectrum should exhibit a peak at the echo delay time. This function was called the "cepstrum" by transposing some letters of the word "spectrum." Noll[12] has traced the evolution of cepstral analysis and also discussed various definitions of the cepstrum which have been employed. Although cepstral methods have been developed from an empirical point of view, we can see that the cepstrum is clearly related to homomorphic deconvolution. The basic difference is that we shall employ a Fourier transform (magnitude and phase), rather than the power or the energy spectrum. We do this because we are concerned with the more general problem of recovery of signals as opposed to detection of echoes. To emphasize this distinction, we shall refer to the output of the characteristic system D as the complex cepstrum.

II. ANALYSIS OF DISCRETE-TIME HOMOMORPHIC DECONVOLUTION

We have introduced the concept of systems that obey a generalized principle of superposition in which addition is replaced by convolution. Since it appears, at present, that such systems can be most easily realized digitally, we shall be concerned henceforth only with discrete-time systems of this class. Thus our signal vectors are sequences of numbers, and convolution is defined as

$$x(n) = \sum_{k=-\infty}^{\infty} x_1(k)\, x_2(n-k). \tag{7}$$

The canonic form for discrete-time homomorphic systems is shown in Fig. 4, where x is the input sequence, and $\hat{x}$ is the complex cepstrum. The system D characterizes all systems of this class. Therefore we shall begin our study of such systems with a study of the system D, and then consider the choice of the linear system L.

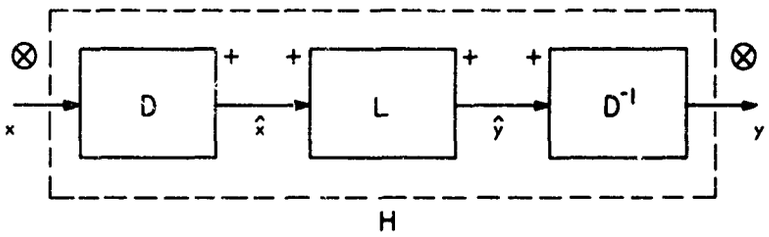


Fig. 4. Canonic form for discrete-time homomorphic deconvolution.

The properties of the transformation D can best be analyzed by considering the z-transforms of x and $\hat{x}$.[23,26] If x is a convolution,

$$x = x_1 \otimes x_2,$$

then

$$X(z) = X_1(z) \cdot X_2(z). \tag{8}$$

(Note that $\otimes$ denotes discrete-time convolution as in Eq. 7.) We require that if x is a convolution as in Eq. 7, then

$$\hat{x} = \hat{x}_1 + \hat{x}_2.$$

Thus the z-transform of $\hat{x}$ must be of the form

$$\hat{X}(z) = \hat{X}_1(z) + \hat{X}_2(z). \tag{9}$$

If we compare (8) and (9), we see that the requirement is that the system D effectively

transform a product of z-transforms into a sum of corresponding z-transforms. We shall show that, under appropriate definition of the complex logarithm,

$$\log [X_1(z)X_2(z)] = \log [X_1(z)] + \log [X_2(z)].$$

Thus we are led to define the system D as one for which the z-transform of the output is the complex logarithm of the z-transform of the input. That is,

$$\hat{X}(z) = \sum_{n=-\infty}^{\infty} \hat{x}(n)\, z^{-n} = \log [X(z)]. \tag{10}$$

Since $\log [X(z)]$ must be a z-transform, it must have the properties of a z-transform. In particular, we must be able to define a region (actually a Riemann surface) in which $\log [X(z)]$ is single-valued and analytic and possesses a Laurent series expansion. Thus before proceeding to the actual definition and discussion of the realization of the system D, it is first appropriate to review some of the properties of the complex logarithm.

2.1 COMPLEX LOGARITHM

The function X(z) can be expressed as

$$X(z) = |X(z)|\, e^{j \arg [X(z)]}.$$

The logarithm of X(z) is defined as

$$\log [X(z)] = \log |X(z)| + j \arg [X(z)]. \tag{11}$$

since $e^{j2\pi q} = 1$ for any positive or negative integer q, it is clear that we may always write $\arg [X(z)]$ as

$$\arg [X(z)] = \mathrm{ARG}\, [X(z)] \pm j2\pi q,$$

where $q = 0, 1, 2, \ldots,$ and

$$-\pi < \mathrm{ARG}\, [X(z)] \leq \pi.$$

Therefore $\log [X(z)]$ may be expressed as

$$\log [X(z)] = \log |X(z)| + j\, \mathrm{ARG}\, [X(z)] \pm j2\pi q. \tag{12}$$

That is, the complex logarithm is multivalued, with infinitely many possible values. The <u>principal value</u> of $\log [X(z)]$ is defined as the value of Eq. 12 when $q = 0$, and $\mathrm{ARG}\, [X(z)]$ is called the <u>principal value</u> of $\arg [X(z)]$. (Henceforth, the principal value of an angle will be denoted by capital letters.)

The transformation D must be unique. Therefore the logarithm must be so defined

that there is no ambiguity with respect to its imaginary part. Furthermore, we require that log $[X(z)]$ be analytic in some annular region of the z plane because the values of the complex cepstrum $\hat{x}$ are defined as

$$\hat{x}(n) = \frac{1}{2\pi j} \oint_C \log [X(z)] z^{n-1} \, dz. \tag{13}$$

In Eq. 13, C is a circular contour specified by

$$z = e^{\sigma + j\omega} \qquad -\pi < \omega \leq \pi,$$

where e^{σ} is the radius of the circle. In Eq. 13 it is assumed that log $[X(z)]$ has a Laurent series expansion as in (10). Thus we must insure that log $[X(e^{\sigma+j\omega})]$ is analytic in an annular region containing the circle with radius e^{σ}. This region is appropriately called the region of convergence of log $[X(z)]$ or of $\hat{X}(z)$.

In general, the principal value of the phase, ARG $[X(e^{\sigma+j\omega})]$ will be a discontinuous function of ω. In fact, ARG $[X(e^{\sigma+j\omega})]$ will be discontinuous for values of ω for which

$$\arg [X(e^{\sigma+j\omega})] = n\pi, \qquad n = \pm 1, +3, +5, \ldots .$$

A typical example of a phase curve and its corresponding principal value is shown in Fig. 5. If the principal value of the phase is used in defining the complex logarithm,

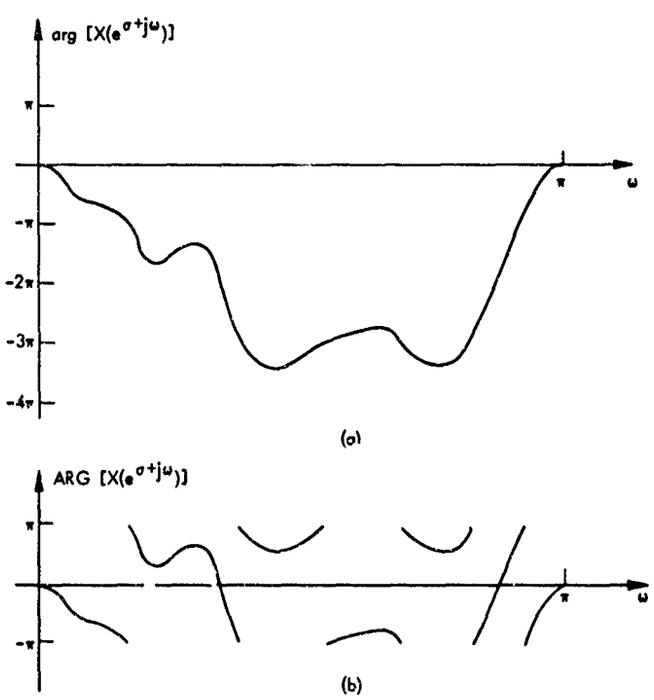


Fig. 5. (a) Typical phase curve for a z-transform evaluated on a circular contour about $z = 0$.
(b) The principal value of the phase curve in (a).

its derivative does not exist at the points of discontinuity of ARG $[X(e^{\sigma+j\omega})]$. Therefore the function log $[X(e^{\sigma+j\omega})]$ would fail to be analytic at these points. Because log $[X(z)]$ must be analytic on the contour C, we must eliminate such singular behavior by computing a phase curve with no discontinuities.

We also recall that if

$$X(z) = X_1(z)\, X_2(z),$$

then we require that

$$\log [X(z)] = \log [X_1(z)] + \log [X_2(z)],$$

on the contour C.

If we write

$$X_1(z) = |X_1(z)|\, e^{j \arg [X(z)]}$$

and

$$X_2(z) = |X_2(z)|\, e^{j \arg [X(z)]},$$

then we require that

$$\log |X(z)| = \log |X_1(z)| + \log |X_2(z)| \tag{14}$$

and

$$\arg [X(z)] = \arg [X_1(z)] + \arg [X_2(z)], \tag{15}$$

where $z = e^{\sigma+j\omega}$ and $-\pi < \omega \leq \pi$. Since log $|X(z)|$ is simply the logarithm of a positive real number, (14) will be satisfied whenever $|X_1(z)|$ and $|X_2(z)|$ are nonzero and finite.

With respect to the phase angles, we can write

$$\arg [X(z)] = \mathrm{ARG}\, [X(z)] \pm j2\pi q \tag{16a}$$

$$\arg [X_1(z)] = \mathrm{ARG}\, [X_1(z)] \pm j2\pi q_1 \tag{16b}$$

$$\arg [X_2(z)] = \mathrm{ARG}\, [X_2(z)] \pm j2\pi q_2, \tag{16c}$$

where q, q_1, and q_2 are integers. Clearly, (15) will hold only if we choose the appropriate value for arg $[X(z)]$. For example, suppose that we choose the principal value for all angles. It can be shown that, in general,

$$\mathrm{ARG}\, [X(z)] \neq \mathrm{ARG}\, [X_1(z)] + \mathrm{ARG}\, [X_2(z)].$$

One way of insuring that (15) will always hold is to assume that all angles are computed so that they are continuous functions of ω as z varies along the contour C specified by $z = e^{\sigma+j\omega}$. This implies that for each value of ω, we have chosen the proper values for

q, q_1, and q_2 in Eqs. 16 so that all angles are continuous functions of ω. In the actual computation we only compute arg $[X(z)]$, so the proper choice of q_1 and q_2 is implicit in the proper choice of q. Thus requiring that the phase curve be continuous also implies that Eq. 15 is satisfied.

Two other restrictions on the form of arg $[X(z)]$ result from considerations that do not have to do with the logarithmic operation. If we require that $\hat{x}(n)$ be real when $x(n)$ is real, the real part of $\hat{X}(e^{\sigma+j\omega})$ must be an even function of ω and the imaginary part of $\hat{X}(e^{\sigma+j\omega})$ must be an odd function of ω. Since $|X(e^{\sigma+j\omega})|$ is even for real $x(n)$, so is

$$\mathrm{Re}\,[X(e^{\sigma+j\omega})] = \log |X(e^{\sigma+j\omega})|.$$

The requirement on the imaginary part implies that we must define

$$\arg\,[X(e^{\sigma+j\omega})] = -\arg\,[X(e^{\sigma-j\omega})].$$

A final condition is required because $\log\,[X(z)]$ is to be the z-transform of the sequence $\hat{x}$; $\log\,[X(e^{\sigma+j\omega})]$ must be periodic in ω with period 2π. That is,

$$\log |X(e^{\sigma+j\omega})| = \log |X(e^{\sigma+j\omega\pm j2\pi k})|$$

and

$$\arg\,[X(e^{\sigma+j\omega})] = \arg\,[X(e^{\sigma+j\omega\pm j2\pi k})],$$

where $k = 0, 1, 2, \ldots$. This periodicity and the even and odd symmetry properties imply that $\log |X(e^{\sigma+j\omega})|$ has even symmetry about $\omega = 0, \pm\pi, \pm 2\pi, \ldots$, and likewise arg $[X(e^{\sigma+j\omega})]$ has odd symmetry about $\omega = 0, \pm\pi, \pm 2\pi, \ldots$.

To summarize, the conditions that are imposed on

$$\mathrm{Im}\,[\hat{X}(z)] = \arg\,[X(z)]$$

are the following.

(C1) arg $[X(z)]$ is a continuous function of ω for $z = e^{\sigma+j\omega}$.

(C2) arg $[X(z)]$ is an odd function of ω for $z = e^{\sigma+j\omega}$.

(C3) arg $[X(z)]$ is periodic in ω, with period 2π for $z = e^{\sigma+j\omega}$.

Conditions similar to (C2) and (C3) apply to $\log |X(z)|$ and follow automatically from the definition of the logarithm of a real number and the symmetry properties of the magnitude of a z-transform. These conditions are the following.

(C5) $\log |X(z)|$ is an even function of ω for $z = e^{\sigma+j\omega}$.

(C6) $\log |X(z)|$ is periodic in ω, with period 2π for $z = e^{\sigma+j\omega}$.

2.2 REALIZATIONS FOR THE SYSTEMS D AND D^{-1}

We have seen that if special care is taken in defining the complex logarithm, the logarithm of a product of z-transforms is the sum of the logarithms. Furthermore, under these conditions, log [X(z)] can also be thought of as the z-transform of the

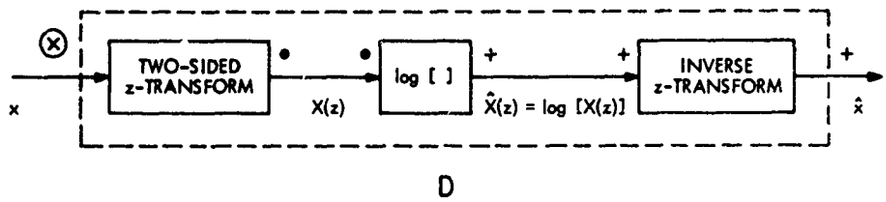


Fig. 6. Realization of the characteristic system for homomorphic deconvolution using z-transforms.

complex cepstrum. Thus, one realization of the system D is that shown in Fig. 6. The complex cepstrum is seen to be the result of the equations

$$X(z) = \sum_{n=-\infty}^{\infty} x(n)\, z^{-n} \tag{17a}$$

$$\hat{X}(z) = \log [X(z)] \tag{17b}$$

$$\hat{x}(n) = \frac{1}{2\pi j} \oint_C \log [X(z)]\, z^{n-1}\, dz, \tag{17c}$$

where the closed contour C lies in a region in which log [X(z)] has been defined as single-valued and analytic.

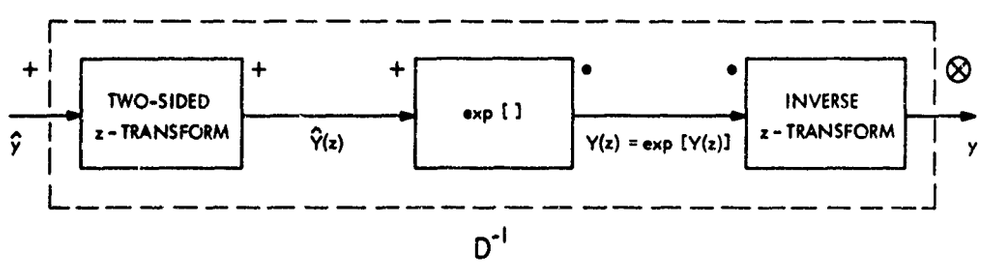


Fig. 7. Realization of the inverse characteristic system for homomorphic deconvolution using the z-transform.

Similarly the inverse of the system D is shown in Fig. 7. Thus, we obtain for the output of D^{-1} the equations

$$\hat{Y}(z) = \sum_{n=-\infty}^{\infty} \hat{y}(n)\, z^{-n} \tag{18a}$$

$$Y(z) = \exp[\hat{Y}(z)] \tag{18b}$$

$$y(n) = \frac{1}{2\pi j} \oint_{C'} Y(z)\, z^{n-1}\, dz. \tag{18c}$$

In (18c) the contour C' must be a closed contour in the region of convergence of the input z-transform X(z). This is required because if the linear system is the identity system, we require that the over-all system be the identity system; that is, if

$$\hat{y}(n) = \hat{x}(n),$$

then

$$y(n) = x(n).$$

2.3 INTEGRAL RELATIONS FOR THE COMPLEX CEPSTRUM

We have shown that the complex cepstrum can be obtained from the set of equations

$$X(z) = \sum_{n=-\infty}^{\infty} x(n)\, z^{-n} \tag{19a}$$

$$\hat{X}(z) = \log [X(z)] = \log |X(z)| + j \arg [X(z)] \tag{19b}$$

$$\hat{x}(n) = \frac{1}{2\pi j} \oint_{C} \hat{X}(z)\, z^{n-1}\, dz. \tag{19c}$$

These equations constitute a definition of the system D and also lead to a computational realization. We shall consider Eq. 19c and show how it may be used in studying the properties of the system D.

We have seen that the circular contour C must lie in the region of convergence of $\hat{X}(z)$. In using these equations for computation, part of the definition of the system D is the choice of the region of convergence of $\hat{X}(z)$. In general, the two-sided transforms X(z) and $\hat{X}(z)$ have regions of convergence which are annular regions of the z plane.[23, 26] For example, we shall usually denote the region of convergence by a relation of the form

$$R_+ < |z| < R_-.$$

By definition, these regions can contain no singularities of the z-transforms. The regions of convergence may, however, contain zeros of the z-transforms, and we shall see that these cases require special handling. Since $\hat{X}(z)$ is the complex logarithm of X(z), $\hat{X}(z)$ will have singularities at all of the singularities <u>and</u> at all of the zeros of X(z). Similarly,

$\hat{X}(z)$ will have zeros at all of the ones ($X(z) = e^{j0}$) of $X(z)$. Therefore we see that the region of convergence of $\hat{X}(z)$ can be the same as the region of convergence of $X(z)$ only if $X(z)$ has no zeros in its region of convergence. On the other hand, it should be clear that we are free to choose any annular region that does not contain singularities or zeros of $X(z)$ as the region of convergence of $\hat{X}(z)$.

The choice of the region of convergence for $\hat{X}(z)$ is based primarily on computational considerations, and at least two different choices have been found useful. In any case, it should be clear that for a given input sequence x, it is possible to obtain many different output sequences $\hat{x}$ depending on the region of convergence that is chosen for $\hat{X}(z)$. This does not mean that the output is not unique because the choice of the contour C (and therefore the choice of the region of convergence of $\hat{X}(z)$) is part of the definition of the characteristic system D. Once this contour is fixed, the output is uniquely determined.

Let us temporarily leave the contour C unspecified and obtain a more useful expression for $\hat{x}(n)$. Using Eqs. 19b and 19c, we obtain

$$\hat{x}(n) = \frac{1}{2\pi j} \oint_C \log [X(z)] z^{n-1} \, dz. \tag{20}$$

If we note that the contour C is specified by $z = e^{\sigma + j\omega}$, with $-\pi < \omega < \pi$, we can write (20)

$$\hat{X}(n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \log [X(e^{\sigma+j\omega})] \, e^{\sigma n} \, e^{j\omega n} \, d\omega. \tag{21}$$

We shall proceed to integrate (21) by parts under the assumption that $\log [X(e^{\sigma+j\omega})]$ is a single-valued periodic function of ω which is everywhere continuous. We shall find that this assumption is somewhat restrictive, but we shall also show how the results derived here apply to more general circumstances.

If we integrate (21) by parts, we obtain, for $n \neq 0$,

$$\hat{x}(n) = \frac{1}{2\pi jn} \left[\log [X(e^{\sigma+j\omega})] e^{j\omega n}\right]_{-\pi}^{\pi} - \frac{1}{2\pi jn} \int_{-\pi}^{\pi} \frac{d}{d\omega} \log [X(e^{\sigma+j\omega})] \, e^{\sigma n} \, e^{j\omega n} \, d\omega.$$

Because both $\log [X(e^{\sigma+j\omega})]$ and $e^{j\omega n}$ are periodic with period 2π, the first term in the expression above vanishes. Since we have assumed that $\log [X(e^{\sigma+j\omega})]$ is continuous everywhere, we obtain

$$\hat{x}(n) = \frac{-1}{2\pi jn} \int_{-\pi}^{\pi} \frac{\frac{d}{d\omega} X(e^{\sigma+j\omega})}{X(e^{\sigma+j\omega})} \, e^{\sigma n} \, e^{j\omega n} \, d\omega. \tag{22}$$

Since the logarithmic derivative is also analytic in the region of convergence of $\hat{X}(z)$ we may write (22)

$$\hat{x}(n) = \frac{-1}{2\pi jn} \oint_C \frac{zX'(z)}{X(z)} z^{n-1} \, dz, \tag{23}$$

where the prime indicates differentiation with respect to z. The contour C is, of course, still in the region of convergence of $\hat{X}(z)$.

The value of $\hat{x}$ at n = 0 is obtained directly from Eq. 21; that is,

$$\hat{x}(0) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \log \left[X(e^{\sigma+j\omega})\right] d\omega.$$

Since arg $[X(e^{\sigma+j\omega})]$ is an odd function of ω and $\log \left|X(e^{\sigma+j\omega})\right|$ is an even function of ω,

$$\hat{x}(0) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \log \left|X(e^{\sigma+j\omega})\right| d\omega. \tag{24}$$

Thus as an alternative to Eqs. 19 for analysis, and possibly for computational purposes, we have Eqs. 23 and 24, under the assumption that $X(e^{\sigma+j\omega})$ is a single-valued and continuous function of ω for all ω. We shall see that this condition must be relaxed in order to include most situations of interest. (This will be done in section 2.9.)

2.4 "TIME-DOMAIN" EXPRESSIONS FOR THE COMPLEX CEPSTRUM

The expressions just derived gave the complex cepstrum $\hat{x}$ explicitly in terms of the z-transform of x. Equation 23 may be used to obtain an implicit expression in terms of x(n) and $\hat{x}(n)$ which, in certain cases, reduces to a recursion formula.

This implicit relation may be derived as follows. If we assume that $\log [X(e^{\sigma+j\omega})]$ is continuous for all ω, we can write

$$\hat{X}'(z) = \frac{X'(z)}{X(z)}.$$

Rearranging this expression, we obtain

$$zX'(z) = z\hat{X}'(z)\, X(z). \tag{25}$$

Since

$$X'(z) = z^{-1} \sum_{n=-\infty}^{\infty} - nx(n)\, z^{-n},$$

we see that the inverse z-transform of (25) is

$$nx(n) = \sum_{k=-\infty}^{\infty} k\hat{x}(k)\, x(n-k). \tag{26}$$

There are several special cases of (26) that are worthy of special consideration.

Case 1: $x(n) = 0$ for $n < 0$, and $x(0) \neq 0$.

In this case we can write

$$nx(n) = \sum_{k=-\infty}^{n} k\hat{x}(k)\, x(n-k),$$

which can be written as

$$\hat{x}(n) = \frac{x(n)}{x(0)} - \frac{1}{n} \sum_{k=-\infty}^{n-1} k\hat{x}(k) \frac{x(n-k)}{x(0)} \qquad n \neq 0. \tag{27}$$

Thus we see that $\hat{x}(n)$ depends on all values of x and the values of $\hat{x}$ for $k < n$.

Case 2: Suppose that $x(0) \neq 0$ and $x(n) = 0$ for $n < 0$. If we further assume that $\hat{x}(n) = 0$ for $n < 0$, we obtain from Eq. 27

$$\hat{x}(n) = \frac{x(n)}{x(0)} - \sum_{k=0}^{n-1} \left(\frac{k}{n}\right) \hat{x}(k) \frac{x(n-k)}{x(0)} \qquad n > 0. \tag{28}$$

The value of $\hat{x}(0)$ for sequences of this type can be shown to be (see section 2.5)

$$\hat{x}(0) = \log x(0). \tag{29}$$

Requiring that $\hat{x}(n) = 0$ for $n<0$ is equivalent to choosing the contour C in (23) so as to enclose all of the poles and zeros of $X(z)$. If $X(z)$ has poles or zeros outside the unit circle, it can be shown[23] that $\hat{x}(n)$ will be unbounded for large n, since we are effectively choosing the region of convergence to be outside of all of the poles and zeros of $X(z)$. This will not be the case, however, if $X(z)$ has all of its poles and zeros inside the unit circle.

Thus when $X(z)$ has all of its poles and zeros inside the unit circle, $\hat{x}(n)$ satisfies a recursion relation that could be used in actually computing $\hat{x}(n)$. (Discussion of the utility of this expression is reserved for Section III.)

Finally, we observe that Eqs. 28 and 29 provide a way of obtaining x from $\hat{x}$, that is, a recursive relation for the inverse characteristic system. By rearranging Eqs. 28 and 29, we obtain

$$x(0) = e^{\hat{x}(0)} \tag{30a}$$

$$x(n) = \hat{x}(n)\, x(0) + \sum_{k=0}^{n-1} \left(\frac{k}{n}\right) \hat{x}(k)\, x(n-k) \qquad n > 0. \tag{30b}$$

Equations 30 represent a realization of the inverse characteristic system for sequences whose z-transforms have no poles and zeros outside the unit circle.

Case 3: Suppose that $x(0) \neq 0$, $\hat{x}(n) = 0$ for $n < 0$, and $x(n) = 0$ for $n < 0$ and $n > M$.

In this case, Eqs. 28 and 29 take the form

$$\hat{x}(0) = \log x(0) \tag{31a}$$

$$\hat{x}(n) = \frac{x(n)}{x(0)} - \sum_{k=0}^{n-1} \left(\frac{k}{n}\right) \hat{x}(k) \frac{x(n-k)}{x(0)} \qquad 0 < n \leq M \tag{31b}$$

$$= \frac{x(n)}{x(0)} - \sum_{k=n-M}^{n-1} \left(\frac{k}{n}\right) \hat{x}(k) \frac{x(n-k)}{x(0)} \qquad n > M. \tag{31c}$$

Case 4: Let $x(n) = \hat{x}(n) = 0$ for $n > 0$ and $x(0) \neq 0$.

These assumptions are equivalent to taking the contour C in (23) to be inside all of the poles and zeros of $X(z)$. Thus for a stable sequence $\hat{x}$, we require that all of the poles and zeros be outside the unit circle.[21] Using Eq. 26, we arrive at

$$\hat{x}(0) = \log x(0) \tag{32a}$$

$$\hat{x}(n) = \frac{x(n)}{x(0)} - \sum_{k=n+1}^{0} \left(\frac{k}{n}\right) \hat{x}(k)\, x(n-k) \qquad n < 0. \tag{32b}$$

2.5 COMPLEX CEPSTRUM FOR SEQUENCES WITH RATIONAL z-TRANSFORMS

In actual computations, we are always restricted to sequences of finite length and hence to z-transforms that are simply polynomials in z^{-1}. Thus it is not a significant restriction if we consider z-transforms of the form

$$X(z) = A \frac{\prod_{k=1}^{m_i} \left(1 - a_k z^{-1}\right) \prod_{k=1}^{m_o} (1 - b_k z)}{\prod_{k=1}^{p_i} \left(1 - c_k z^{-1}\right) \prod_{k=1}^{p_o} (1 - d_k z)}, \tag{33}$$

where A is a positive-real constant, and the a_k, b_k, c_k and d_k are nonzero complex numbers whose magnitudes are less than one. If x is a real sequence, then the a_k, b_k, c_k and d_k occur in complex conjugate pairs. Careful examination of Eq. 33 shows that there are m_i zeros and p_i poles inside the unit circle, and m_o zeros and p_o poles outside the unit circle. Clearly, (33) is not the most general rational z-transform, since A could be negative and in general we must include a factor of the form z^r to account for all shifted versions of the sequence x. Since our method in computation will be to deal with these issues separately, we shall defer discussion of these points.

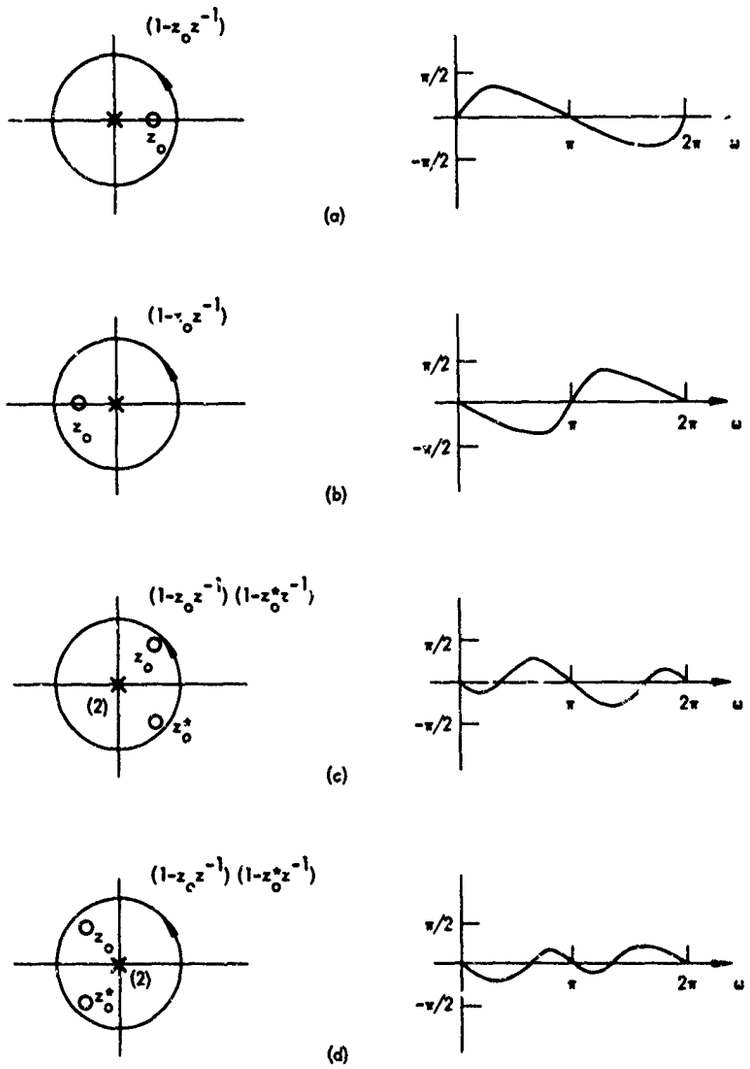


Fig. 8. Phase curves for zeros inside the unit circle.

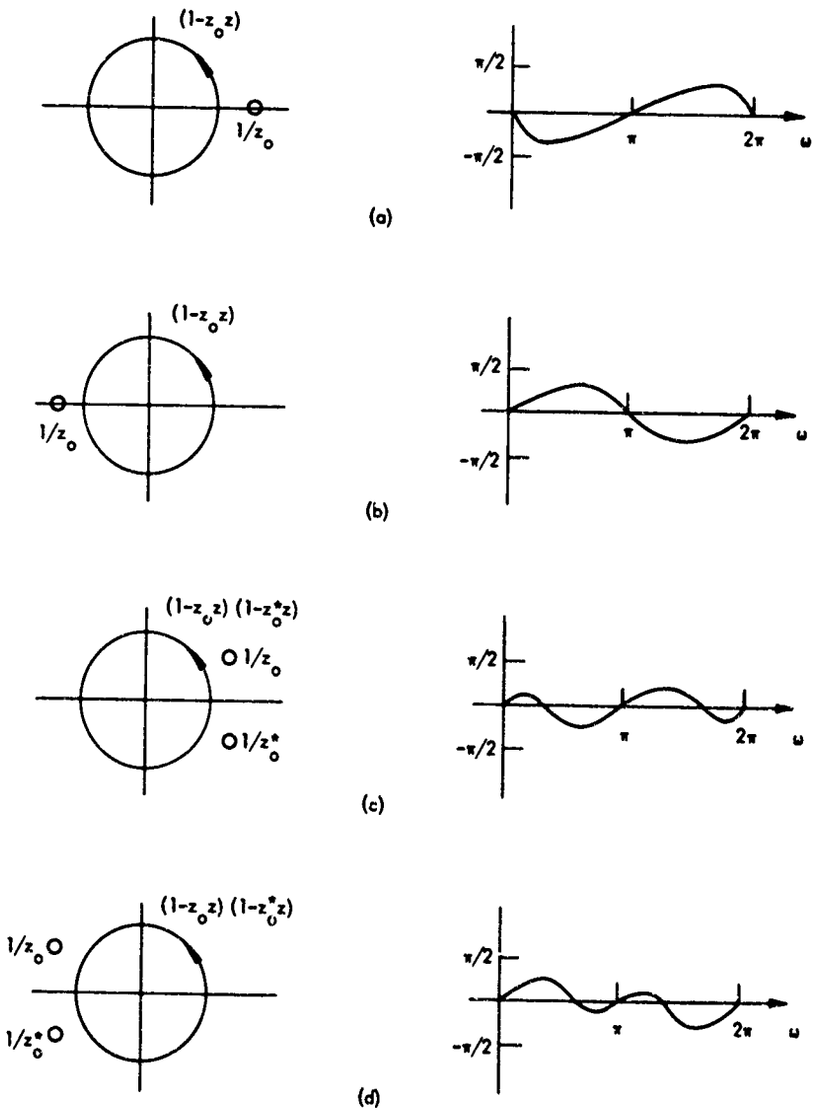


Fig. 9. Phase curves for zeros outside the unit circle.

We have shown that the phase curve must be a continuous function of ω. Since $\arg [X(e^{\sigma+j\omega})]$ will be the sum of the arguments of each multiplicative factor in Eq. 33, it is helpful to consider the contributions from each of these factors. Figures 8 and 9 show the typical pole-zero plots and one period of phase curves when $z = e^{j\omega}$ for each type of numerator factor in Eq. 33. The corresponding denominator factors produce phase curves that are the same except for sign. In all cases, the peak value of these phase curves is less than or equal to $\pi/2$. The value $\pi/2$ is attained only when the zeros (or poles) lie on the unit circle. If the zeros (or poles) are on the unit circle, the phase curves become discontinuous. We also observe from Figs. 8 and 9 that all of the phase curves of these factors are zero at $\omega = 0, \pm\pi, \pm 2\pi, \ldots$.

Since the total phase curve for Eq. 33 is the sum of the phase curves of each factor, the total phase curve will be zero at $\omega = 0, \pm\pi, \pm 2\pi, \ldots$. Furthermore, it is clear that $\arg [X(e^{\sigma+j\omega})]$ will in general be greater in magnitude than π. Therefore in computing the phase, we must use an algorithm that enables us to determine the correct phase curve, that is, one without discontinuities.

One such algorithm computes the principal value of the phase and then determines the correct multiple of 2π to add to or subtract from the principal value for each value of ω. This algorithm is discussed in Section III.

We are also interested in $\log |X(e^{\sigma+j\omega})|$, since this is the real part of the complex logarithm. Since the magnitude is an even function of ω, it will have the same general form for poles and zeros both inside and outside the unit circle. Let us consider a factor such as $\left(1 - z_o z^{-1}\right)$ for $z = e^{j\omega}$, and $z_o = |z_o| e^{j\phi_o}$. The magnitude of such a factor is

$$|1 - z_o e^{-j\omega}| = \left[1 + |z_o|^2 - 2|z_o| \cos(\omega - \phi_o)\right]^{1/2}.$$

Taking the logarithm, we obtain

$$\log |1 - z_o e^{-j\omega}| = \frac{1}{2} \log \left[1 + |z_o|^2 - 2|z_o| \cos(\omega - \phi_o)\right].$$

This function is sketched in Fig. 10. We note that it is periodic with period 2π. The maximum positive value is $\frac{1}{2} \log \left(1 + 2|z_o| + |z_o|^2\right)$ which approaches log (2) as $|z_o|$ approaches 1. Similarly, the most negative value is $\frac{1}{2} \log \left(1 - 2|z_o| + |z_o|^2\right)$ which approaches log (0) or $-\infty$ as $|z_o|$ approaches 1.

Since $\log |X(e^{j\omega})|$ is the sum of terms such as this (with negative signs for denominator factors), we would expect that $\log |X(e^{j\omega})|$ would have an appearance similar to that of Fig. 10, except that in general there will be peaks corresponding to each of the poles and zeros of $X(e^{j\omega})$. A typical example of $\log |X(e^{j\omega})|$ is shown in Fig. 11.

We have seen that z-transforms having the form of Eq. 33 satisfy the requirement that $\log [X(e^{\sigma+j\omega})]$ be continuous everywhere. Thus we may employ Eq. 23 to evaluate the complex cepstrum. The integrand $zX'(z)/X(z)$, in this case, is

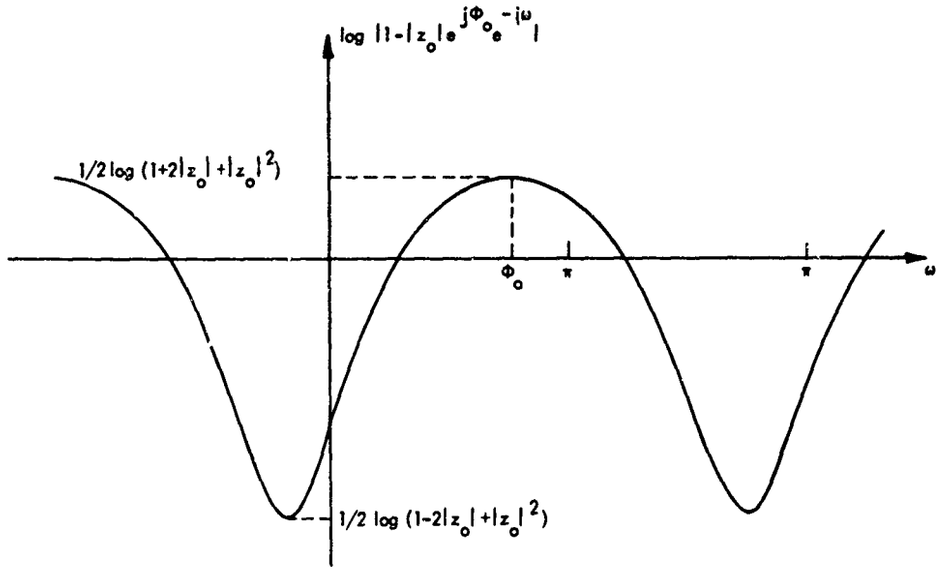


Fig. 10. Logarithm of the magnitude of the complex factor $\left(1 - z_o\, e^{-j\omega}\right)$ for $z_o = |z_o|\, e^{j\phi_o}$.

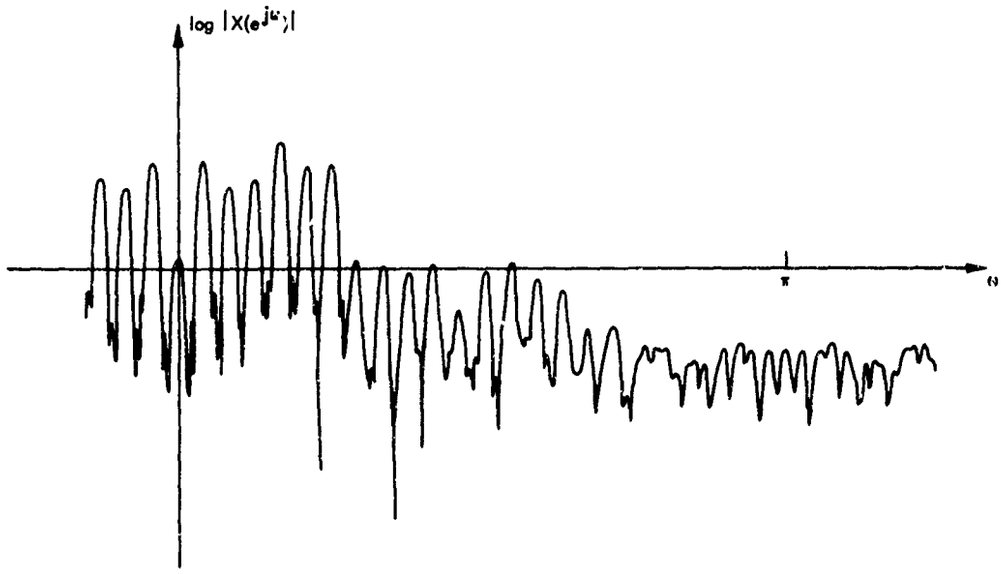


Fig. 11. Typical curve for the logarithm of the magnitude of the z-transform of a finite-length sequence.

$$z\frac{X'(z)}{X(z)} = \sum_{k=1}^{m_i} \frac{a_k z^{-1}}{1 - a_k z^{-1}} - \sum_{k=1}^{m_o} \frac{b_k z}{1 - b_k z} - \sum_{k=1}^{p_i} \frac{c_k z^{-1}}{1 - c_k z^{-1}} + \sum_{k=1}^{p_o} \frac{d_k z}{1 - d_k z}. \tag{34}$$

Since

$$\hat{x}(n) = \frac{-1}{2\pi jn} \oint_C \frac{zX'(z)}{X(z)} z^{n-1}\, dz,$$

we see that if we desire a stable sequence (one whose values approach zero for large n), we must choose the region of convergence to include the unit circle. Each factor in (34) is the z-transform of an exponential sequence. Therefore, if the contour is taken as the unit circle, $\hat{x}(n)$ is given by

$$x(n) = \sum_{k=1}^{p_i} \frac{c_k^n}{n} - \sum_{k=1}^{m_i} \frac{a_k^n}{n} \qquad n \geq 1 \tag{35a}$$

$$= \sum_{k=1}^{p_o} \frac{b_k^{-n}}{n} - \sum_{k=1}^{m_o} \frac{d_k^{-n}}{n} \qquad n \leq -1. \tag{35b}$$

The value of $\hat{x}(0)$ is obtained from Eq. 24 with $\sigma = 0$. Therefore

$$\hat{x}(0) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \log \left|X(e^{j\omega})\right| d\omega.$$

Each factor of $\left|X(e^{j\omega})\right|$ has the form

$$\left|1 - a\, e^{\pm j\omega}\right| = 1 + |a|^2 - 2|a| \cos(\omega \mp \arg[a]),$$

and it can be shown that

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} \log\left(1 + |a|^2 - 2|a| \cos(\omega \mp \arg[a])\right) d\omega = 0,$$

if $|a| \leq 1$. Therefore we see that

$$\hat{x}(0) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \log A\, d\omega = \log A. \tag{36}$$

Equations 35 and 36 express $\hat{x}(n)$ in terms of the poles and zeros of the rational z-transform. They also illustrate an important property of the complex cepstrum. It is clear from Eqs. 35 that

$$|\hat{x}(n)| \leq B \frac{a^{|n|}}{|n|} \qquad n \neq 0,$$

where B is a positive constant and a is the magnitude of the pole or zero that is closest to the unit circle.

In many simple cases, it is not necessary or desirable to use the integral formulas for purposes of analysis. This is particularly true when the z-transform is a rational function. In this case a power-series expansion of log [X(z)] is usually more convenient.

Under the assumptions that log [X(z)] is defined to be single-valued and analytic in the region of convergence, and that X(z) has the form of Eq. 33, we may write

$$\log [X(z)] = \log A + \sum_{k=1}^{m_i} \log\left(1-a_k z^{-1}\right)$$

$$+ \sum_{k=1}^{m_o} \log (1-b_k z) - \sum_{k=1}^{p_i} \log\left(1-c_k z^{-1}\right) - \sum_{k=1}^{p_o} \log (1-d_k z). \tag{37}$$

Since we define $\hat{X}(z)$ as a z-transform, it must be true that

$$\hat{X}(z) = \log [X(z)] = \sum_{n=-\infty}^{\infty} \hat{x}(n)\, z^{-n}. \tag{38}$$

Thus we immediately see that

$$\hat{x}(0) = \log A.$$

If we effectively take the contour C to be the unit circle, then each of the remaining terms in (37) can be expanded in a Laurent series about z = 0. For example, we can write

$$\log\left(1-a_k z^{-1}\right) = -\sum_{n=1}^{\infty} \frac{a_k^n}{n} z^{-n} \qquad \text{for } |z| > |a_k|$$

$$-\log\left(1-c_k z^{-1}\right) = \sum_{n=1}^{\infty} \frac{c_k^n}{n} z^{-n} \qquad \text{for } |z| > |c_k|$$

$$\log (1-b_k z) = \sum_{n=-\infty}^{-1} \frac{b_k^{-n}}{n} z^{-n} \qquad \text{for } |z| < |b_k^{-1}|$$

$$-\log (1-d_k z) = -\sum_{n=-\infty}^{-1} \frac{d_k^{-n}}{n} z^{-n} \qquad \text{for } |z| < |d_k^{-1}|.$$

Therefore if we add these convergent series and collect the coefficients of z^{-n}, we can determine $\hat{x}(n)$. In general, we see that $\hat{x}(n)$ can be written

$$\hat{x}(0) = \log A \tag{39a}$$

$$\hat{x}(n) = \sum_{k=1}^{p_i} \frac{c_k^n}{n} - \sum_{k=1}^{m_i} \frac{a_k^n}{n} \qquad n \geq 1 \tag{39b}$$

$$= \sum_{k=1}^{m_o} \frac{b_k^{-n}}{n} - \sum_{k=1}^{p_o} \frac{d_k^{-n}}{n} \qquad n \leq -1. \tag{39c}$$

Equations 39 agree with Eqs. 35 and 36, as we would expect. The real value of the power-series approach is best illustrated by our use of it in discussing echo removal applications in Section V.

2.6 MINIMUM-PHASE AND MAXIMUM-PHASE SEQUENCES

We have considered a realization of the system D which was based on the z-transform. In some cases it is possible to take advantage of the properties of the z-transform to obtain simplified results. For example, we have seen (section 2.4) that under certain conditions, $\hat{x}(n)$ obeys a recursion formula. We shall now consider these cases in detail and present an alternative computation scheme.

A minimum-phase sequence is defined as a sequence whose z-transform has no poles or zeros outside the unit circle. Furthermore, the region of convergence for the z-transform includes the unit circle. For rational z-transforms, $X(z)$ is of the form

$$X(z) = A \frac{\prod_{k=1}^{m_i} \left(1 - a_k z^{-1}\right)}{\prod_{k=1}^{p_i} \left(1 - c_k z^{-1}\right)},$$

where the a_k and c_k are complex numbers whose magnitudes are less than one, and the region of convergence is specified by

$$|z| > \max_k |c_k|.$$

Such sequences have the properties

$$x(n) = 0 \qquad n < 0 \tag{40a}$$

$$x(0) \neq 0 \tag{40b}$$

$$\sum_{n=0}^{\infty} |x(n)| < \infty. \tag{40c}$$

We have shown that the complex cepstrum of such a sequence has the properties

$$\hat{x}(n) = 0 \qquad n < 0 \tag{41a}$$

$$\hat{x}(0) = \log x(0) \tag{41b}$$

$$\sum_{n=0}^{\infty} |\hat{x}(n)| < \infty. \tag{41c}$$

Since Eqs. 40 and 41 are necessary and sufficient conditions, these equations could be taken as the definition of a minimum-phase sequence.

An entirely analogous situation is called maximum-phase. In this case, X(z) has all poles and zeros outside of the unit circle, and the region of convergence includes the unit circle. In this case, x has the properties

$$x(n) = 0 \qquad n > 0 \tag{42a}$$

$$x(0) \neq 0 \tag{42b}$$

$$\sum_{n=-\infty}^{0} |x(n)| < \infty. \tag{42c}$$

Similarly, the complex cepstrum has the properties

$$\hat{x}(0) = 0 \qquad n > 0 \tag{43a}$$

$$\hat{x}(0) = \log x(0) \tag{43b}$$

$$\sum_{n=-\infty}^{0} |\hat{x}(n)| < \infty. \tag{43c}$$

For rational z-transforms, X(z) has the form

$$X(z) = \frac{B \prod_{k=1}^{m_o} (1-b_k z)}{\prod_{k=1}^{p_o} (1-d_k z)},$$

where the b_k and d_k are all less than one in magnitude, and the region of convergence is

$$|z| < \min_k |d_k^{-1}|.$$

In general, there may be poles and/or zeros on the unit circle. These cases will be formally excluded from either class; however, we shall see (section 2.7) that it is possible to move such poles and zeros inside or outside the unit circle by exponential weighting of the sequence.

If the input sequence is known to be minimum-phase, we can obtain significant simplifications in our results. We have already seen that the characteristic system D and its inverse can be realized through a recursion formula. We now wish to show that the properties of minimum-phase sequences allow other simplifications in the computation of the complex cepstrum.

Let us introduce some definitions. We define the even part of a sequence to be the sequence whose values are

$$\text{Ev}\,[x(n)] = \frac{x(n) + x(-n)}{2}. \tag{44}$$

The sequence $\text{Ev}\,[x(n)]$ is seen to have even symmetry; that is,

$$\text{Ev}\,[x(n)] = \text{Ev}\,[x(-n)].$$

Similarly, we define the odd part of a sequence as

$$\text{Odd}\,[x(n)] = \frac{x(n) - x(-n)}{2}, \tag{45}$$

which has odd symmetry; that is,

$$\text{Odd}\,[x(n)] = -\text{Odd}\,[x(-n)].$$

It can be shown that if x is a real sequence and

$$X(e^{j\omega}) = X_r(e^{j\omega}) + jX_i(e^{j\omega}),$$

then $X_r(e^{j\omega})$ is the transform of $\text{Ev}\,[x(n)]$ and similarly, $jX_i(e^{j\omega})$ is the transform of $\text{Odd}\,[x(n)]$.

Let us assume that $x(n) = 0$ for $n < 0$. In this case we can see from Eq. 44 that

$$\begin{aligned} x(n) &= 2\,\text{Ev}\,[x(n)] && n > 0 \\ &= \text{Ev}\,[x(n)] && n = 0 \\ &= 0 && n < 0. \end{aligned}$$

That is, knowledge of the even part of a sequence that is zero for $n < 0$ is sufficient to determine the entire sequence.

These properties represent a real part sufficiency theorem for z-transforms of

sequences that are zero for $n < 0$. For example, suppose we are given the real part $X_r(e^{j\omega})$ of the transform of a sequence x. Since $X_r(e^{j\omega})$ is the transform of the even part of the sequence, we can determine Ev $[x(n)]$ from $X_r(e^{j\omega})$. If $x(n) = 0$ for $n < 0$, we can determine the sequence x and therefore $X(e^{j\omega})$. Thus knowledge of $X_r(e^{j\omega})$ is sufficient to completely determine $X(e^{j\omega})$.

Similar relations hold between the logarithm of the magnitude and the phase of a z-transform, but under more restricted conditions. Rather than focus our attention on relations between the magnitude and phase, let us consider the complex cepstrum first and return to this question eventually. If $\hat{x}(n) = 0$ for $n < 0$, we can apply the previous results to the complex cepstrum to obtain

$$\hat{x}(n) = 2\,\mathrm{Ev}\,[\hat{x}(n)] \qquad n > 0 \tag{46a}$$

$$= \mathrm{Ev}\,[\hat{x}(n)] \qquad n = 0 \tag{46b}$$

$$= 0 \qquad n < 0. \tag{46c}$$

Since the real part of $\hat{X}(e^{j\omega})$ is just

$$\hat{X}_r(e^{j\omega}) = \log |X(e^{j\omega})|,$$

we see that

$$\mathrm{Ev}\,[\hat{x}(n)] = \frac{1}{2\pi}\int_{-\pi}^{\pi} \log |X(e^{j\omega})|\, e^{j\omega n}\, d\omega.$$

Thus, if $\hat{x}(n) = 0$ for $n < 0$, we only need to compute

$$\hat{X}_r(e^{j\omega}) = \log |X(e^{j\omega})|,$$

and we do not need to compute the phase.

If we wish to obtain the original sequence from $\log |X(e^{j\omega})|$, we can do so by first computing Ev $[\hat{x}(n)]$, then $\hat{x}$ by using Eqs. 46, and then obtain the original sequence by using the inverse characteristic system to compute x. Since the condition $\hat{x}(n) = 0$ for $n < 0$ has been shown to be equivalent to the condition that x be minimum-phase, we see that for <u>minimum-phase sequences</u>, the complex cepstrum can be computed from

$$X(e^{j\omega}) = \sum_{n=0}^{\infty} x(n)\, e^{-j\omega n} \tag{47}$$

$$\mathrm{Ev}\,[\hat{x}(n)] = \frac{1}{2\pi}\int_{-\pi}^{\pi} \log |X(e^{j\omega})|\, e^{j\omega n}\, d\omega \tag{48}$$

$$\hat{x}(n) = \mathrm{Ev}\,[\hat{x}(n)]\, u(n), \tag{49}$$

where

$$u(n) = 2 \qquad n > 0$$
$$= 1 \qquad n = 0$$
$$= 0 \qquad n < 0.$$

Thus the characteristic system D for minimum-phase sequences can be realized as shown in Fig. 12, where "Fourier Transform" means $X(e^{j\omega})$.

Clearly, the discussion above also indicates a unique relation between $\log |X(e^{j\omega})|$

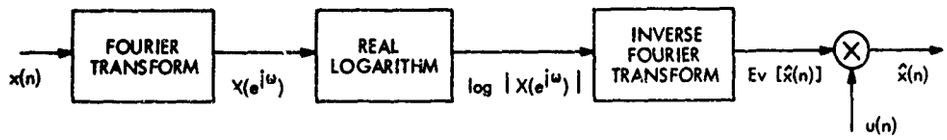


Fig. 12. A realization of the characteristic system D for minimum phase sequences.

and $\arg [X(e^{j\omega})]$. In fact, the operations illustrated in Fig. 12 are equivalent to using the Hilbert transform[24] to obtain the proper phase curve for $\log |X(e^{j\omega})|$ when $X(e^{j\omega})$ is minimum-phase.

As an example of the use of this result, let us consider the minimum-phase sequence

$$x(n) = a^n \qquad n \geq 0$$
$$= 0 \qquad n < 0.$$

The z-transform of this sequence is

$$X(z) = \frac{1}{1 - az^{-1}} \qquad \text{for } |z| > |a|,$$

and

$$\log |X(e^{j\omega})| = -\frac{1}{2}\log (1+a^2-2a\cos\omega).$$

Therefore the even part of the complex cepstrum is

$$\mathrm{Ev}\,[\hat{x}(n)] = -\frac{1}{4\pi}\int_{-\pi}^{\pi} \log (1+a^2-2a\cos\omega)\, e^{j\omega n}\, d\omega$$

which can be written

$$\mathrm{Ev}\,[\hat{x}(n)] = -\frac{1}{2\pi}\int_{0}^{\pi} \log (1+a^2-2a\cos\omega)\, \cos \omega n\, d\omega.$$

At this point, let us note two useful equations which can be found in Carslaw.[21]

$$\int_0^{\pi} \log (1+a^2-2a\cos\omega)\, d\omega = 0 \qquad |a| \leq 1$$

$$= \pi \log a^2 \qquad a \geq 1 \tag{50}$$

and

$$\int_0^{\pi} \log (1+a^2-2a\cos\omega) \cos \omega n\, d\omega = -\pi \frac{a^{|n|}}{|n|} \qquad |a| \leq 1$$

$$= -\pi \frac{a^{-|n|}}{|n|} \qquad |a| \geq 1. \tag{51}$$

These equations hold for n an integer. We can use (50) and (51) to show that

$$\mathrm{Ev}\,[\hat{x}(n)] = 0 \qquad n = 0$$

$$= \frac{1}{2} \frac{a^{|n|}}{|n|} \qquad n \neq 0.$$

Therefore we see that

$$\hat{x}(n) = 2\,\mathrm{Ev}\,[\hat{x}(n)] = \frac{a^n}{n} \qquad n > 0$$

$$= 0 \qquad n \leq 0.$$

In conclusion, we wish to call attention to an interesting representation of the input sequence, and an interesting result for finite-length sequences. It is clear from the properties of minimum-phase and maximum-phase sequences that every sequence x may be expressed as

$$x = x_{min} \circledast x_{max},$$

where

$$\hat{x}_{min}(n) = \hat{x}(n) \qquad n \geq 0$$

$$\hat{x}_{max}(n) = \hat{x}(n) \qquad n < 0.$$

For rational z-transforms, this is equivalent to

$$X(z) = X_{min}(z)\, X_{max}(z),$$

where

$$X_{min}(z) = \frac{A \prod_{k=1}^{m_i} \left(1-a_k z^{-1}\right)}{\prod_{k=1}^{p_i} \left(1-c_k z^{-1}\right)}$$

$$X_{max}(z) = \frac{\prod_{k=1}^{m_o} (1-b_k z)}{\prod_{k=1}^{p_o} (1-d_k z)} .$$

The results given here have particular significance for finite-length sequences. Suppose that X(z) has the form

$$X(z) = A \prod_{k=1}^{m_i} \left(1-a_k z^{-1}\right) \prod_{k=1}^{m_o} (1-b_k z),$$

where the a_k and b_k are all less than one in magnitude. Clearly,

$$x_{min}(n) \neq 0 \qquad 0 \leq n \leq m_i$$

$$= 0 \qquad \text{elsewhere}$$

and

$$x_{max}(n) \neq 0 \qquad -m_o \leq n \leq 0$$

$$= 0 \qquad \text{elsewhere.}$$

From Eqs. 30 and 32, we obtain the relations

$$x_{min}(n) = e^{\hat{x}(0)} \qquad n = 0$$

$$= \hat{x}(n)\, x(0) + \sum_{k=0}^{n-1} \left(\frac{k}{n}\right) \hat{x}(k)\, x_{min}(n-k) \qquad n > 0$$

$$x_{max}(n) = 1 \qquad n = 0$$

$$= \hat{x}(n) + \sum_{k=n+1}^{0} \left(\frac{k}{n}\right) \hat{x}(k)\, x_{max}(n-k) \qquad n < 0.$$

We see, therefore, that only $m_o + m_i + 1$ values of the complex cepstrum are required to completely determine the $m_o + m_i + 1$ values of the sequence x. This

result implies that even though $\hat{x}$ is of infinite duration, only a number of samples of $\hat{x}$ equal to the length of the input sequence x is required to completely determine the sequence x from the complex cepstrum.

2.7 EXPONENTIAL WEIGHTING OF SEQUENCES

Because of the special properties of minimum-phase sequences, it is of interest to consider ways of obtaining minimum-phase sequences from nonminimum-phase sequences. One way of doing this is to weight the nonminimum-phase sequence with a decaying exponential. By this we mean multiplication of the values of a sequence by α^n to obtain a new sequence whose values are

$$w(n) = \alpha^n x(n).$$

There are two important points to consider. First, we shall consider the effect of exponential weighting on a convolution, and then the effect on the z-transform and its region of convergence.

Suppose that x(n) is given by

$$x(n) = \sum_{k=-\infty}^{\infty} x_1(k)\, x_2(n-k).$$

For the exponentially weighted sequence, we obtain

$$w(n) = \alpha^n \sum_{k=-\infty}^{\infty} x_1(k)\, x_2(n-k) = \sum_{k=-\infty}^{\infty} \alpha^k x_1(k)\, \alpha^{n-k} x_2(n-k)$$

$$= \sum_{k=-\infty}^{\infty} w_1(k)\, w_2(n-k).$$

Therefore exponential weighting of a convolution of two sequences x_1 and x_2 is seen to be equivalent to the convolution of the exponentially weighted sequences w_1 and w_2 whose values are

$$w_1(n) = \alpha^n x_1(n)$$

$$w_2(n) = \alpha^n w_2(n).$$

The second interesting point is the effect of exponential weighting on the z-transform of a sequence. The z-transform of the weighted sequence w is given by

$$W(z) = \sum_{n=-\infty}^{\infty} \alpha^n x(n)\, z^{-n} = X(\alpha^{-1} z).$$

Therefore we see that if $X(z)$ has a pole or zero at $z = z_o$, then $W(z)$ has a pole or zero at αz_o. Thus if the region of convergence of $X(z)$ is

$$R_+ < |z| < R_-,$$

then the region of convergence of $W(z)$ is

$$\alpha R_+ < |z| < \alpha R_-.$$

If we have a sequence x for which $x(n) = 0$ for $n < 0$ but is nonminimum-phase, then the sequence can be made minimum-phase by appropriate exponential weighting. We simply need to multiply $x(n)$ by α^n, where α is less than one and small enough to move the pole or zero with greatest magnitude inside the unit circle.

We see, then, that exponential weighting may be very useful because convolutions are preserved and it permits a more desirable pole-zero distribution. We should point out, however, that if the required value of α is too small, we shall often be troubled with rounding errors in carrying out such weighting on numbers stored in a computer.

A final point should be made. Exponential weighting clearly changes the complex cepstrum. As the value of α approaches the reciprocal of the magnitude of the pole or zero that is farthest from the origin, the complex cepstrum becomes zero for negative n. If, on the other hand, α is close to 1, it will not significantly affect the complex cepstrum, unless the z-transform of the input has poles or zeros on or close to the unit circle. Since convolutions are preserved in the weighted sequence, the complex cepstrum will always have the form

$$\hat{w}(n) = \hat{w}_1(n) + \hat{w}_2(n).$$

In general, there is not a simple relationship between $\hat{w}_1(n)$ and $\hat{x}_1(n)$; however, if the poles and zeros of $X_1(z)$ are inside the unit circle, then clearly those of $W_1(z)$ will also be inside the unit circle if $|\alpha| < 1$. Thus, in this special case, if

$$w_1(n) = \alpha^n x_1(n),$$

then

$$\hat{w}_1(n) = \alpha^n \hat{x}_1(n).$$

Although there may not always be such a simple relationship between $\hat{w}_1(n)$ and $\hat{x}_1(n)$ and $\hat{w}_2(n)$ and $\hat{x}_2(n)$, we do still have a way of recovering $x_1(n)$ or $x_2(n)$ if we are given $\hat{w}_1(n)$ or $\hat{w}_2(n)$. This is so because

$$x_1(n) = \alpha^{-n} w_1(n)$$

$$x_2(n) = \alpha^{-n} w_2(n).$$

These relations are useful in practice, since they allow us to work with exponentially weighted sequences as inputs to the system D. If we desire the original sequences at the output, we can simply unweight, using the relations above. In practice, this idea is useful if we are dealing with finite-length sequences such that $x(n) = 0$ for $n < 0$ and $n > M$, and if we can choose α so that α^M is not so small as to introduce excessive rounding error.

2.8 MORE GENERAL RATIONAL z-TRANSFORMS

Let us assume that the z-transform of the input to the system D has the form

$$X(z) = Az^r \frac{\prod_{k=1}^{m_i} \left(1-a_k z^{-1}\right) \prod_{k=1}^{m_o} (1-b_k z)}{\prod_{k=1}^{p_i} \left(1-c_k z^{-1}\right) \prod_{k=1}^{p_o} (1-d_k z)}. \tag{52}$$

In all of our previous results based on rational z-transforms, we assumed that A was positive and real and $r = 0$. This was to insure that arg $[X(z)]$ could be defined as single-valued and continuous. Clearly, there are many interesting sequences that do not have z-transforms of this form. For example, if we allow A to be positive or negative and $r \neq 0$, we can include most sequences of interest for computation. In fact, finite-length sequences have z-transforms of the form of Eq. 52, with the c_k and d_k all equal to zero. (Note that we have excluded zeros on the unit circle. These could be included in our discussion if we were willing to consider discontinuous phase curves and logarithmic infinities in the log magnitude. For simplicity, we shall take the point of view that zeros on the unit circle have been shifted inside by exponential weighting.) Let us now see how the results previously presented can be applied in this more general situation.

When $X(z)$ is actually the product of two or more z-transforms, we shall assume that each term in the product is written in the form of (52). Thus the constant A will be the product of the corresponding constants of the individual factors of $X(z)$. For example, if

$$X(z) = X_1(z) \cdot X_2(z),$$

then

$$A = A_1 \cdot A_2.$$

Clearly, A will be positive if A_1 and A_2 are both of the same sign, and A will be negative if the signs of A_1 and A_2 differ. That is, by consideration of the sign of A it will only be possible to determine the sign of A_1 relative to the sign of A_2. In most situations,

the appropriate signs will be clear from consideration of the source of the signals or, in many cases, the signs will not be important.

The constant term A contributes an integer multiple of π to the phase. Since we can only determine whether A is positive or negative, we normally test to see if A is positive or negative before computing the phase. If A is negative, we can change the sign of $X(e^{\sigma+j\omega})$ to effectively remove any contribution to the phase which is due to the sign of A. Whether A was positive or negative can be remembered if this information is of interest. The sign of A can be determined by noting that

$$X(1) = X(e^{j0}) = A \frac{\prod_{k=1}^{m_i} (1-a_k) \prod_{k=1}^{m_o} (1-b_k)}{\prod_{k=1}^{p_i} (1-c_k) \prod_{k=1}^{p_o} (1-d_k)} . \tag{53}$$

Since the a_k, b_k, c_k and d_k are all less than one in magnitude, all of the factors in (53) are positive; therefore, the sign of A is the sign of X(1).

Let us now consider the effect of the factor z^r in (52). Assuming that the phase is computed as specified in section 2.1, we can write formally

$$\hat{X}(z) = \log [z^r] + \log \left[A \frac{\prod_{k=1}^{m_i} \left(1-a_k z^{-1}\right) \prod_{k=1}^{m_o} (1-b_k z)}{\prod_{k=1}^{p_i} \left(1-c_k z^{-1}\right) \prod_{k=1}^{p_o} (1-d_k z)} \right] . \tag{54}$$

Thus, the complex cepstrum $\hat{x}$ consists of a component having all of the properties that we have previously discussed and a component that is due to the term $\log [z^r]$. To see how our results are modified by this term, let us consider the phase contribution for $z = e^{\sigma+j\omega}$. We are tempted to write

$$\log [z^r] = \log [e^{\sigma r} e^{j\omega r}] = \sigma r + j\omega r.$$

If we recall, however, that the phase angle must be periodic in ω (since it is the imaginary part of a z-transform), we see that $\arg [e^{(\sigma+j\omega)r}]$ must be defined as in Fig. 13. This factor then adds a nonanalytic component to the imaginary part of $\log [X(e^{\sigma+j\omega})]$.

Formally, the contribution to the complex cepstrum of this type of term is

$$\theta(n) = \frac{e^{\sigma n}}{2\pi} \int_{-\pi}^{\pi} r \log [e^{\sigma+j\omega}] e^{j\omega n} d\omega.$$

Performing the indicated integration shows that

$$\theta(n) = r e^{\sigma n} \frac{\cos \pi n}{n} \qquad n \neq 0 \tag{55a}$$

$$= \sigma r \qquad n = 0. \tag{55b}$$

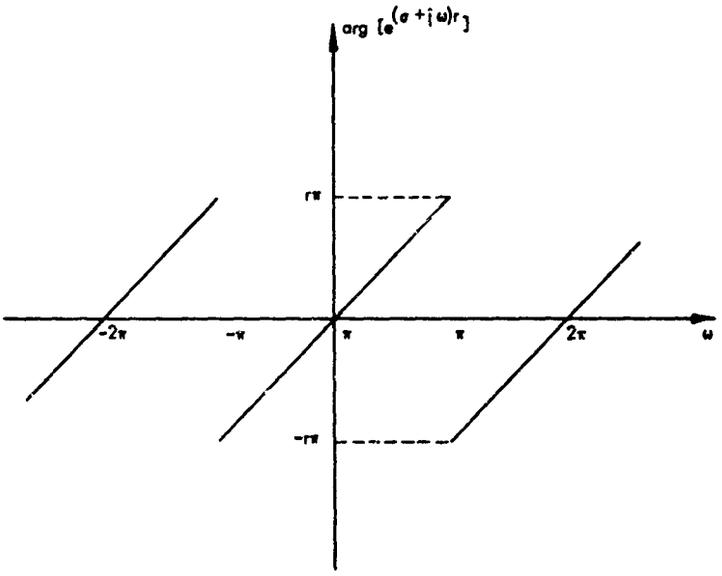


Fig. 13. Phase curve attributable to a factor z^r when $z = e^{\sigma+j\omega}$.

This sequence is stable only if the contour of integration is the unit circle ($\sigma = 0$). In fact, the sequence, strictly speaking, only has a discrete Fourier transform

$$\Theta(e^{j\omega}) = \sum_{n=-\infty}^{\infty} \theta(n)\, e^{-j\omega n},$$

since log z has no Laurent series expansion about $z = 0$. This situation is analogous to the continuous-time function $(1+t^2)^{-1}$, which has no two-sided Laplace transform but does have a Fourier transform.

Usually, we prefer to remove the linear-phase component before computing the complex cepstrum. This is easily done once the phase curve is computed, and clearly its removal simply corresponds to a shift of r samples in the input sequence. This value of r can be saved and used to shift the output of D^{-1}, if this is appropriate. The parameter r is very much like the sign of A, in that if X(z) is the product of $X_1(z)$ and $X_2(z)$, each having the form of Eq. 52, then $r = r_1 + r_2$ and it will only be possible to determine r_1 and r_2 from consideration of the source of the sequence x.

Thus, the complex cepstrum of a sequence whose z-transform is of the form of (52) is normally obtained as follows. Choose the contour C to be the unit circle, and find the contributions that are due to all of the factors except z^r, using either the power series expansion or the integral relations. We may then simply add to this the component $\theta(n)$ given by Eq. 55. For example, we could write for sequences whose z-transforms are of the form of (52), and choosing $\sigma = 0$

$$\hat{x}(n) = \frac{r \cos \pi n}{n} - \frac{1}{2\pi jn} \oint_C z \frac{X'(z)}{X(z)} z^{n-1}\, dz \qquad n \neq 0 \tag{56}$$

$$= \log A \qquad n = 0. \tag{57}$$

We note that the integral in (56) may be evaluated for $n = 0$ if we do not divide by n. In fact, it is easily shown from the principle of the argument that the value of r can be obtained from

$$r = \frac{1}{2\pi j} \oint_C \frac{X'(z)}{X(z)}\, dz.$$

2.9 EXAMPLES OF COMPLEX CEPSTRA

We shall give several examples of the analytical determination of the complex cepstrum. These examples, although simple, are somewhat typical of the kinds of sequences that will be encountered in practice.

Example 1: Minimum-phase sequence

Let $x(n)$ have the form

$$x(n) = 0 \qquad n < 0$$

$$= a^n \qquad n \geqslant 0,$$

where $|a| < 1$.

The z-transform of this sequence is

$$X(z) = \sum_{n=0}^{\infty} a^n z^{-n} = \frac{1}{1 - az^{-1}} \qquad \text{for } |z| > |a|.$$

The complex logarithm of $X(z)$ is

$$\hat{X}(z) = -\log (1 - az^{-1}).$$

Since $|a| < 1$, we see that x is a minimum-phase sequence.

Let us choose the region of convergence of $\hat{X}(z)$ to include the unit circle so that this can be the contour of integration in determining the complex cepstrum. In this case we can obtain $\hat{x}(n)$ by three different methods.

(a) Integral Formula

Since we observe that $\arg [X(e^{j\omega})]$ is continuous everywhere for this example, $\hat{x}(n)$ is determined by the equation

$$\hat{x}(n) = \frac{1}{2\pi jn} \oint_C \frac{-zX'(z)}{X(z)} z^{n-1}\, dz,$$

where $-zX'(z)/X(z)$ is given by

$$\frac{-zX'(z)}{X(z)} = \frac{az^{-1}}{1 - az^{-1}}.$$

Thus we see that

$$\hat{x}(n) = \frac{1}{2\pi jn} \oint \frac{az^{-1}}{1 - az^{-1}} z^{n-1}\, dz = \frac{1}{2\pi jn} \oint \frac{az^{n-1}}{z - a}\, dz.$$

This integral can be evaluated by using the residue theorem, to obtain

$$\begin{aligned} \hat{x}(n) &= \frac{a^n}{n} && n > 0 \\ &= \log(1) = 0 && n = 0 \\ &= 0 && n < 0 \end{aligned}$$

(b) Power Series Expansion for $\hat{X}(z)$

Recall that $\hat{X}(z)$ is given by

$$\hat{X}(z) = -\log(1-az^{-1}) = \sum_{n=1}^{\infty} \frac{a^n}{n} z^{-n},$$

where the power series expansion for $-\log(1-az^{-1})$ is valid in the region $|z| > |a|$. By definition, $\hat{X}(z)$ is also given by

$$\hat{X}(z) = \sum_{n=-\infty}^{\infty} \hat{x}(n)\, z^{-n}.$$

By comparison of the two power series, we see that

$$\begin{aligned} \hat{x}(n) &= 0 && n \leq 0 \\ &= \frac{a^n}{n} && n > 0. \end{aligned}$$

(c) Recursive Formula

Let us compute several values of $\hat{x}(n)$, using the recursive formula. Since $x(0) = 1$,

$\hat{x}(0) = \log x(0) = 0.$

By applying the recursion formula, we obtain

$$\hat{x}(1) = \frac{x(1)}{x(0)} = a$$

$$\hat{x}(2) = \frac{x(2)}{x(0)} - \frac{1}{2}\hat{x}(1)\,x(1) = a^2 - \frac{1}{2}a^2 = \frac{a^2}{2}$$

$$\hat{x}(3) = \frac{x(3)}{x(0)} - \frac{1}{3}(\hat{x}(1)x(2)+2\hat{x}(2)x(1)) = a^3 - \frac{1}{3}(a^3+a^3) = \frac{a^3}{3}.$$

We note that because the sequence is minimum-phase, the complex cepstrum is zero for $n < 0$, and that our result agrees with that obtained by using the Hilbert transform.

Example 2: Nonminimum-phase finite length sequence

Let $x(n)$ be given by

$$x(n) = 0 \qquad \text{for } n < 0 \text{ and } n \geq M$$

$$= b^n \qquad 0 \leq n < M.$$

The z-transform is given by

$$X(z) = \sum_{n=0}^{M-1} b^n z^{-n} = \frac{1 - b^M z^{-M}}{1 - bz^{-1}}.$$

Let us assume that $|b| > 1$. Then $X(z)$ has $M - 1$ zeros located as shown in Fig. 14.

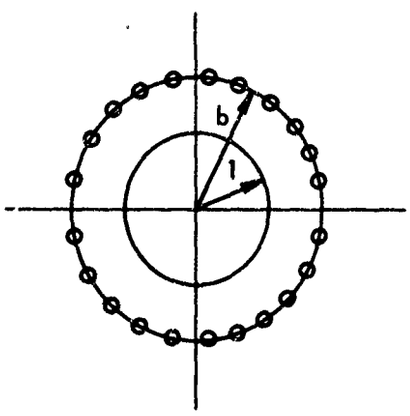


Fig. 14. Zeros of $X(z) = \dfrac{1 - b^M z^{-M}}{1 - bz^{-1}}$.

Since the sequence is finite-length, the region of convergence of $X(z)$ is the entire z plane except for $z = 0$. If we write $X(z)$ in the form of Eq. 52, we obtain

$$X(z) = b^{M-1} z^{-(M-1)} \frac{(1-b^{-M}z^{M})}{(1-b^{-1}z)}.$$

The complex logarithm of this expression is

$$\hat{x}(z) = \log b^{M-1} + \log z^{-(M-1)} + \log (1-b^{-M}z^{M}) - \log (1-b^{-1}z).$$

If the region of convergence of $\hat{X}(z)$ is chosen to be the region $|z| < |b|$, then

$$\hat{X}(z) = \log b^{M-1} + \log z^{-(M-1)} + \sum_{n=1}^{\infty} \frac{b^{-n}}{n} z^{n} - \sum_{n=1}^{\infty} \frac{b^{-Mn}}{n} z^{Mn}.$$

If we introduce the symbol $\delta(n)$

$$\delta(n) = 1 \qquad n = 0$$
$$\quad = 0 \qquad n \neq 0,$$

we can write the inverse transform on the unit circle as

$$\hat{x}(n) = (\log b^{M-1})\, \delta(n) + \sum_{k=-\infty}^{\infty} \frac{(1-M)\cos \pi k}{k} \delta(n-k) + \sum_{k=-\infty}^{-1} \frac{b^{Mk}}{k} \delta(n-kM)$$
$$- \sum_{k=-\infty}^{-1} \frac{b^{k}}{k} \delta(n-k).$$

Therefore we observe that we obtain a stable sequence that is nonzero for both positive and negative n. We can also see that if the sequence were shifted to the left by $M-1$ samples, we would remove the term

$$\sum_{k=-\infty}^{\infty} \frac{(1-M)\cos k}{k} \delta(n-k)$$

from the complex cepstrum. That is, the sequence whose values are

$$s(n) = x(n+M-1)$$

will have the z-transform

$$S(z) = b^{M-1} \frac{(1-b^{-M}z^{M})}{(1-b^{-1}z)}.$$

Thus we can easily see that the complex cepstrum $\hat{s}(n)$ will be zero $n > 0$. This is so

because the shifted sequence is maximum-phase. The shifting of sequences to remove linear phase components is quite important and will be discussed in detail in Section III.

Example 3: Repeated pulses

Suppose that the sequence x has values

$$x(n) = s(n) + s(n-n_o) + \ldots + s(n-(P-1)n_o),$$

where s is an aperiodic sequence like those of Examples 1 and 2. The sequence x can also be expressed as

$$x = s \otimes u,$$

where u is a sequence whose values are given by

$$u(n) = \sum_{k=0}^{P-1} \delta(n-kn_o).$$

The z-transform $X(z)$ will have the form

$$X(z) = S(z) \cdot U(z),$$

where $U(z)$ is

$$U(z) = \sum_{k=0}^{P-1} z^{-n_o k} = \frac{\left(1-z^{-Pn_o}\right)}{\left(1-z^{-n_o}\right)}.$$

The region of convergence of $U(z)$ is the entire z-plane except for $z = 0$. We note that $U(z)$ has zeros at equal angular spacing around the unit circle.

The complex logarithm of $X(z)$ is

$$\hat{X}(z) = \log [S(z)] + \log [U(z)] = \hat{S}(z) + \hat{U}(z).$$

Let us now choose a region of convergence for $\hat{X}(z)$. In this case it is not possible to choose the region of convergence to contain the unit circle, since $U(z)$ has all its zeros there. This means that if $S(z)$ is non minimum-phase, it will not be possible to obtain a stable complex cepstrum. We can remove this difficulty by using exponential weighting of the sequence x. We know that if

$$x_1(n) = a^n x(n),$$

then

$$X_1(z) = X(a^{-1}z) = S(a^{-1}z) \cdot U(a^{-1}z) = S_1(z) \cdot U_1(z),$$

where $S_1(z)$ is the z-transform of the sequence whose values are

$$s_1(n) = a^n s(n),$$

and $U_1(z)$ is the z-transform of the sequence

$$u_1(n) = a^n u(n) = \sum_{k=0}^{P-1} a^{n_o k} \delta(n-kn_o).$$

Thus $U_1(z)$ is

$$U_1(z) = \frac{\left(1-a^{Pn_o} z^{-Pn_o}\right)}{\left(1-a^{n_o} z^{-n_o}\right)}.$$

Now, we see that the complex logarithm of $X_1(z)$ is

$$\hat{X}_1(z) = \log [S_1(z)] + \log [U_1(z)] = \hat{S}_1(z) + \log \left(1-a^{Pn_o} z^{-Pn_o}\right) - \log \left(1-a^{n_o} z^{-n_o}\right).$$

Since the zeros of $U_1(z)$ now lie on a circle with radius $|a|$, we see that the sequence u_1 can be made minimum-phase by making $|a| < 1$. In this case it would be possible to use the unit circle as the contour of integration for obtaining $\hat{x}_1(n)$, which would be

$$\hat{x}_1(n) = \hat{s}_1(n) + \sum_{k=1}^{\infty} \frac{a^{n_o k}}{k} \delta(n-kn_o) - \sum_{k=1}^{\infty} \frac{a^{Pn_o k}}{k} \delta(n-Pn_o k).$$

Suppose that s is the sequence of Example 2, after shifting it left by M - 1 samples, that is,

$$s(n) = b^{n+M-1} \qquad M-1 < n \leq 0$$

$$= 0 \qquad \text{elsewhere.}$$

Then the sequence x would appear as in Fig. 15a, where P = 4 and $M > n_o$. The samples of the discrete-time function have been connected by a smooth curve for convenience in plotting. The weighted sequence x_1 is shown in Fig. 15b, and the resulting complex cepstrum is shown in Fig. 15c.

This simple example points out several things that are true in general:

1. The input x_1 was the convolution of a minimum-phase sequence with a maximum-phase sequence. The resulting complex cepstrum consists of a part that is zero for $n < 0$ because of the minimum-phase part of the input and a part that is zero for $n > 0$ because of the maximum phase part of the input. This is always true because any sequence can be expressed as the convolution of a minimum-phase sequence and a maximum-phase sequence (with possibly some time shift). In some cases either the

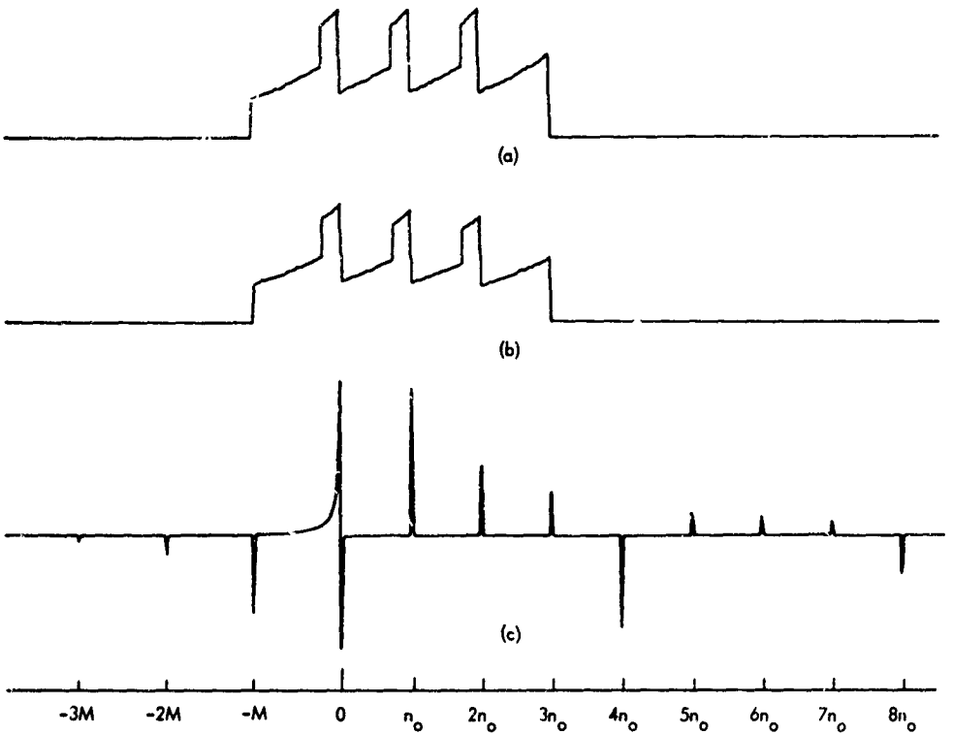


Fig. 15. (a) Input sequence x. (b) Exponentially weighted sequence $x_1(n) = a^n x(n)$. (c) Complex cepstrum of x_1.

minimum-phase part or the maximum-phase part may be all that is of interest, while in general this may not be true. When it is true, considerable simplification results if we take advantage of the properties of minimum-phase sequences.

2. The input sequence u_1 had samples spaced at intervals of n_o, rather than 1. The resulting complex cepstrum has its samples at the same spacing n_o. This can be shown to be true in general if the spacing of the samples is uniform. It is also true approximately if the samples are unequally spaced, but it is difficult to obtain precise results on this problem. Examples of this are given in Section V.

3. The input sequence s_1 was "pulselike"; that is, it had most of its samples concentrated in a small region relative to the total duration of u_1. The samples of s_1 were spaced at unit intervals. The complex cepstrum of s_1 is seen to approach 0 as $(ab)^{-n}/n$, so that it dies out at a fairly rapid rate. We have seen that it is true in general that the complex cepstrum dies out at least as rapidly as $|1/n|$ for all n.

2.10 LINEAR SYSTEM IN THE CANONIC REPRESENTATION

We have discussed in detail the analysis and realization of the characteristic system D (and therefore its inverse D^{-1}). We wish now to discuss the general type of linear system that has been found useful in separating convolved signals. We shall leave specific examples to be covered in the sections on applications.

As a very general comment, we can say that the type of signals for which homomorphic deconvolution has thus far been found useful are those that are convolutions of signals whose complex cepstra, in some sense, do not overlap. An obvious example of this is when one signal is minimum-phase and the other is maximum-phase. This was the case in Example 3. The second basic situation is also indicated by Example 3. We observe that the complex cepstrum of the "impulse train" u_1 has isolated samples occurring with spacing n_o. Suppose that s_1 is a sequence whose complex cepstrum dies out rapidly so that $s_1(n)$ is small for $n \leq n_o$ (s_1 may not in general be maximum-phase as in the example). If we have as the input

$$x_1 = s_1 \otimes u_1,$$

then we see that the complex cepstra $\hat{s}_1$ and $\hat{u}_1$ will be in a sense separated in "time" in the complex cepstrum of x_1.

Both of these situations suggest that the kind of linear system that we should use is of the form

$$\hat{y}(n) = \ell(n)\, \hat{x}(n). \tag{58}$$

Such systems will be called frequency-invariant, in analogy with time-invariant linear systems in which we multiply z-transforms and convolve time functions. For frequency invariant linear systems we multiply time functions as in Eq. 58 and, therefore, we convolve frequency functions as in

$$\hat{Y}(e^{j\omega}) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \hat{X}(e^{j\xi})\, L(e^{j(\omega-\xi)})\, d\xi, \tag{59}$$

where $L(z)$ is the z-transform of the sequence whose values are $\ell(n)$.

As a general comment, let us consider an interesting possibility that results if the input is of the form

$$\hat{x}(n) = a\hat{x}_1(n) + b\hat{x}_2(n).$$

The transform of this equation is

$$\hat{X}(e^{j\omega}) = a\hat{X}_1(e^{j\omega}) + b\hat{X}_2(e^{j\omega}).$$

(See the Appendix for a discussion of scalar multiplication.) Suppose that we filter the real and imaginary parts of $\hat{X}(e^{j\omega})$ with different linear systems $L_r(e^{j\omega})$ and $L_i(e^{j\omega})$, respectively. Thus $\hat{Y}(e^{j\omega})$ would be of the form

$$\hat{Y}(e^{j\omega}) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \hat{X}_r(e^{j\xi})\, L_r(e^{j(\omega-\xi)})\, d\xi + \frac{1}{2\pi} \int_{-\pi}^{\pi} j\hat{X}_i(e^{j\xi})\, L_i(e^{j(\omega-\xi)})\, d\xi.$$

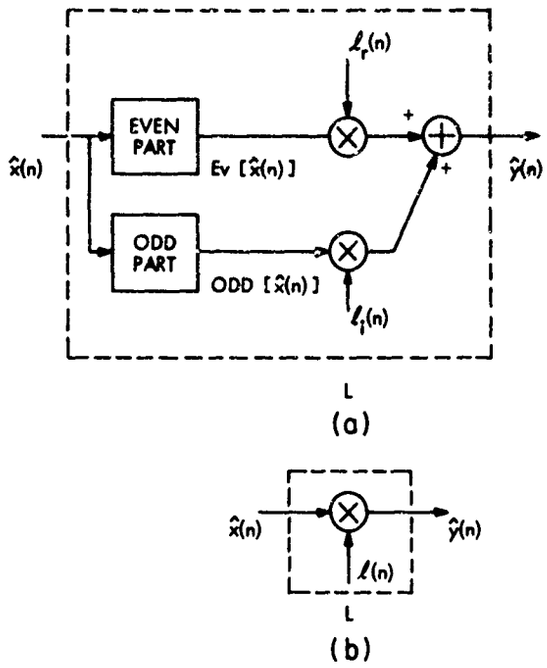


Fig. 16. (a) Frequency-invariant system that is linear for real scalars.
(b) General linear frequency-invariant linear system.

It can be shown that if and only if a and b are real numbers,

$$\hat{Y}(e^{j\omega}) = a\hat{Y}_1(e^{j\omega}) + b\hat{Y}_2(e^{j\omega}).$$

Thus for real scalars we may filter the magnitude and phase with separate filters. Therefore we see that in general the linear system can be of the form

$$\hat{y}(n) = \ell_r(n)\ \mathrm{Ev}\ [\hat{x}(n)] + \ell_i(n)\ \mathrm{Odd}\ [\hat{x}(n)], \tag{60}$$

where

$$\hat{x}(n) = \mathrm{Ev}\ [\hat{x}(n)] + \mathrm{Odd}\ [\hat{x}(n)],$$

and $\ell_r(n)$ and $\ell_i(n)$ are the inverse transforms of $L_r(e^{j\omega})$ and $L_i(e^{j\omega})$, respectively. The operations suggested by Eq. 60 are summarized in Fig. 16. Figure 16a illustrates the general case and Fig. 16b, the case

$$\ell_r(n) = \ell_i(n) = \ell(n).$$

III. COMPUTATIONAL CONSIDERATIONS IN HOMOMORPHIC DECONVOLUTION

We have given a detailed analysis of the canonic form for homomorphic deconvolution. This analysis was carried out for discrete convolutions and it leaned heavily on the z-transform both in the analysis and realization of the system. When we turn to the actual computational realization (in the form of digital-computer programs), we find that we must somewhat modify the results of Section II. The main reason for this is that the z-transform (usually its unit circle evaluation) is a function of a continuous complex variable z (or ω). Since digital computers deal with finite collections of numbers rather than functions, it is clear that we must be content with only a finite number of values of the z-transform. Thus we are led to the study of the sampled z-transform and its properties.

The other major consideration is the calculation of the phase of the sampled z-transform. We shall apply the results of Section II to show what properties the sampled phase function must have and then we shall show an algorithm with which we can compute the proper phase.

Therefore our present major purpose is to show how the ideas of Section II can be translated into programs for a digital computer. These programs will constitute our realization of the canonic system of Fig. 4.

3.1 SAMPLED z-TRANSFORM

As we can see, the use of the z-transform is very convenient in the analysis of homomorphic deconvolution. If we wish, however, to use a computer to evaluate the relations that we obtained in Section II, we must deal with only a finite number of values of the z-transform. For the same reason, we shall be limited at the outset to finite-length sequences.

Let us consider a sequence x whose z-transform has a region of convergence including the unit circle. (This will always be true for finite length sequences.) We can therefore evaluate X(z) for $z = e^{j\omega}$ so that we obtain

$$X(e^{j\omega}) = \sum_{n=-\infty}^{\infty} x(n)\, e^{-j\omega n}. \tag{61}$$

Henceforth, (61), which is a function of the continuous variable ω, will be referred to as the Fourier transform (FT) of the sequence x. The inverse Fourier transform (IFT) of $X(e^{j\omega})$ is

$$x(n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} X(e^{j\omega})\, e^{j\omega n}\, d\omega = \frac{1}{2\pi} \int_{0}^{2\pi} X(e^{j\omega})\, e^{j\omega n}\, d\omega. \tag{62}$$

We note that the limits of integration in (62) can be any convenient interval of length 2π,

since both $X(e^{j\omega})$ and $e^{j\omega n}$ are periodic with period 2π. For computation on a digital computer, we must be content with only a finite collection of samples of Eq. 61. This leads us to consider a sampling theorem for the transform $X(e^{j\omega})$. Let us suppose that we have samples at exactly N values of ω, that is, at N points around the unit circle spaced at equal angular increments of $2\pi/N$. Thus we obtain the sequence of samples

$$X\left(e^{j\frac{2\pi}{N}k}\right) = \sum_{n=-\infty}^{\infty} x(n)\, e^{-j\frac{2\pi}{N}kn} \qquad \text{for } k = 0, 1, \ldots, N-1. \tag{63}$$

(We choose to use positive values of k for computational convenience.) Corresponding to (62), we have

$$\underset{\sim}{x}(n) = \frac{1}{N}\sum_{k=0}^{N-1} X\left(e^{j\frac{2\pi}{N}k}\right) e^{j\frac{2\pi}{N}kn}. \tag{64}$$

To see how $\underset{\sim}{x}(n)$ is related to $x(n)$, let us substitute (63) in (64). The result is

$$\underset{\sim}{x}(n) = \frac{1}{N}\sum_{k=0}^{N-1}\left(\sum_{m=-\infty}^{\infty} x(m)\exp\left(-j\frac{2\pi}{N}km\right)\right)\exp\left(j\frac{2\pi}{N}kn\right).$$

If we interchange the order of summation, we obtain

$$\underset{\sim}{x}(n) = \sum_{m=-\infty}^{\infty} x(m)\left(\frac{1}{N}\sum_{k=0}^{N-1}\exp\left(j\frac{2\pi}{N}k(n-m)\right)\right). \tag{65}$$

Let us define

$$d(n-m) = \frac{1}{N}\sum_{k=0}^{N-1}\exp\left(j\frac{2\pi}{N}k(n-m)\right). \tag{66}$$

It can be shown that the sequence d is periodic with period N and that it can be represented as

$$d(n-m) = \sum_{r=-\infty}^{\infty}\delta(n-m+rN), \tag{67}$$

where

$$\delta(n) = \begin{cases} 0 & n \neq 0 \\ 1 & n = 0. \end{cases}$$

Therefore (65) becomes

$$\underset{\sim}{x}(n) = \sum_{m=-\infty}^{\infty} x(m)\, d(n-m) = \sum_{r=-\infty}^{\infty} x(n+rN). \tag{68}$$

Equation 68 shows that $\underset{\sim}{x}(n)$ is periodic with period N, and that it is possible for $\underset{\sim}{x}(n)$ to be equal to x(n) for certain values of n if the original sequence is of finite length. For example, let us assume that the sequence x has values x(n) such that x(n) = 0 for n < 0 and n > M. It is clear from (68) that if N >M, then

$$\underset{\sim}{x}(n) = x(n) \qquad \text{for } 0 \leq n < N.$$

On the other hand, if N ≤ M, we encounter aliasing in attempting to return to the sequence from the samples of $X(e^{j\omega})$. This is shown in the simple example of Fig. 17. Figure 17a shows the original sequence x, wherein we can see that the length of the sequence is 5 samples (M=5). Using (68), we have plotted the sequence $\underset{\sim}{x}$ for two different values of N. In Fig. 17b we show the sequence $\underset{\sim}{x}$ when we have sampled $X(e^{j\omega})$

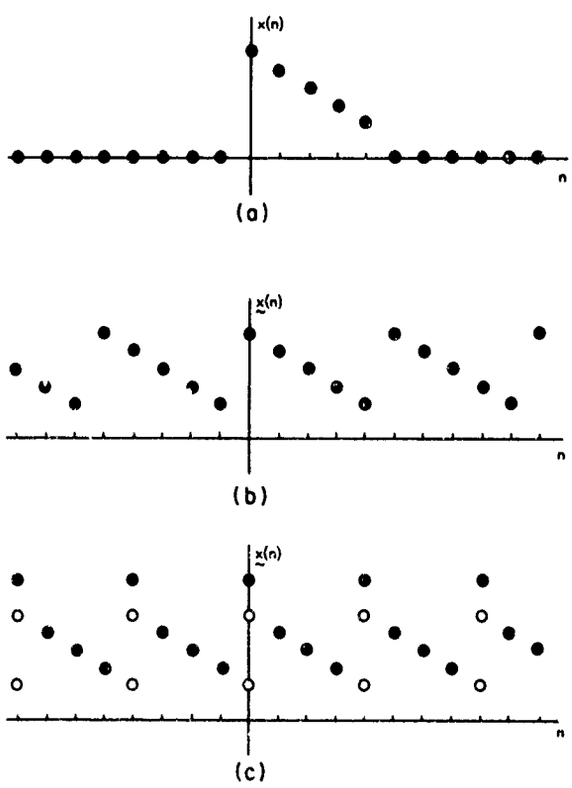


Fig. 17. (a) Finite-length sequence of 5 samples. (b) Periodic sequence corresponding to sampling the z-transform at 5 points on the unit circle. (c) Periodic sequence corresponding to sampling the z-transform at 4 points on the unit circle. (Open circles indicate that two values overlap at n = 0, ±4, ±8, ... This effect is called aliasing.)

five times (N=5). In the interval $0 \leq n < 5$, $\underset{\sim}{x}(n) = x(n)$, and the sequence $\underset{\sim}{x}$ is periodic with N = 5. In Fig. 17c $\underset{\sim}{x}$ is shown for the case in which we have sampled $X(e^{j\omega})$ only four times. We note that $\underset{\sim}{x}$ is periodic with period N = 4; however, two points of the original sequence occur at the same value of n in $\underset{\sim}{x}$ so that $\underset{\sim}{x}(0) \neq x(n)$. This overlap in the periodic sequence x is known as aliasing. Clearly, the way to avoid aliasing is to insure that N > M; that is, we must sample $X(e^{j\omega})$ at a high enough rate.

If we assume that N > M, we can write the following pair of relationships for finite length sequences:

$$\underset{\sim}{X}(k) = \sum_{n=0}^{N-1} \underset{\sim}{x}(n)\, e^{-j\frac{2\pi}{N}kn} \qquad k = 0, 1, \ldots, N-1, \tag{69}$$

$$\underset{\sim}{x}(n) = \frac{1}{N} \sum_{k=0}^{N-1} \underset{\sim}{X}(k)\, e^{j\frac{2\pi}{N}kn} \qquad n = 0, 1, \ldots, N-1. \tag{70}$$

We note that both $\underset{\sim}{x}$ and $\underset{\sim}{X}$ are periodic with period N. That is,

$$\underset{\sim}{x}(n) = \underset{\sim}{x}(n+rN) \qquad \text{for } r = 0, \pm 1, \pm 2, \ldots$$

$$\underset{\sim}{X}(k) = \underset{\sim}{X}(k+rN) \qquad \text{for } r = 0, \pm 1, \pm 2, \ldots .$$

Another way of expressing this fact is to say that in Eq. 69 and Eq. 70, all integers n, k, and kn are to be interpreted modulo N. Equations 69 and 70 have been referred to as a Discrete Fourier Transform (DFT) pair.[16]

Let us now consider how (69) and (70) can be used in the realization of the characteristic system D. We shall replace all z-transforms by Eq. 69 (the DFT) and all inverse z-transforms by the inverse discrete Fourier transform (IDFT) of Eq. 70. Since our interest is in convolutions of sequences, we must consider the effect of sampling the z-transform when the input is a convolution. Let us assume that x_1 and x_2 are finite-length sequences such that

$$x_1(n) = 0 \qquad \text{for } n < 0 \text{ and } n > M_1,$$

$$x_2(n) = 0 \qquad \text{for } n < 0 \text{ and } n > M_2.$$

The convolution of x_1 and x_2 has values given by

$$x(n) = \sum_{r=0}^{n} x_1(n-r)\, x_2(r) = \sum_{r=0}^{n} x_1(r)\, x_2(n-r). \tag{71}$$

We can see that the sequence x is also finite in length and that

$$x(n) = 0 \qquad \text{for } n < 0 \text{ and } n > M_1 + M_2.$$

The FT of the sequence x is $X(e^{j\omega})$ and is given by

$$X(e^{j\omega}) = X_1(e^{j\omega}) \cdot X_2(e^{j\omega}).$$

Let us assume that we sample $X(e^{j\omega})$ at N points to obtain the samples

$$X\left(e^{j\frac{2\pi}{N}k}\right) = X_1\left(e^{j\frac{2\pi}{N}k}\right) \cdot X_2\left(e^{j\frac{2\pi}{N}k}\right), \qquad k = 0, 1, \ldots, N-1. \tag{72}$$

We have seen that if N is not large enough, aliasing will occur. To see the nature of this aliasing for convolutions, let us note that (72) can be written in terms of the DFT.

$$\underset{\sim}{X}(k) = \underset{\sim}{X}_1(k) \cdot \underset{\sim}{X}_2(k) \qquad k = 0, 1, \ldots, N-1, \tag{73}$$

where k is taken modulo N. The IDFT of (73) can be shown to be

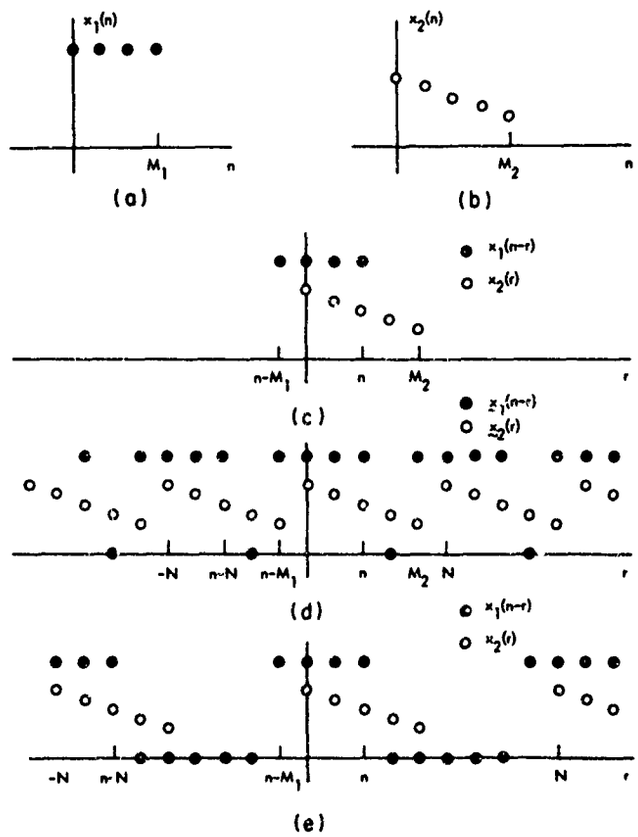


Fig. 18. (a) and (b) Aperiodic finite-length sequences. (c) Discrete convolution of x_1 and x_2. (d) Periodic sequences for circular convolution when $N < M_1 + M_2$. (e) Periodic sequences for circular convolution when $N > M_1 + M_2$.

$$\underset{\sim}{x}(n) = \sum_{r=0}^{N-1} \underset{\sim}{x}_1(n-r)\, \underset{\sim}{x}_2(r) = \sum_{r=0}^{N-1} \underset{\sim}{x}_1(r)\, \underset{\sim}{x}_2(n-r), \tag{74}$$

where r and n-r are both to be taken modulo N, and n = 0, 1, . . . , N-1. That is, $\underset{\sim}{x}$ is the result of a circular or periodic convolution as opposed to the original sequence x which was the result of an aperiodic convolution. This point is illustrated by Fig. 18. In Fig. 18a and 18b we show two finite length sequences of lengths M_1 and M_2, respectively. Note that $M_1 < M_2$. Figure 18c shows $x_1(n-r)$ and $x_2(r)$ plotted as a function of r, as is required for convolution. The value of x(n) is obtained by adding the products $x_1(n-r)\, x_2(r)$. From this figure, we can clearly see that x(n) will be zero for $n < 0$ and $n > M_1 + M_2$, since the sequences do not overlap for these values. Figure 18d shows $\underset{\sim}{x}_1(n-r)$ and $\underset{\sim}{x}_2(r)$ as would be obtained for $N = M_2 + 1$. Although neither of the sequences $\underset{\sim}{x}_1$ or $\underset{\sim}{x}_2$ differ from x_1 and x_2, respectively, in the interval $0 \leq n < N$, it is clear that the circular convolution will not be equal to the aperiodic convolution for this value of N. Figure 18e shows $\underset{\sim}{x}_1(n-r)$ and $\underset{\sim}{x}_2(r)$ for $N = M_1 + M_2 + 1$. In this case, it is clear that the periodic convolution $\underset{\sim}{x}$ will equal the aperiodic convolution x in the interval $0 \leq n < N$. Thus, if we sample the z-transform at a high enough rate, the DFT $\underset{\sim}{X}(k)$ can represent an aperiodic convolution with no aliasing.

Let us now consider how we can compute the complex cepstrum using the DFT. We recall that we have defined

$$\hat{X}(e^{j\omega}) = \log\left[X(e^{j\omega})\right].$$

If we sample $\hat{X}(e^{j\omega})$ at N equally spaced values of ω, we obtain

$$\hat{X}\left(e^{j\frac{2\pi}{N}k}\right) = \log\left[X\left(e^{j\frac{2\pi}{N}k}\right)\right] \qquad k = 0, 1, \ldots, N-1.$$

The samples of $X(e^{j\omega})$, of course, can be directly evaluated by using the DFT relationship of (69). Therefore, we can write

$$\underset{\sim}{\hat{X}}(k) = \log\left[\underset{\sim}{X}(k)\right] \qquad k = 0, 1, \ldots, N-1, \tag{75}$$

If we apply the IDFT, we obtain

$$\underset{\sim}{\hat{x}}(n) = \frac{1}{N} \sum_{k=0}^{N-1} \log\left[X(k)\right] e^{j\frac{2\pi}{N}kn} \qquad n = 0, 1, \ldots, N-1. \tag{76}$$

In general, $\hat{x}$ is not a finite-length sequence so that we should expect some aliasing to occur in $\underset{\sim}{\hat{x}}$. Equation 68 shows, in fact, that we must write $\underset{\sim}{\hat{x}}(n)$ as

$$\underset{\sim}{\hat{x}}(n) = \sum_{r=-\infty}^{\infty} \hat{x}(n+rN), \tag{77}$$

where $\hat{\underset{\sim}{x}}(n)$ is periodic with period N. The aliasing of the complex cepstrum may be a problem in some cases and in others it may not. We recall that in general

$$|\hat{x}(n)| < A\left|\frac{1}{n}\right| \qquad \text{for all } n \neq 0,$$

so that it is possible, by making N large, to eliminate most of the aliasing of the complex cepstrum. By this we mean that if $x(n) = 0$ for $n < 0$ and $n > M$, then by choosing $N \gg M$ and defining

$$\underset{\sim}{x}(n) = x(n) \qquad 0 \leq n < M$$

$$= 0 \qquad M \leq n < N,$$

there will not be as much overlap between periods of $\hat{\underset{\sim}{x}}$ as if we had chosen $N = M + 1$, as it is clearly possible to do. Therefore the aliasing of $\hat{\underset{\sim}{x}}$ is minimized by choosing N as large as possible consistent with the computer storage and computation time constraints and then padding the $M + 1$ samples of x with zeros.

The numerical operations of computing the complex cepstrum are summarized in the following equations.

$$\underset{\sim}{X}(k) = \sum_{n=0}^{N-1} x(n)\, e^{-j\frac{2\pi}{N}kn} \qquad k = 0, 1, \ldots, N-1, \tag{78a}$$

$$\hat{\underset{\sim}{X}}(k) = \log\left[\underset{\sim}{X}(k)\right] \qquad k = 0, 1, \ldots, N-1, \tag{78b}$$

$$\hat{\underset{\sim}{x}}(n) = \frac{1}{N}\sum_{k=0}^{N-1} \hat{\underset{\sim}{X}}(k)\, e^{j\frac{2\pi}{N}kn} \qquad n = 0, 1, \ldots, N-1, \tag{78c}$$

We recognize that the complex cepstrum that we compute by using these equations will differ because of aliasing from the theoretical complex cepstrum obtained through the z-transform.

3.2 FAST FOURIER TRANSFORM

In 1965, Cooley and Tukey[14] disclosed an algorithm for high-speed calculation of the DFT. Since that time, there has been tremendous interest in the application of this algorithm in many diverse areas. One of the fields in which it has been used with great success is that of digital filtering or waveform processing.[16,17] In fact, the availability of this algorithm allows us to consider realizing this scheme for deconvolution.

To see the nature of this new method of evaluation of Fourier transforms, let us recall the DFT pair

$$\underset{\sim}{X}(k) = \sum_{n=0}^{N-1} \underset{\sim}{x}(n)\, e^{-j\frac{2\pi}{N}kn} \qquad k = 0, 1, \ldots, N-1, \tag{79a}$$

$$\underset{\sim}{x}(n) = \frac{1}{N} \sum_{k=0}^{N-1} \underset{\sim}{X}(k)\, e^{j\frac{2\pi}{N}kn} \qquad n = 0, 1, \ldots, N-1. \tag{79b}$$

The fast Fourier transform (FFT) is just Eqs. 79a or 79b done in a high-speed way. We note that to evaluate Eqs. 79 in a straightforward manner involves N^2 complex multiplications and additions. Cooley and Tukey showed that if $N = 2^m$, it is possible to evaluate either (79a) or (79b) by using m iterations of a process involving N complex multiplications and additions. Therefore the total number of operations is $N \log_2 N$, rather than N^2. Clearly, if N is only moderately large, the amount of computation (and thus computing time) is considerably reduced by using the FFT algorithm.

Since the publications[14-17] on the subject of the FFT have mushroomed in the several years since the appearance of the original Cooley-Tukey paper, we shall not discuss the algorithm and its programming. We shall, however, be interested in the properties of the FFT (or DFT), many of which are slightly different from corresponding ones for the z-transform or Fourier transform of a sequence. This difference is usually a result of the periodicity of both $\underset{\sim}{X}(k)$ and $\underset{\sim}{x}(n)$. We have already seen an example of this in the case of convolution. A table of the useful relations and symmetries is given by Gentleman and Sande.[15]

3.3 PROPERTIES OF THE SAMPLED-PHASE CURVES

We shall consider the problem of computing the samples of $\arg[X(e^{j\omega})]$ from the samples of $X(e^{j\omega})$. We have shown that $\arg[X(e^{j\omega})]$ must be continuous for $-\pi < \omega < \pi$ and that it must be odd and periodic with period 2π. We recall that for rational z-transforms, $\arg[X(e^{j\omega})]$ is in general discontinuous at $\omega = \pm\pi, \pm 3, \ldots$, because of the linear phase component. These conditions imply a similar set of conditions on the samples of $\arg[X(e^{j\omega})]$, that is, on $\arg[\underset{\sim}{X}(k)]$, for $k = 0, 1, 2, \ldots, N-1$. These conditions are given below under the assumption that we have sampled $X(e^{j\omega})$ at a sufficiently high rate so that the complex cepstrum will not be severely aliased. These conditions are as follows.

C1. The inequality

$$|\arg[\underset{\sim}{X}(k)] - \arg[\underset{\sim}{X}(k+1)]| < \epsilon$$

must hold for $0 \leq k < N/2 - 1$ and for $N/2 + 1 \leq k < N-1$, where ϵ is a tolerance depending on N (that is, the rate of sampling of $X(e^{j\omega})$).

C2. $\arg[\underset{\sim}{X}(k)]$ is an odd function of k; that is,

$$\arg[\underset{\sim}{X}(k)] = -\arg[\underset{\sim}{X}(N-k)]$$

for $k = 0, 1, \ldots, N-1$, with k and N-k taken modulo N.

C3. $\arg[\underset{\sim}{X}(k)]$ is periodic with period N; that is,

$$\arg[\underset{\sim}{X}(k)] = \arg[\underset{\sim}{X}(k+rN)] \qquad r = 0, \pm 1, \pm 2, \ldots,$$

where $k = 0, 1, \ldots, N-1$, and k and k+rN are taken modulo N.

To see what these conditions mean, let us consider the type of phase curves to be expected for finite-length sequences. Consider a sequence x whose values x(n) satisfy

$$x(n) = 0 \qquad \text{for } n < 0 \text{ and } n > M.$$

The corresponding z-transform is

$$X(z) = \sum_{n=0}^{M} x(n)\, z^{-n} = Az^{-m_o} \prod_{r=1}^{m_i} \left(1-a_r z^{-1}\right) \prod_{r=1}^{m_o} (1-b_r z), \tag{80}$$

where the a_r and b_r are all less than one in magnitude, and $M = m_i + m_o$. Equation 80 places in evidence the fact that in general there will be m_i zeros inside the unit circle and m_o zeros outside the unit circle. Since $\arg[X(e^{j\omega})]$ is the sum of the angles of each factor in Eq. 80, consideration of Figs. 8, 9, and 13 shows that $\arg[X(e^{j\omega})]$ has the properties

$$\arg[X(e^{j\omega})] = 0 \qquad \text{for } \omega = 0, 2\pi, 4\pi, \ldots \tag{81a}$$

$$\arg[X(e^{j\omega})] = -m_o\pi \qquad \text{for } \omega = \pi-\gamma, 3\pi-\gamma, \ldots, \tag{81b}$$

$$\arg[X(e^{j\omega})] = m_o\pi \qquad \text{for } \omega = \pi+\gamma, 3\pi+\gamma, \ldots, \tag{81c}$$

where γ is an arbitrarily small positive number. In obtaining these equations, we have assumed that A is positive.

Let us now consider the corresponding sampled Fourier transform and its sampled phase. We obtain $\underset{\sim}{X}(k)$ from (80) simply by replacing z by W^k, where $W = e^{j\frac{2\pi}{N}}$. Therefore we obtain

$$\underset{\sim}{X}(k) = AW^{-km_o} \prod_{r=1}^{m_i} \left(1-a_r W^{-k}\right) \prod_{r=1}^{m_o} \left(1-b_r W^{k}\right) \tag{82}$$

for $k = 0, 1, \ldots, N-1$. We can determine the sign of A by looking at

$$\underset{\sim}{X}(0) = A \prod_{r=1}^{m_i} (1-a_r) \prod_{r=1}^{m_o} (1-b_r),$$

since the sign of A is the same as the sign of $\underset{\sim}{X}(0)$. If A is negative, we normally change the sign of $\underset{\sim}{X}(k)$ before computing $\arg[\underset{\sim}{X}(k)]$, so that we remove any constant phase

component caused by A. Henceforth we shall assume that A is positive.

The properties corresponding to Eqs. 81 are easily seen to be

$$\arg\left[\underset{\sim}{X}(k)\right] = 0 \qquad \text{for } k = 0, \frac{N}{2}, N, \frac{3N}{2}, 2N, \ldots, \tag{83a}$$

$$\arg\left[\underset{\sim}{X}(k)\right] \approx -m_o\pi \qquad \text{for } k = \frac{N}{2} - 1, \frac{3N}{2} - 1, \ldots, \tag{83b}$$

$$\arg\left[\underset{\sim}{X}(k)\right] \approx +m_o\pi \qquad \text{for } k = \frac{N}{2} + 1, \frac{3N}{2} + 1, \ldots. \tag{83c}$$

We note that $\arg\left[\underset{\sim}{X}(k)\right]$ is given the value 0 at $k = N/2;\ 3N/2, \ldots$ so as to satisfy the requirement of odd symmetry. These properties are exhibited by Fig. 19a where we have shown only one half cycle of $\arg\left[\underset{\sim}{X}(k)\right]$.

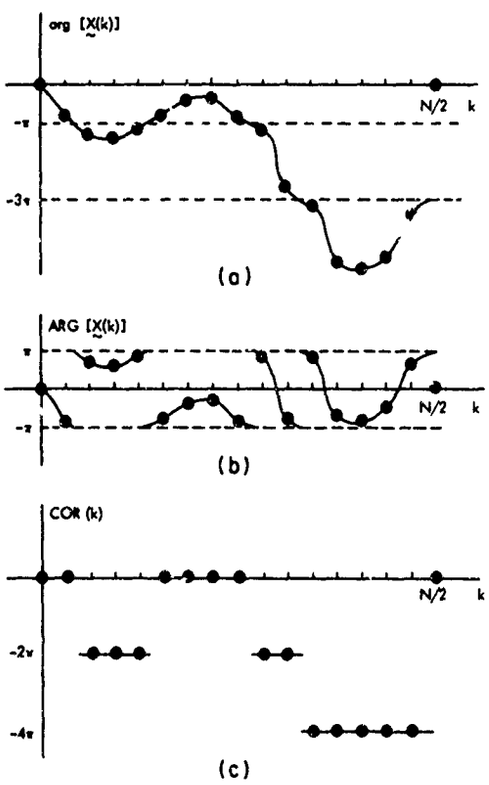


Fig. 19.

(a) Samples of $\arg\left[X(e^{j\omega})\right]$. (b) Principal value of $\arg\left[X(e^{j\omega})\right]$. (c) Correction sequence for obtaining arg from ARG.

These properties of the sampled phase curve will be used to discuss an algorithm for computing $\arg\left[\underset{\sim}{X}(k)\right]$ from $\underset{\sim}{X}(k)$ which in turn is obtained by using the FFT algorithm.

3.4 AN ALGORITHM FOR COMPUTING $\arg\left[\underset{\sim}{X}(k)\right]$

The properties of $\arg\left[\underset{\sim}{X}(k)\right]$ were given in section 3.3. These properties must be satisfied by any phase curve that we compute if we want to insure that

$$\hat{\underset{\sim}{x}}(n) = \hat{\underset{\sim}{x}}_1(n) + \hat{\underset{\sim}{x}}_2(n),$$

where x is a finite length sequence of the form

$$x = x_1 \otimes x_2.$$

A typical sampled phase curve is shown in Fig. 19a. In Fig. 19b, we show the principal value of the curve in Fig. 19a. Let us recall that we can obtain $\underset{\sim}{X}(k)$ from

$$\underset{\sim}{X}(k) = \sum_{n=0}^{N-1} x(n)\, W^{-kn} = \underset{\sim}{X}_r(k) + j\underset{\sim}{X}_i(k), \tag{84}$$

where N is a power of 2 that is greater than the length of the sequence x. Thus we are given $\underset{\sim}{X}_r(k)$ and $\underset{\sim}{X}_i(k)$, and we must compute arg $[\underset{\sim}{X}(k)]$ so as to satisfy the properties given in section 3.3. This is not as simple as it may appear at first glance. The principal value may be easily obtained by using standard routines based on a polynomial approximation. Let us assume that we have computed

$$-\pi < \mathrm{ARG}\,[\underset{\sim}{X}(k)] \leq \pi \tag{85}$$

for k = 0, 1, ..., N-1. Although ARG $[\underset{\sim}{X}(k)]$ does not satisfy our requirements, it can be used as a basis for computing the correct phase curve. To see this, consider Fig. 19b. We note that arg $[\underset{\sim}{X}(k)]$ can be expressed as the sum

$$\arg\,[\underset{\sim}{X}(k)] = \mathrm{ARG}\,[\underset{\sim}{X}(k)] + \mathrm{COR}\,(k), \tag{86}$$

where COR (k) is shown in Fig. 19c for that example. In general, it is clear that

$$\mathrm{COR}\,(k) = 2\pi q,$$

where q is a positive or negative integer that depends on k so that the properties of section 3.3 are satisfied for all k.

Our discussion suggests that we can compute arg $[\underset{\sim}{X}(k)]$ from ARG $[\underset{\sim}{X}(k)]$ by computing the correction sequence COR (k) and then adding it to ARG $[\underset{\sim}{X}(k)]$. This can be done if the sampling rate is high enough. Then ARG $[\underset{\sim}{X}(k)]$ contains all information required to compute COR (k). In order to see this, it is convenient to make several definitions. We say that ARG $[\underset{\sim}{X}(k)]$ has a positive jump of 2π (PJ of 2π) at k if

$$\mathrm{ARG}\,[\underset{\sim}{X}(k+1)] - \mathrm{ARG}\,[\underset{\sim}{X}(k)] > 2\pi - \epsilon_1, \tag{87a}$$

where ϵ_1 is a positive number whose value depends to some extent on the rate at which we sample the phase. Similarly, we say that ARG $[\underset{\sim}{X}(k)]$ has a negative jump of 2π at k (NJ of 2π) if

$$\text{ARG}\,[\underset{\sim}{X}(k+1)] - \text{ARG}\,[\underset{\sim}{X}(k)] < -(2\pi-\epsilon_1). \tag{87b}$$

By carefully considering the example of Fig. 19, we see that COR (k) can be computed by using the following algorithm:

$$\text{COR}\,(k+1) = \text{COR}\,(k) - 2\pi \qquad \text{for a PJ of } 2\pi \text{ at } k \tag{88a}$$

$$\text{COR}\,(k+1) = \text{COR}\,(k) + 2\pi \qquad \text{for a NJ of } 2\pi \text{ at } k \tag{88b}$$

$$\text{COR}\,(k+1) = \text{COR}\,(k) \qquad \text{otherwise,} \tag{88c}$$

where COR (0) = 0 and k = 0, 1, . . . , N-1. Therefore we see that we can compute arg $[\underset{\sim}{X}(k)]$ in the following steps: Compute ARG $[\underset{\sim}{X}(k)]$ for k = 0, 1, . . . , N-1, using a suitable routine based on an inverse tangent approximation; use Eqs. 87 and 88 to compute COR (k) for k = 0, 1, . . . , N-1; and add COR (k) to ARG $[\underset{\sim}{X}(k)]$ for k = 0, 1, . . . , N-1 to obtain arg $[\underset{\sim}{X}(k)]$. (We note that this step can be done, as COR (k) is computed so that extra storage for COR (k) is not required.)

When we are dealing with input sequences with many samples, we quite often find that there may be several hundred zeros of X(z) outside the unit circle. Since

$$\arg\left[\underset{\sim}{X}\left(\frac{N}{2}-1\right)\right] \approx -m_o\pi,$$

where m_o is the number of zeros outside the unit circle, we often find that the linear phase component is so large that it dominates the rest of the phase components. Let us call the sampled linear phase component $\underset{\sim}{\Theta}(k)$. Therefore we find that in order to have properties of section 3.4, we require

$$\underset{\sim}{\Theta}(k) = -\frac{2\pi}{N}\, m_o k \qquad 0 \leq k < \frac{N}{2} \tag{89a}$$

$$= 0 \qquad k = \frac{N}{2} \tag{89b}$$

$$= \frac{2\pi}{N}\, m_o(k-N) \qquad \frac{N}{2} < k < N. \tag{89c}$$

It can be shown that the contribution to $\hat{\underset{\sim}{x}}(n)$, which is due to $j\underset{\sim}{\Theta}(k)$, is

$$\underset{\sim}{\theta}(n) = \frac{\pi}{N} m_o \frac{\cos \pi n}{\tan \frac{\pi n}{N}} \qquad n = 1, 2, \ldots, N-1 \tag{90}$$

$$= 0 \qquad n = 0. \tag{90}$$

These sequences are shown in Fig. 20. Just as $\underset{\sim}{\Theta}(k)$ dominates the phase when m_o is large, we find that $\underset{\sim}{\theta}(n)$ dominates the complex cepstrum and obscures much of the interesting information in $\hat{\underset{\sim}{x}}$.

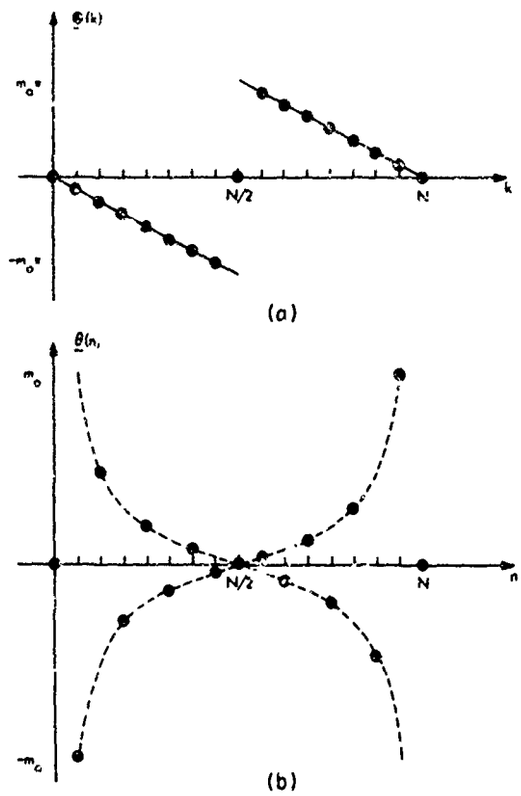


Fig. 20. (a) Linear phase component attributable to a factor W^{-km_o}.
(b) IDFT of (a).

It is possible to remove this component before computing the complex cepstrum. Clearly, we could simply subtract the sequence in Fig. 20a from arg $[\underset{\sim}{X}(k)]$ before computing the complex cepstrum. This can be done, since we note that

$$m_o \approx -\frac{\arg\left[\hat{\underset{\sim}{X}}\left(\frac{N}{2}-1\right)\right]}{\pi} \tag{91}$$

if we have sampled the phase curve at a high enough rate. Thus we can compute the right-hand side of (91) and round off the result to the nearest integer to obtain m_o.

Removing the linear phase component in arg $[\underset{\sim}{X}(k)$ is equivalent to removing the factor W^{-km_o} in (80) by multiplying by W^{km_o}. This in turn is equivalent to rotating (since all integers are taken modulo N) the input sequence $\underset{\sim}{x}(n)$ to the left by m_o positions. This fact may be used at the output of the inverse characteristic system to reposition the output sequence when this is necessary.

As an example, let us consider the phase curves of Fig. 19. We see that we can compute Figs. 19b and 19c by using the methods discussed previously. If we add these two, we obtain Fig. 19a which contains the undesirable linear component. On the other hand, if we first note that

$$m_o \approx -\frac{\mathrm{ARG}\left[\underset{\sim}{X}\left(\frac{N}{2}-1\right)\right] + \mathrm{COR}\left(\frac{N}{2}-1\right)}{\pi}, \tag{92}$$

then we could subtract $\underset{\sim}{\Theta}(k)$ from COR (k) before adding it to ARG $[\underset{\sim}{X}(k)]$. This is shown in Fig. 21 wherein we have repeated Fig. 19b as Fig. 21a. Figure 21a shows the resulting correction for this example, and Fig. 21c shows the sum of Fig. 21a and 21b. The result is the phase curve of the rotated sequence whose values are $\underset{\sim}{x}(n+m_o)$, where $n + m_o$ is taken modulo N.

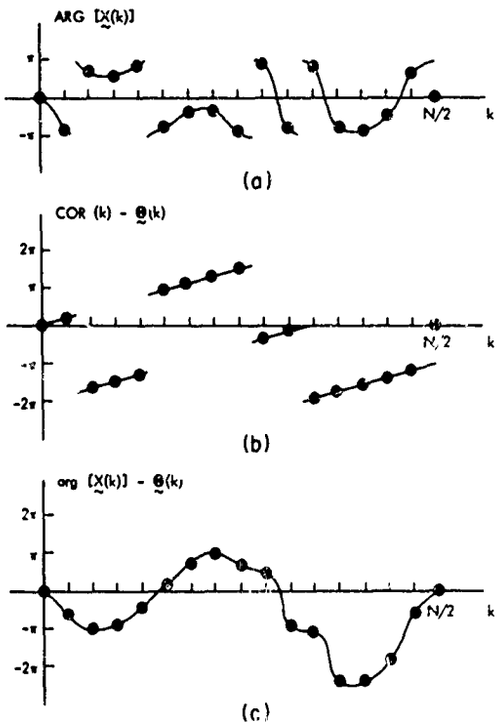


Fig. 21.

(a) Typical principal value curve. (b) Correction curve for obtaining arg $[\underset{\sim}{X}(k)]$ and at the same time removing the linear phase component. (c) Result of adding (a) to (b).

Let us summarize the results of the previous discussion. Our procedure in computing the phase is as follows.

(A1) Compute the principal value ARG $[\underset{\sim}{X}(k)]$ from the DFT $\underset{\sim}{X}(k)$ for k = 0, 1, . . . , N-1.

(A2) Compute the step correction function COR (k) for k = 0, 1, . . . , N-1 using Eqs. 88.

(A3) Determine how far to the left the sequence should be rotated by using Eq. 92.

(A4) Subtract $\underset{\sim}{\Theta}(k)$ as defined in Eqs. 89 from COR (k) for k = 0, 1, . . . , N-1.

(A5) Add the result of (A4) to ARG $[\underset{\sim}{X}(k)]$ for k = 0, 1, . . . , N-1 to obtain the phase for the rotated sequence.

If one has enough computer storage for an extra table of N values, we recommend that each step be done for all values of k before proceeding to the next step. This is because of the differences in the sizes of the values of the three sequences that we are adding or subtracting. The maximum value of ARG $[\underset{\sim}{X}(k)]$ is clearly π. The maximum magnitude of COR (k) and of $\underset{\sim}{\Theta}(k)$ is $m_o\pi$; however, the maximum magnitude of the difference of the two is generally much less than $m_o\pi$.

If storage is not available for an extra table, we can accomplish the same result by essentially doing (A2) twice. We can see from Fig. 19 that the value COR (N-1) is equal to

$$\mathrm{COR}\left(\frac{N}{2}-1\right) = 2\pi(\#NJ-\#PJ),$$

where #PJ is the number of positive jumps, and #NJ is the number of negative jumps in the interval $0 \leq k < N/2$. Thus we can compute m_o from Eq. 92 and form COR (k) - $\underset{\sim}{\Theta}(k)$ as it is added to ARG $[\underset{\sim}{X}(k)]$ and then store the result in the same location.

3.5 OTHER COMPUTATIONAL CONSIDERATIONS

We shall discuss some simple techniques that can be used to minimize the required computation time for realizing the characteristic system and its inverse.

Gentleman and Sande[15] summarize some of the useful properties of the DFT. Many authors refer to the symmetries inherent in the DFT relationships. For example, we find that if $\underset{\sim}{x}(n)$ is real, then the real and imaginary parts of the DFT

$$\underset{\sim}{X}(k) = \underset{\sim}{X}_r(k) + j\underset{\sim}{X}_i(k),$$

have the properties

$$\underset{\sim}{X}_r(k) = \underset{\sim}{X}_r(N-k) \tag{93a}$$

$$\underset{\sim}{X}_i(k) = -\underset{\sim}{X}_i(N-k) \tag{93b}$$

for k = 0, 1, . . . , N-1 and k and N-k taken modulo N. Thus we say that the real part of $\underset{\sim}{X}(k)$ is an even function of k, and imaginary part of $\underset{\sim}{X}(k)$ is an odd function of k.

Let us consider a sequence whose values are purely imaginary such as

$$\underset{\sim}{x}(n) = j\underset{\sim}{q}(n) \qquad \text{for } n = 0, 1, \ldots, N-1,$$

where the $\underset{\sim}{q}(n)$ are real numbers. The resulting DFT would be

$$\underset{\sim}{X}(k) = j\underset{\sim}{Q}(k) = -\underset{\sim}{Q}_i(k) + j\underset{\sim}{Q}_r(k),$$

where

$$\underset{\sim}{Q}(k) = \underset{\sim}{Q}_r(k) + j\underset{\sim}{Q}_i(k)$$

is the DFT of the real sequence $\underset{\sim}{q}$ whose values are $\underset{\sim}{q}(n)$. Therefore, if

$$\underset{\sim}{x}(n) + \underset{\sim}{p}(n) + j\underset{\sim}{q}(n),$$

where p and q are both finite length sequences, then we see that

$$\underset{\sim}{X}_r(k) = \underset{\sim}{P}_r(k) - \underset{\sim}{Q}_i(k)$$

$$\underset{\sim}{X}_i(k) = \underset{\sim}{P}_i(k) + \underset{\sim}{Q}_r(k).$$

Because of the symmetry properties, we see that the following relations are true.

$$\underset{\sim}{P}(k) - \underset{\sim}{P}_r(k) + j\underset{\sim}{P}_i(k) = \text{Ev}\left[\underset{\sim}{X}_r(k)\right] + j\ \text{Odd}\left[\underset{\sim}{X}_i(k)\right] \tag{94a}$$

$$\underset{\sim}{Q}(k) = \underset{\sim}{Q}_r(k) + j\underset{\sim}{Q}_i(k) = \text{Ev}\left[\underset{\sim}{X}_i(k)\right] - j\ \text{Odd}\left[\underset{\sim}{X}_r(k)\right], \tag{94b}$$

where, for example,

$$\text{Ev}\left[\underset{\sim}{X}_r(k)\right] = \frac{\underset{\sim}{X}_r(k) + \underset{\sim}{X}_r(N-k)}{2} \qquad k = 0, 1, \ldots, N-1 \tag{95a}$$

$$\text{Odd}\left[\underset{\sim}{X}_r(k)\right] = \frac{\underset{\sim}{X}_r(k) - \underset{\sim}{X}_r(N-k)}{2} \qquad k = 0, 1, \ldots, N-1. \tag{95b}$$

From these relationships, several comments are in order.

Comment 1: For real sequences, if we are given $\underset{\sim}{X}_r(k)$ for $k = 0, 1, \ldots, N/2$, and $\underset{\sim}{X}_i(k)$ for $k = 1, 2, \ldots, N/2-1$, then we can use Eqs. 93 to determine $\underset{\sim}{X}_r(k)$ and $\underset{\sim}{X}_i(k)$ for all k. This fact can be used to conserve memory when we must store several complex transforms for some reason.

Comment 2: We can obtain the DFT of two sequences p and q by using only one evaluation of the DFT relationships by evaluation of

$$\underset{\sim}{X}(k) = \sum_{n=0}^{N-1} (\underset{\sim}{p}(n)+j\underset{\sim}{q}(n))\, W^{-kn} = \underset{\sim}{P}(k) + j\underset{\sim}{Q}(k).$$

We have seen that Eqs. 94 and 95 can be used to recover $\underset{\sim}{P}(k)$ and $\underset{\sim}{Q}(k)$ from $\underset{\sim}{X}(k)$. Comment 1 can be applied here to allow us to store the transforms $\underset{\sim}{P}(k)$ and $\underset{\sim}{Q}(k)$ in 2N locations, rather than in the 4N locations required to store all values of these transforms.

Comment 3: In computing $|\underset{\sim}{X}(k)|^2$, $\frac{1}{2}\log|\underset{\sim}{X}(k)|^2$, ARG $[\underset{\sim}{X}(k)]$, COR (k), and arg $[\underset{\sim}{X}(k)]$, we recognize that each of these are even or odd so that we only need compute N/2 appropriate values of the sequences, and then we can find the rest of the values by symmetry. Thus we can save almost one half of the time in performing these operations on $\underset{\sim}{X}(k)$.

3.6 COMPUTATION TIME REQUIREMENTS

We have shown the numerical techniques that must be used to carry out the transformations D and D^{-1}. A major consideration is the amount of time required for these operations. We shall provide rough guidelines for estimating the computation time.

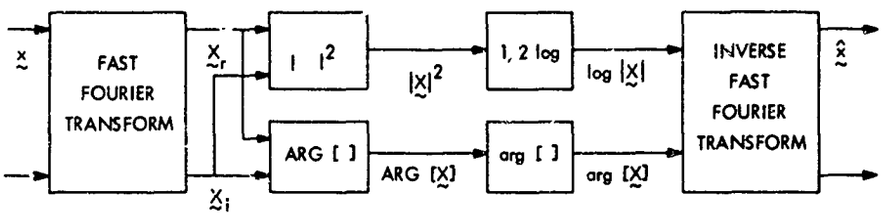


Fig. 22. Computational realization of the characteristic system.

The operations required for the transformations D and D^{-1} are summarized in Figs. 22 and 23. In these figures, all operations are performed on either one or two tables of real numbers. The length of these tables is N, which is a power of two. As we can see, these numbers are combined together in various ways to obtain the operations of FFT, log, ARG, etc. These numerical operations consist of additions, subtractions, multiplications, and divisions, coupled together by logical operations and indexing through the tables. Most of the total computation time is due to the arithmetic operations.

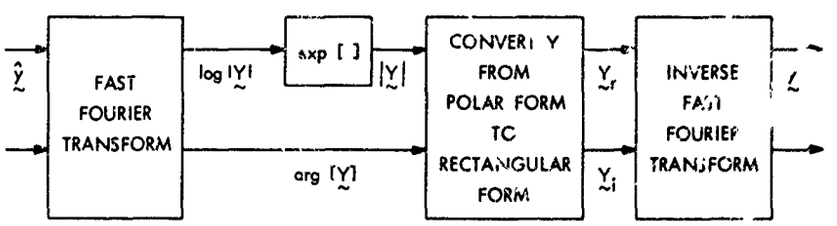


Fig. 23. Computational realization of the inverse characteristic system.

In realizing the transformations D or D^{-1}, we require 2 FFT's. We have stated that the number of complex additions and multiplications is equal to $N \log_2 N$. The remaining operations; log, ARG, arg, exp, etc. all require a number of additions and multiplications proportional to N. Thus it is reasonable to write

$$T_D = 2T_{FFT} N \log_2 N + T_{AD} N \tag{96}$$

and

$$T_{D^{-1}} = 2T_{FFT} N \log_2 N + T_{AD^{-1}} N, \tag{97}$$

where

T_D = time required to compute the complex cepstrum

$T_{D^{-1}}$ = the time required to compute the output

$T_{FFT} N \log_2 N$ = the time required to compute the FFT

$T_{AD} N$ = the time required to compute the log magnitude and the phase curve

$T_{AD^{-1}} N$ = the time required to exponentiate the transform of $\hat{y}$.

For one realization on the TX-2 computer, we obtained the values

T_{FFT} = 60 μsec

T_{AD} = 9 msec

$T_{AD^{-1}}$ = 6 msec.

Thus, for example, for N = 4096, we obtain

T_D = 40 sec

$T_{D^{-1}}$ = 30 sec

We should point out that Eqs. 96 and 97 would be divided by 2 if we exploit all of the symmetry properties.

3.7 MINIMUM-PHASE COMPUTATIONS

We have discussed certain simplifications that occur when the input sequence is minimum-phase. As we have seen, for minimum-phase finite-length sequences, all zeros of X(z) are inside the unit circle, and we have as many poles at zero as we have nonzero zeros. We have shown that, in this case, a recursive formula can be derived for $\hat{x}(n)$ and we can also compute $\hat{x}(n)$ from $\log |X(e^{j\omega})|$ alone without computing the phase. We shall now consider the actual computational realization of this second method and compare it with the recursive method.

We have seen that the samples of $\log |X(e^{j\omega})|$ may be computed by first calculating

$$\underset{\sim}{X}(k) = \underset{\sim}{X}_r(k) + j\underset{\sim}{X}_i(k) = \sum_{n=0}^{N-1} x(n)\, W^{-kn}$$

for k = 0, 1, . . . , N-1. (We use the FFT algorithm to evaluate these equations.) We may use a polynomial approximation to compute

$$X_r(k) = \log |\underset{\sim}{X}(k)| = \frac{1}{2} \log \left[\underset{\sim}{X}_r^2(k) + \underset{\sim}{X}_i^2(k) \right]$$

for k = 0, 1, . . ., N-1, as we normally do even if the phase is to be used. Because of

the periodicity of $\underset{\sim}{x}$ and $\underset{\sim}{\hat{x}}$, the even part of $\underset{\sim}{\hat{x}}$ is

$$\mathrm{Ev}\,[\underset{\sim}{\hat{x}}(n)] = \frac{\underset{\sim}{\hat{x}}(n) + \underset{\sim}{\hat{x}}(N-n)}{2}, \tag{98}$$

where n and N-n are taken modulo N. Since $\underset{\sim}{\hat{X}}_r(k)$ is an even function, its IDFT is $\mathrm{Ev}\,[\underset{\sim}{\hat{x}}(n)]$ so that

$$\mathrm{Ev}\,[\underset{\sim}{\hat{x}}(n)] = \frac{1}{N}\sum_{k=0}^{N-1} \frac{1}{2}\log\left[\underset{\sim}{X}_r^2(k) + \underset{\sim}{X}_i^2(k)\right] W^{kn} \tag{99}$$

for n = 0, 1, . . . , N-1. We note that $\hat{x}(n) = 0$ for $n < 0$ for minimum-phase sequences. We have also seen that if $\hat{x}(n) = 0$ for $n \geq N$, then no aliasing will occur in $\underset{\sim}{\hat{x}}(n)$, that is,

$$\underset{\sim}{\hat{x}}(n) = \hat{x}(n) \qquad \text{for } n = 0, 1, \ldots, N-1.$$

We note that if $\hat{x}(n) = 0$ for $n \geq N/2$ and $n < 0$, it is clearly true that

$$\underset{\sim}{\hat{x}}(n) = \hat{x}(n) \qquad \text{for } n = 0, 1, \ldots, N-1.$$

Furthermore, we see from (98) that in this case we can obtain $\underset{\sim}{\hat{x}}(n)$ from $\mathrm{Ev}\,[\underset{\sim}{\hat{x}}(n)]$ by using the relation

$$\underset{\sim}{\hat{x}}(n) = \mathrm{Ev}\,[\underset{\sim}{\hat{x}}(n)]\,\underset{\sim}{u}(n), \tag{100}$$

where

$$\begin{aligned} \underset{\sim}{u}(n) &= 1 \quad n = 0 \\ &= 2 \quad 0 < n < \frac{N}{2} \\ &= 0 \quad \frac{N}{2} \leq N < N. \end{aligned} \tag{101}$$

If $\hat{x}(n) \neq 0$ for $n \geq N/2$, as is usually the case, we can still use Eqs. 99 and 100 to compute an aliased approximation to $\hat{x}(n)$. Clearly, the more rapidly $\hat{x}(n)$ approaches zero, for large positive and negative values of n, the better will be the approximation for a given value of N. Alternatively, the larger we make N, the better will be the approximation to $\hat{x}(n)$. In the case of minimum-phase sequences, if we exponentially weight the input so that

$$w(n) = a^n x(n),$$

then the complex cepstrum is also exponentially weighted by the same exponential. That is, if $|a| < 1$,

$$\hat{w}(n) = a^n \hat{x}(n).$$

Therefore exponential weighting can be used to help reduce the aliasing of the complex cepstrum.

In summary, we can compute an aliased complex cepstrum when x is minimum-phase by using the following equations which are the counterparts of Eqs. 60-62.

$$\underset{\sim}{X}(k) = \sum_{n=0}^{N-1} \underset{\sim}{x}(n)\, W^{-kn} \qquad k = 0, 1, \ldots, N-1 \tag{102a}$$

$$\mathrm{Ev}\,[\hat{\underset{\sim}{x}}(n)] = \frac{1}{N} \sum_{k=0}^{N-1} \log |\underset{\sim}{X}(k)|\, W^{kn} \qquad n = 0, 1, \ldots, N-1, \tag{102b}$$

$$\hat{\underset{\sim}{x}}(n) = \mathrm{Ev}\,[\hat{\underset{\sim}{x}}(n)]\, \underset{\sim}{u}(n). \tag{102c}$$

We have defined $\underset{\sim}{u}(n)$ in (101) and we recall that $W = e^{j\frac{2\pi}{N}}$.

As an alternative, of course, we have Eqs. 44. These equations are repeated below for the case $x(n) = 0$ for $n < 0$ and $n > M$.

$$\hat{x}(n) = \log x(0) \qquad n = 0$$

$$= \frac{x(n)}{x(0)} - \sum_{k=0}^{n-1} \frac{k}{n}\, \hat{x}(k)\, \frac{x(n-k)}{x(0)} \qquad 0 \leq n \leq M$$

$$= \frac{x(n)}{x(0)} - \sum_{k=n-M}^{n-1} \frac{k}{n}\, \hat{x}(k)\, \frac{x(n-k)}{x(0)} \qquad M < n.$$

It is clear that the recursive algorithm has the advantage over Eqs. 102, in that no aliasing is introduced. Furthermore, to compute M samples of $\hat{x}(n)$ we require only 2M data storage locations. The recursive algorithm, however, suffers greatly when compared with Eqs. 102 on the basis of computation time. If we exploit all the symmetry properties in carrying out Eqs. 102 as discussed in section 3.6, we find that for values of M greater than 64, the evaluation of Eqs. 102 for N = 2M is faster than evaluating M points by using the recursive algorithm. This is an estimate based on a third-order approximation to the logarithm and the assumption that a multiplication takes twice as long as addition. In cases in which a higher order approximation to the logarithm is required, or a multiplication takes longer than twice the time for an addition, this crossover point will be higher, but not significantly higher. Therefore for minimum-phase sequences we find that even though we have obtained a recursive algorithm in which aliasing is not a problem it is usually preferable to use Eqs. 102 with N such that aliasing is not significant. The recursive algorithm is still useful conceptually, and it

may find application when only a small number of values of $\hat{x}(n)$ is required or when aliasing cannot be tolerated.

3.8 SAMPLING OF CONTINUOUS-TIME SIGNALS

It can be shown that in the case of continuous-time signals whose Fourier transforms are bandlimited, the samples of a continuous time convolution are equal to the discrete convolution of the samples of the individual signals. That is, if time and frequency are normalized so that the Nyquist frequency is 1 Hz, then

$$x(n) = \int_{-\infty}^{\infty} x_1(\tau)\, x_2(n-\tau)\, d\tau = \sum_{k=-\infty}^{\infty} x_1(k)\, x_2(n-k). \tag{103}$$

In our calculations, we have assumed that all sequences are of finite length. This assumption would imply that the corresponding continuous time function is time-limited. It is well known that a time-limited signal cannot have a frequency-limited Fourier transform. In practice, however, the Fourier transform approaches zero quite rapidly in most cases, even when the signal is time-limited. Therefore we shall use the approximation that both the continuous time signal and its transform are zero outside of some finite interval.

Let us consider a continuous-time function x(t) whose Fourier transform $X_c(\omega)$ is such that

$$X_c(\omega) = 0 \qquad |\omega| > \pi.$$

The Fourier transform of the sequence of samples of x(t) is

$$X(e^{j\omega}) = X_c(\omega) \qquad \text{for } -\pi < \omega < \pi$$

and $X(e^{j\omega})$ is periodic with period equal to 2π.

Such an example is shown in Fig. 24. There we see that if we sample exactly at the Nyquist rate, we encounter no aliasing of $X(e^{j\omega})$, and $|X(e^{j\omega})|$ is nonzero for all ω. This means that $\log |X(e^{j\omega})|$ will be finite for all ω. This is shown in Fig. 25a. On the other hand, if we sample at a rate higher than the Nyquist rate as in Fig. 24c, we see that $X(e^{j\omega})$ will be zero over a finite interval, so that $\log |X(e^{j\omega})|$ will be undefined in that interval. This is shown in Fig. 25b.

The previous example is idealized; however, the essential points remain true in practice. If we are dealing with a finite-length sequence that is the result of sampling some continuous-time signal, we normally would try to sample at a rate that is as high as possible so as to avoid aliasing in the Fourier transform. Since the high-frequency content of many signals (speech or seismic signals) is quite low, we shall normally find that if the sampling rate is high, there is a significant interval of frequency over which

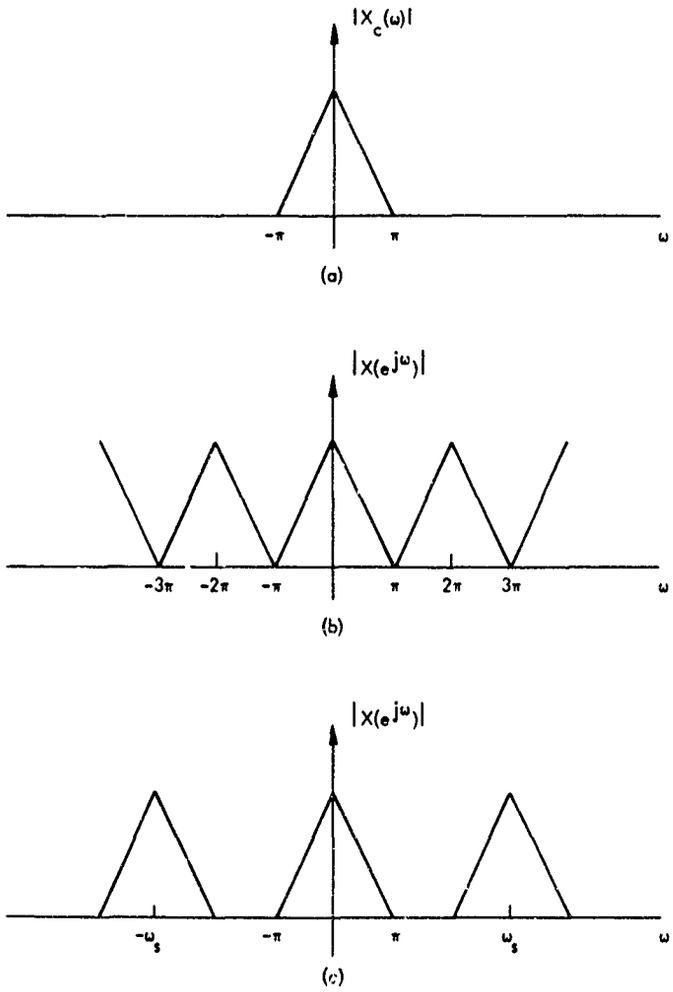


Fig. 24. (a) Magnitude of the Fourier transform of a bandlimited continuous-time function x(t). (b) Transform of the samples of x(t) for sampling just at the Nyquist rate. (c) Transform of the samples of x(t) for sampling at a rate higher than the Nyquist rate.

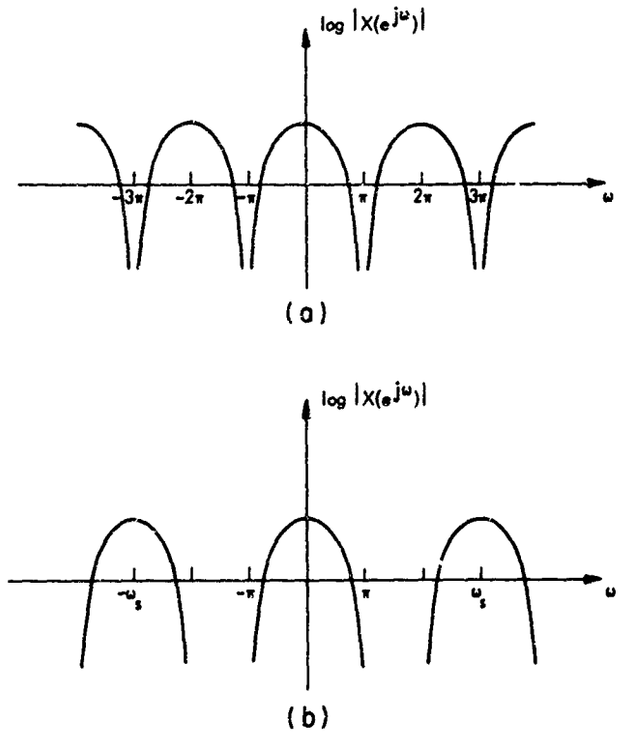


Fig. 25. (a) Log magnitude for Fig. 24b.
(b) Log magnitude for Fig. 24c.

$X(e^{j\omega})$ is very small. This means that if we compute the sampled transform

$$\underset{\sim}{X}(k) = \underset{\sim}{X}_r(k) + j\underset{\sim}{X}_i(k) = \sum_{n=0}^{N-1} \underset{\sim}{x}(n)\, W^{-kn} \qquad k = 0, 1, \ldots, N-1,$$

then both $\underset{\sim}{X}_r(k)$ and $\underset{\sim}{X}_i(k)$ will normally be quite small for values of k around N/2. Since we must compute both $\log |\underset{\sim}{X}(k)|$ and ARG $[\underset{\sim}{X}(k)]$, using polynomial approximations involving divisions, we shall normally encounter severe accuracy problems when both $\underset{\sim}{X}_r(k)$ and $\underset{\sim}{X}_i(k)$ are very small. This is especially true in using fixed-point machines in which we must effectively keep the scale factor of $\underset{\sim}{X}_r(k)$ and $\underset{\sim}{X}_i(k)$ the same for all values of k. Similarly, the scale factor must be the same for $\log |\underset{\sim}{X}(k)|$ and $\arg [\underset{\sim}{X}(k)]$ for all values of k. Therefore it is clear that in sampling a continuous-time signal so as to apply discrete homomorphic deconvolution, we must effectively sample as in Fig. 24b, rather than as in Fig. 24c. In most cases we can usually lowpass filter the signal before sampling. Therefore if we precede the sampler with a very sharp cutoff filter, and then sample at a little more than twice the nominal cutoff frequency of the filter, we shall obtain a sequence in which aliasing is not excessive in the Fourier transform, while at the same time we shall make it possible to accurately compute $\log |\underset{\sim}{X}(k)|$ and $\arg [\underset{\sim}{X}(k)]$ for all values of k.

IV. ANALYSIS OF SPEECH WAVEFORMS

We have shown that it is possible to transform a convolution of two or more signals into a sum of related signals. We have seen that in some cases these signals are separated in "time" so that a frequency-invariant linear system can be used to advantage in separating the signals from one another.

We shall now discuss a simple model for the production of speech waveforms. This model results in a representation of certain segments of speech waveforms as a convolution of an aperiodic pulse with a quasi-periodic impulse train. Thus speech waveforms are examples of the class of signals to which this technique of deconvolution is particularly applicable.

Our purpose in discussing speech is twofold. First, we shall see that speech waveforms serve as very interesting examples of the techniques presented in this report. Second, Section V will focus on the problem of echo removal and detection. As a particular example of the application of the results of Section V, we shall discuss the removal and detection of echoes in speech signals. Thus we shall also obtain an understanding of the characteristics of the complex cepstrum of speech which will be applied in another context in Section V.

The application of homomorphic deconvolution to speech analysis by Oppenheim[19,20] has paralleled our application to echo removal. This section will attempt to give a brief introduction to this field of application. The reader who is interested in this is directed to Oppenheim and Schafer[19] and Oppenheim[20] for more detail. We shall first discuss a model for the speech waveform and then consider the complex cepstrum of speech. We shall give some examples illustrating the applicability of our techniques to the recovery of the separate components of the speech waveform.

4.1 SPEECH PRODUCTION AND THE SPEECH WAVEFORM

The speech signal is an acoustic disturbance that is generated by air escaping from the lungs of the speaker. The mouth and throat form an acoustic resonator called the vocal tract which is excited by the air that is supplied by the lungs. We may think of the lungs as a source of a steady flow of air which is converted into a varying flow either by the vocal cords or by constrictions of the vocal tract. In the first case, the vocal cords may convert the steady flow of air into a series of quasi-periodic pulses by rapidly opening and closing the passage to the lungs. Sounds generated in this way are called voiced speech sounds. Examples are the vowel sounds. The other principal class of speech sounds is termed unvoiced and these are generated by creating constrictions in the vocal tract which cause turbulence to occur at these points. Many of the consonants are generated by using this type of "noise" signal as excitation for the vocal tract.

In Fig. 26a, we see a schematic representation of voiced speech production.[18] The lungs supply a steady flow of air that is modulated by the vocal cords to give the excitation function e(t) as shown in Fig. 26b. The vocal tract is modeled by a cascade of linear

resonators whose combined impulse response v(t) is typically like Fig. 26c. The resonant frequencies of the resonators are called the formant frequencies. The speech

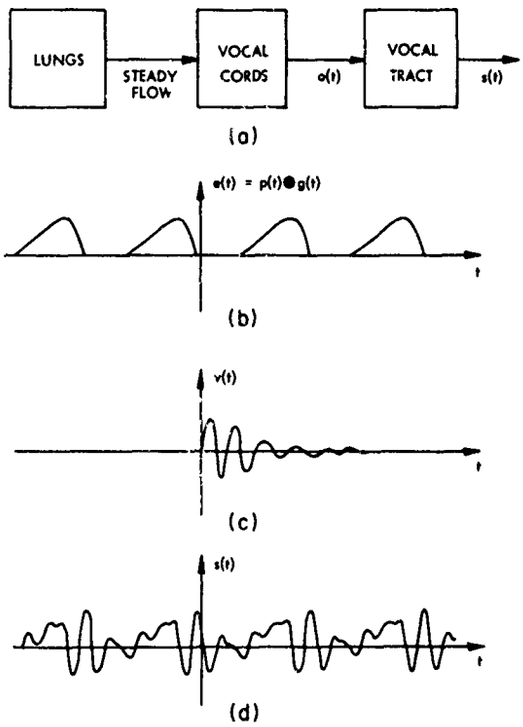


Fig. 26. (a) Model of voiced-speech production. (b) Excitation function generated by vocal cords. (c) Impulse response of vocal tract. (d) Resulting speech waveform.

signal is the response of the vocal tract to the quasi-periodic excitation e(t), and therefore is given by

$$s(t) = \int v(\tau)\, e(t-\tau)\, d\tau.$$

Clearly, the resulting speech signal is also quasi-periodic. Since e(t) is quasi-periodic, we could further represent e(t) as the convolution of a basic pulse shape g(t) (called the glottal pulse) with an impulse train p(t) such that

$$e(t) = g(t) \otimes p(t).$$

Therefore we obtain

$$s(t) = v(t) \otimes g(t) \otimes p(t).$$

In the case of unvoiced sounds, the vocal tract is effectively excited by noise which is not pulselike or quasi-periodic. A similar model can be used, however, with the excitation being a noise source.

We have seen that speech can be thought of as a convolution of two or more continuous-time signals. In general, the Fourier transforms of the separate components

of speech are bandlimited (with the exception of the impulse train). Thus it is reasonable to assume that the samples of the speech waveform can be obtained from discrete convolutions of the samples of the individual components. Thus, a reasonable model for the samples of a segment of voiced speech is

$$s(n) = v(n) \otimes g(n) \otimes p(n),$$

where p(n) is a sequence that is approximately of the form

$$p(n) = \sum_r \delta(n-rn_p),$$

where n_p corresponds to the "pitch" of the sound.

4.2 SHORT-TIME TRANSFORM

Speech production can be modeled as we have just discussed, but implicit in our discussion was the fact that the character of the speech waveform changes as time proceeds. That is, human speech is a string of sounds that are continuously changing. Thus the model that we have discussed must be time-variant in the sense that both the excitation function and the vocal-tract impulse response change as a function of time.

The fact that the character of the speech waveform changes with time, together with the fact that for reasonably good quality speech we need at least 10,000 samples per second to represent the waveform, requires that we adjust our notions about Fourier analysis. We find that it is neither possible or desirable to speak of the Fourier transform of even an entire sentence, let alone the transform of longer segments. Rather, the notion of a short-time Fourier transform is more appropriate.

Suppose we are interested in a segment of the speech waveform in the vicinity of $n = \xi$. (The time origin is arbitrary.) We define the short-time z-transform as

$$S(z, \xi) = \sum_{n=0}^{L-1} s(n+\xi)\, w(n)\, z^{-n}. \tag{104}$$

In Eq. 104, w(n) is a "window" containing L nonzero samples. We "view" the speech signal through this window by changing the parameter ξ, in order to move the segment of interest under the window. The short-time Fourier transform of the sampled speech is obtained by letting $z = e^{j\omega}$, that is,

$$S(e^{j\omega}, \xi) = \sum_{n=0}^{L-1} s(n+\xi)\, w(n)\, e^{-j\omega n}. \tag{105}$$

Since $S(e^{j\omega}, \xi)$ is really just the transform of a sequence of finite length, all of the properties of Fourier transforms of such sequences apply to (105). We note that the

sequence $s(n+\xi)\ w(n)$ can be recovered by using the inverse Fourier transform relation

$$s(n+\xi)\ w(n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} S(e^{j\omega}, \xi)\ e^{j\omega n}\ d\omega. \tag{106}$$

In speech analysis, the length of the window is usually comparable to the duration of the shortest speech sounds. We note that $S(e^{j\omega}, \xi)$ is also equal to the convolution of the transform of $s(n+\xi)$ with the transform of $w(n)$. Therefore if one requires good frequency resolution, longer windows may be required. This is the case in the echo removal and detection applications for which we shall see that the window must be quite long compared with the temporal detail of the signal. When we are interested in speech analysis, however, we find that we must make the window only a few pitch periods long so that the character of the speech signal remains essentially the same within the window. Typically, the duration of the window in this case is equal to a number of samples equivalent to 20-40 msec of speech. At a 10-kHz sampling rate this is from 200 to 400 samples.

For speech analysis, we have found the so-called Hamming window to be quite useful, because of its good frequency resolution properties. The discrete Hamming window is

$$\begin{aligned} w(n) &= \frac{1}{2}\left(1 - \cos\frac{2\pi}{L} n\right) && 0 \leq n < L \\ &= 0 && \text{elsewhere} \end{aligned} \tag{107}$$

For echo removal and detection, we shall see in Section V that a truncated exponential window of the form

$$\begin{aligned} w(n) &= a^n && 0 \leq n < L \\ &= 0 && \text{elsewhere} \end{aligned}$$

has properties that are very much suited to that application.

4.3 SHORT-TIME COMPLEX CEPSTRUM OF SPEECH

We have seen that speech may be modeled as a convolution if we consider short segments of the waveform. We have also introduced the notion of a short-time transform with an appropriate window. We shall now discuss how short-time transforms may be used to obtain a short-time realization of homomorphic deconvolution.

Let us consider a segment of voiced speech that is multiplied by a window; that is,

$$s(n+\xi)\ w(n) = [vg(n) \otimes p(n+\xi)]\ w(n), \tag{108}$$

where $vg(n) = v(n) \otimes g(n)$. (We could, of course, assume without loss of generality that the segment of interest occurs at $\xi = 0$.) In practice, the component $vg(n)$ is of finite duration, and we shall assume that it remains the same over the entire window. We observe that the sequence whose values are $s(n+\xi)\ w(n)$ for $0 \leq n < L$, however, is not strictly a

convolution. It should be clear that this is so because the "tails" of previous periods overlap into the window and the "tails" of periods within the window are truncated by the window. Nevertheless, under appropriate conditions this sequence will be approximately a convolution so that the short-time transform for such a segment will be approximately the product of the transform of vg(n) and the transform of a segment of p(n). This is best illustrated as follows.

Let us assume that vg has a duration of not more than two pitch periods. We shall assume that the window varies slowly so that $w(n) \approx w(n+2n_p)$, where n_p is the pitch period. Then we may write Eq. 108 as

$$s(n+\xi)\ w(n) = vg(n) \otimes p_w(n, \xi) + e(n, \xi),$$

where $e(n, \xi)$ accounts for the overlaps at the beginning and the end of the window, and

$$p_w(n, \xi) = p(n+\xi)\ w(n).$$

Thus we have assumed that the window is such that vg remains essentially the same across the whole window, with the pitch pulses being the only part that is weighted by the window. If we evaluate the short-time transform we obtain

$$S(e^{j\omega}, \xi) = VG(e^{j\omega})\ P_w(e^{j\omega}, \xi) + E(e^{j\omega}, \xi).$$

Therefore we see that only if $E(e^{j\omega}, \xi)$ is negligible, is $S(e^{j\omega}, \xi)$ simply a product. If it is true that $E(e^{j\omega}, \xi)$ is negligible, however, we see that

$$\hat{S}(e^{j\omega}, \xi) = \log\left[S(e^{j\omega}, \xi)\right] \approx \log\left[P_w(e^{j\omega}, \xi)\right] + \log\left[VG(e^{j\omega})\right].$$

Windows that taper to zero at the ends are quite effective in minimizing the end effects.

We define the short-time complex cepstrum to be

$$\hat{s}(n, \xi) = \frac{1}{2\pi}\int_{\pi}^{\pi} \log\left[S(e^{j\omega}, \xi)\right] e^{j\omega n}\, d\omega \approx \hat{p}_w(n, \xi) + \widehat{vg}(n).$$

Thus the short-time complex cepstrum is the sum of $\hat{p}_w(n, \xi)$, which contains the pitch information in the interval spanned by the window, and the complex cepstrum of vg. The latter can actually be thought of as the sum of two components, one that is due to the vocal tract and the other to the glottal pulse. The impulse response of the vocal tract is minimum phase; however, the glottal pulse is not. Therefore $\widehat{vg}(n)$ will be nonzero for both positive and negative values of n. Nevertheless, it is clear from the aperiodic nature of vg, that the component $\widehat{vg}(n)$ will tend to be concentrated around n = 0 and will be bounded by

$$\left|\widehat{vg}(n)\right| \leq A\left|\frac{1}{n}\right| \qquad \text{for all } n.$$

On the other hand, $p_w(n, \xi)$ is an impulse train which in general will give an impulse train in the complex cepstrum, with the impulses occurring at multiples of the pitch period. Thus we have a situation in which the complex cepstrum can be divided into two regions, each corresponding to a different component of the speech signal. To recover the component $p_w(n, \xi)$, we zero those samples of the complex cepstrum for $|n| < n_p$, and then employ the inverse of the characteristic system to obtain $p_w(n, \xi)$. To recover vg, we replace with zero those samples for $|n| \geq n_p$, and transform the result with the inverse system. Examples of this are given in section 4.4.

If the speech segment under consideration is unvoiced, the short-time complex cepstrum does not have the impulse train component as in the voiced speech. Thus, the complex cepstrum can be used as a pitch detector and voiced/unvoiced detector for vocoder and speech analysis applications. Noll[11, 12] has shown that the cepstrum is very successful in this application. (The cepstrum as defined by others is essentially the even part of the complex cepstrum.) It appears that the complex cepstrum may offer advantages over the cepstrum, since it is possible to actually extract the component $p_w(n, \xi)$ so that we obtain information about variations in pitch across the window, rather than an average pitch period as is obtained in the cepstrum. The reason for this is that if the pitch is not constant, the impulses in the cepstrum and complex cepstrum either become smeared out around some average pitch period, or impulses appear at longer times than the fundamental period. The following examples illustrate this point.

4.4 EXAMPLES

We shall first consider two examples. We shall consider the recovery of pitch, and then the recovery of the component vg, which contains the impulse response of the vocal tract and the glottal pulse.

As an example of the recovery of pitch, let us consider the segment of the vowel "ah" as in father, shown in Fig. 27a. This waveform was sampled at a 10-kHz rate, and in Fig. 27a the samples have been connected by straight lines to form a smooth curve. (This is the case in all of the curves that we show in this section.) Note that the pitch is quite constant in the interval shown.

In Fig. 27c is shown the complex cepstrum for the segment of Fig. 27a when weighted by the Hamming window of Eq. 107, where L = 256 (25.6 msec). Since the Hamming window is very small in value at its ends, it tends to minimize the error caused by overlap. We note significant peaks in the complex cepstrum at approximately ±5 msec (50 samples). This is the period of the waveform in Fig. 27a. We also note that the complex cepstrum is relatively small for values beyond 50 samples. Figure 27b shows the output of the inverse characteristic system after having replaced the samples for $|n| \leq 40$ with zero. We note that the impulses are spaced at the pitch period and are weighted by the shape of the Hamming window.

As an example of what happens when the pitch varies across the window, we

replaced the samples of Fig. 27c with zero for $|n| \geq 40$. Thus the pitch was removed, and the resulting output of the inverse characteristic system was a pulse containing the glottal pulse and vocal tract information. This pulse is shown in Fig. 29c. This pulse was used in the computer to synthesize a new waveform in which the spacing of impulses in p(n) was alternating between 35 samples and 40 samples. This waveform is shown in Fig. 28a. (Such pitch fluctuations have been reported by Noll.[12])

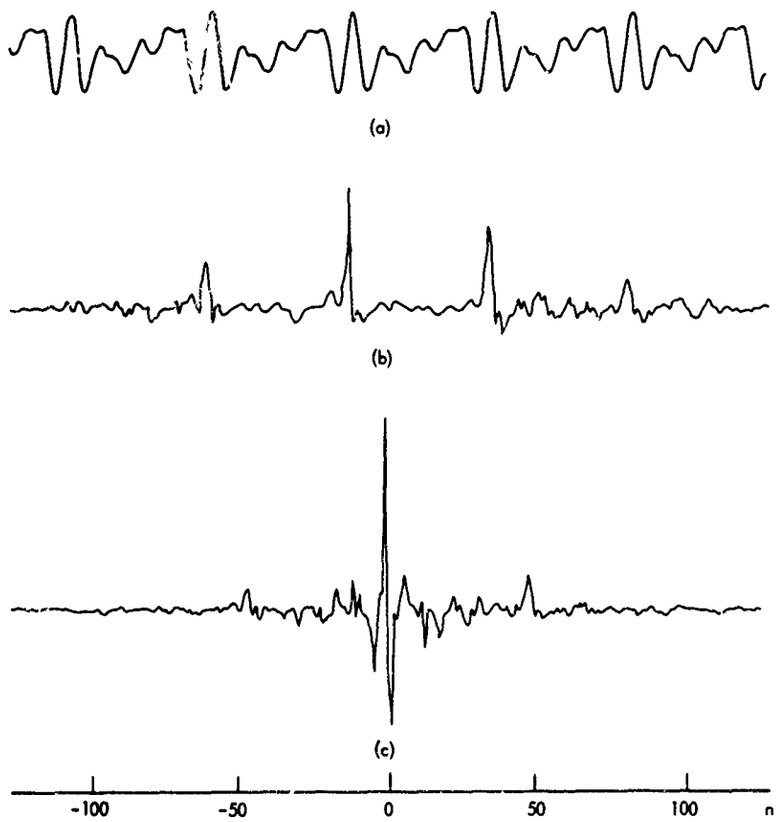


Fig. 27. Pitch-extraction example. (a) Segment of the vowel "ah." (b) The output for a long-pass system. (c) Complex cepstrum for Hamming window.

The complex cepstrum for Fig. 28a weighted by a Hamming window with L = 256 samples is shown in Fig. 28c. This time, we note that there are significant peaks at $n = -75$, -40, $+35$, and $+75$. The values of the complex cepstrum were replaced by zero for $|n| < 32$ and the resulting output of the inverse characteristic system is shown in Fig. 28b. It is clear that in this case, the output of the overall system shows very clearly how the pitch changes in this time interval. On the other hand, things are not so clear in the complex cepstrum. In the even part of

the complex cepstrum, that is, the cepstrum, we would find peaks at n = ±35, ±40, and ±74. Thus we would probably have to be content with an average constant pitch over this interval, since the cepstrum does not tell in what order the long and short periods occur.

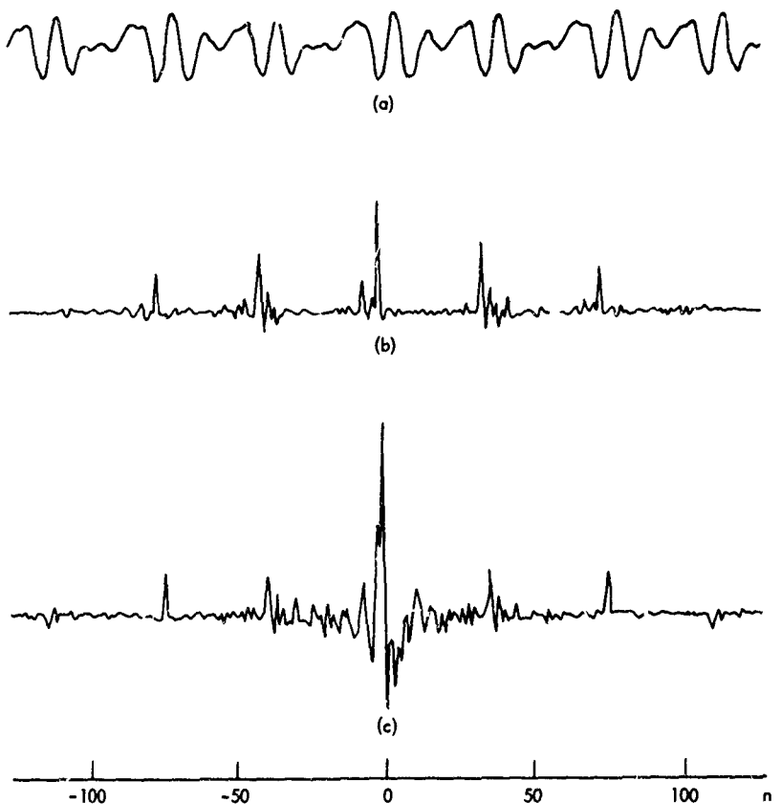


Fig. 28. Pitch extraction example. (a) Synthetic speech with fluctuating pitch. (b) The output for a long-pass system. (c) Complex cepstrum for a Hamming window.

The first two examples indicate that the use of the complex cepstrum may be desirable in situations in which we are interested in very accurate and synchronous pitch extraction. This remark is based, however, on only limited experimentation and can only be verified through more extensive effort in this direction.

As a third example, we note that in obtaining the waveform of Fig. 28a we recovered a pulse from the complex cepstrum of Fig. 27c. This pulse is shown in Fig. 29c. The original waveform is repeated in Fig. 29a with expanded time and amplitude scales. To show that this pulse, together with pitch information for the original waveform, is sufficient to recover the waveform, we synthesized the waveform of Fig. 29b, using pitch information obtained from Fig. 27b. We note that the

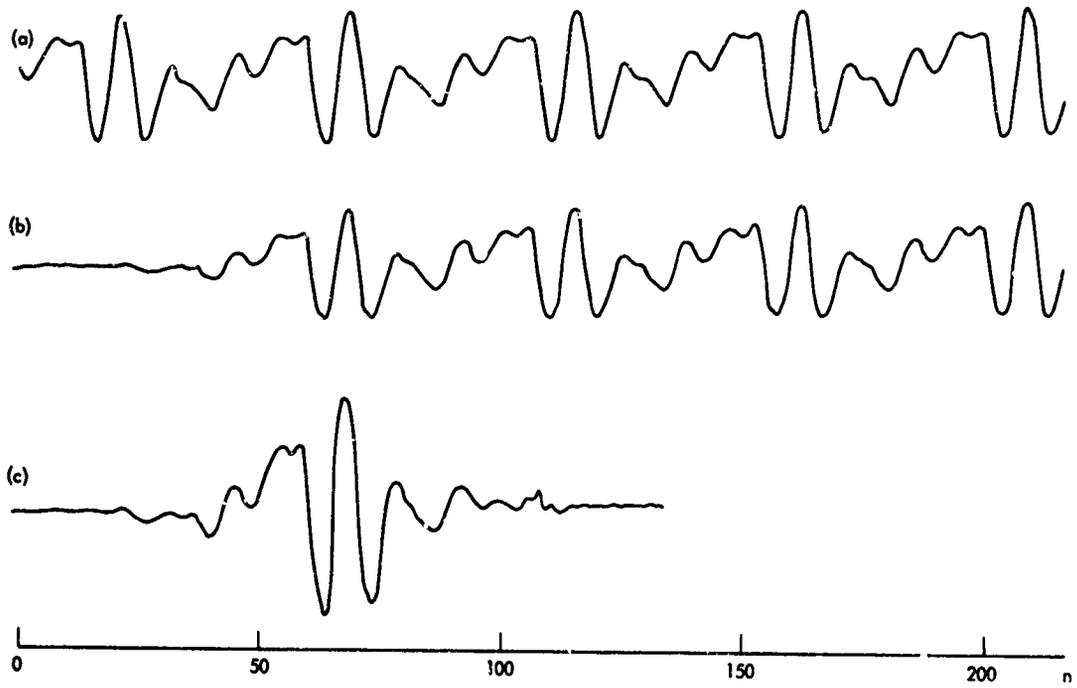


Fig. 29. Example of speech deconvolution. (a) Original speech (same as Fig. 27a). (b) Synthesized speech using the pulse in (c) and pitch information in Fig. 27b. (c) Pulse obtained from the cepstrum of Fig. 27c, using the values for $|n| \leq 40$. (Note: sampling rate is 10 kHz.)

waveforms (a) and (b) are not exactly the same, but they are quite similar in most details. From this example and Fig. 27b, it seems clear that the assumptions employed in section 4.3 are justified in this case. That is, the component vg is not affected to a great extent by the weighting, while the pitch impulses retain the shape of the window.

V. APPLICATIONS TO ECHO REMOVAL AND DETECTION

5.1 A SIMPLE EXAMPLE

We have presented the theoretical and computational details of homomorphic deconvolution. Now we shall be concerned with the application of these techniques to the processing of signals containing echoes. To illustrate our approach, let us consider a simple example. We can easily obtain analytical results for this example, and we shall also present computational results for comparison.

Suppose we have a sequence x whose values are

$$x(n) = s(n) + \alpha s(n-n_1) = s(n) \otimes p(n), \tag{108}$$

where the sequence s has values

$$s(n) = na^n \qquad 0 \leq n < M$$
$$= 0 \qquad \text{elsewhere.}$$

Furthermore, $0 < a < 1$, and

$$p(n) = \delta(n) + \alpha\delta(n-n_1).$$

The z-transform of Eq. 108 is

$$X(z) = S(z)\left(1+\alpha z^{-n_1}\right), \tag{109}$$

where S(z) can be shown to be

$$S(z) = \frac{az^{-1}(1-a^M z^{-M})}{(1-az^{-1})^2} - \frac{Ma^M z^{-M}}{1-az^{-1}}.$$

If M is large, a^M approaches zero, so that for large M

$$S(z) \approx \frac{az^{-1}}{(1-az^{-1})^2}.$$

The logarithm of X(z) is

$$\hat{X}(z) = \log\left[S(z)\right] + \log\left[1+\alpha z^{-n_1}\right],$$

which can be written

$$\hat{X}(z) = \log a + \log z^{-1} - 2\log\left[1-az^{-1}\right] + \log\left[1+\alpha z^{-n_1}\right]. \tag{110}$$

Using the Laurent series expansion of the logarithm, we can write Eq. 110

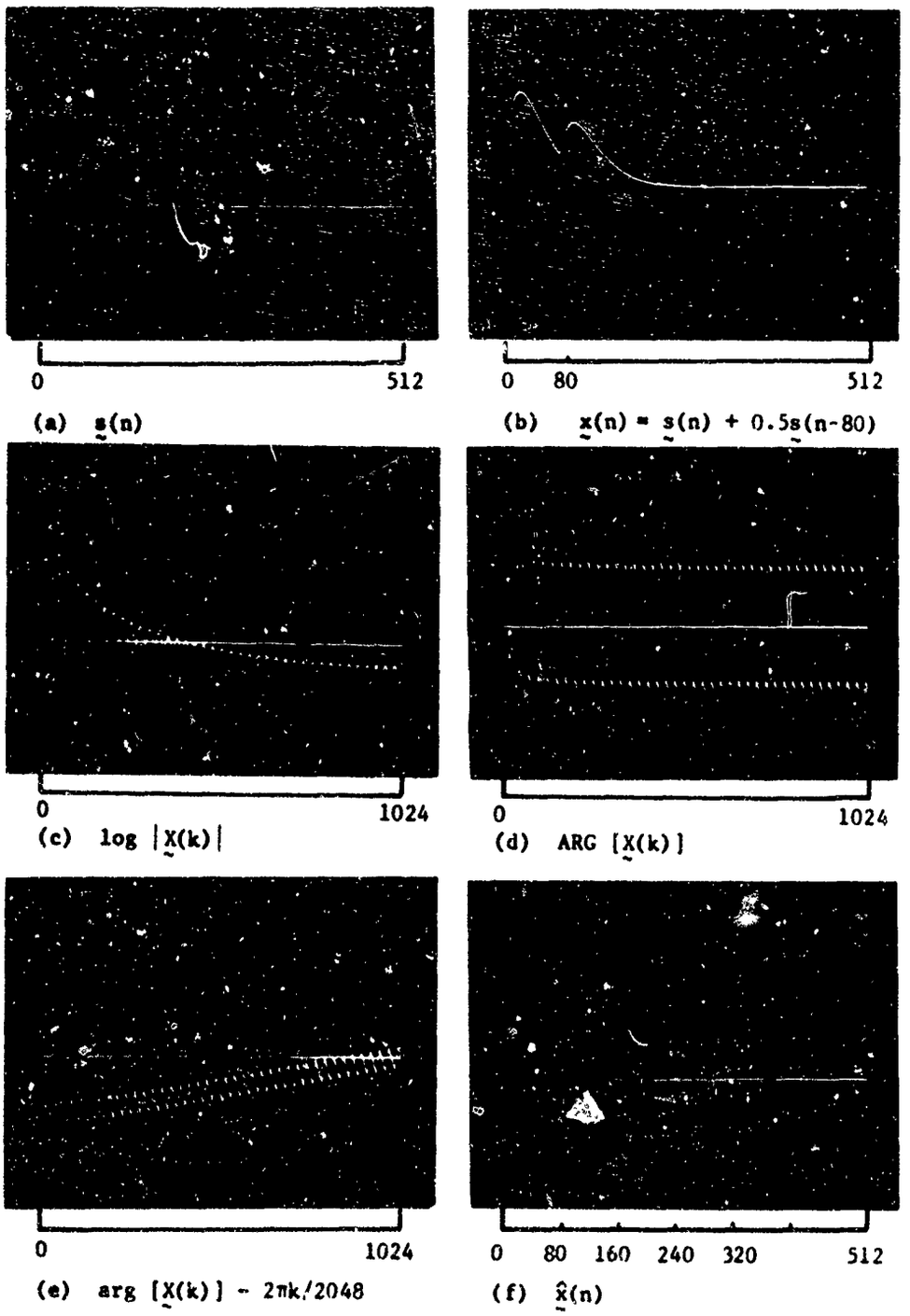


Fig. 30. Waveform and transform for a simple example.

$$\hat{X}(z) = \log a + \log z^{-1} + \sum_{m=1}^{\infty} \frac{2a^m}{m} z^{-m} + \sum_{m=1}^{\infty} (-1)^{m+1} \frac{\alpha^m}{m} z^{-mn_1}. \tag{111}$$

If we choose the contour of integration to be the unit circle, and remove the term $\log z^{-1}$ by shifting the sequence one place to the left, we obtain for the complex cepstrum

$$\hat{x}(n) = (\log a)\, \delta(n) + \sum_{k=1}^{\infty} \frac{2a^k}{k} \delta(n-k) + \sum_{k=1}^{\infty} (-1)^{k+1} \frac{\alpha^k}{k} \delta(n-kn_1). \tag{112}$$

From Eq. 112 we see that the contribution due to p has samples spaced at intervals of n_1 while the samples due the sequence s have unit spacing. Clearly both approach zero faster than $\frac{1}{n}$, but since the samples due to p have greater spacing, the part due to s will occupy primarily the "short-time" region while the part due to the echoes will be in the "long time" region.

This example was actually computed as discussed in Section III for $a = .96$, $\alpha = .5$, $M = 800$, and $n_1 = 80$. The value of N for the DFT was 2048. The sequences s and x are shown in Fig. 30a and 30b, respectively. In Fig. 30c, we show the samples of the real part of Eq. 110 for $z = e^{j\omega}$. That is, Fig. 30c is

$$\log |\underset{\sim}{X}(k)| = \log a - 2 \log \left| 1 - a e^{-j \frac{2\pi}{N} k} \right| + \log \left| 1 + \alpha e^{-j \frac{2\pi}{N} n_1 k} \right|.$$

In Fig. 30d we show the principal value of the phase of $\underset{\sim}{X}(k)$. This includes the linear phase component, as we can see, since

$$\mathrm{ARG}\left[\underset{\sim}{X}\left(\frac{N}{2} - 1\right)\right] \approx -\pi,$$

and the net number of positive and negative jumps between $k = 0$ and $k = \frac{N}{2} - 1$ is zero. Figure 30e shows the phase curve after adding the corrections involved in removing the discontinuities and rotating the sequence. Figure 30f shows the complex cepstrum for this example. We note that the part attributable to s(n) is primarily concentrated between $n = 0$ and $n = 80$, while the impulses attributable to the echo occur at $n = 80$, 160, etc. We should point out that after rotating to the left one sample, the input sequence is minimum phase, so that $\hat{x}(n) = 0 \quad n < 0$.

Let us now consider how we might choose a frequency-invariant linear system to recover each of the components s and p from the complex cepstrum. From Fig. 30b, we see that to recover the sequence s, we must remove the impulses caused by p. One way of doing this is to form

$$\hat{y}(n) = \ell(n)\, \hat{x}(n),$$

where $\ell(n)$ is as in Fig. 31a, and $n_c < 80$. This type of linear system is appropriately

termed a short-pass system. With this system, we would remove all of the contribution caused by p, and part of that caused by s. The analogy to lowpass filtering should

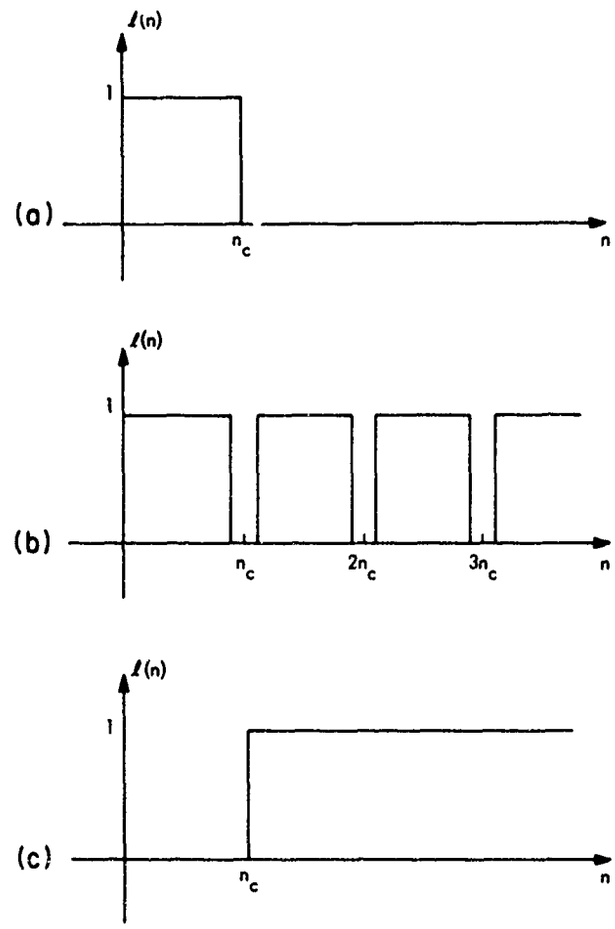


Fig. 31. Frequency-invariant linear systems for (a), (b) echo removal, and (c) detection.

be clear. If we consider Fig. 30c and 30e, we see that both the log magnitude and the phase are made up of the sum of a slowly varying component caused by s, and a more rapidly varying component caused by p. In attempting to recover s, we must remove the rapidly varying components while leaving the slowly varying component relatively unaltered. Alternatively, if we have accurate knowledge of the echo time (80 samples in this case), a comb system such as that shown in Fig. 31b would remove the contribution from p and would retain almost all of the part from s. In either case, operating on $\hat{y}(n)$ with the inverse system should produce an approximation to s(n). The computational example was carried out for these choices and the results are shown in Fig. 32. In Fig. 32c we show the output for the comb system with $n_c = 80$. Clearly, it is not possible to distinguish Fig. 32c from Fig. 30a. Similarly, we chose $n_c = 79$ in the short-pass system and obtained the output shown in Fig. 32d. Again, Fig. 32d is not

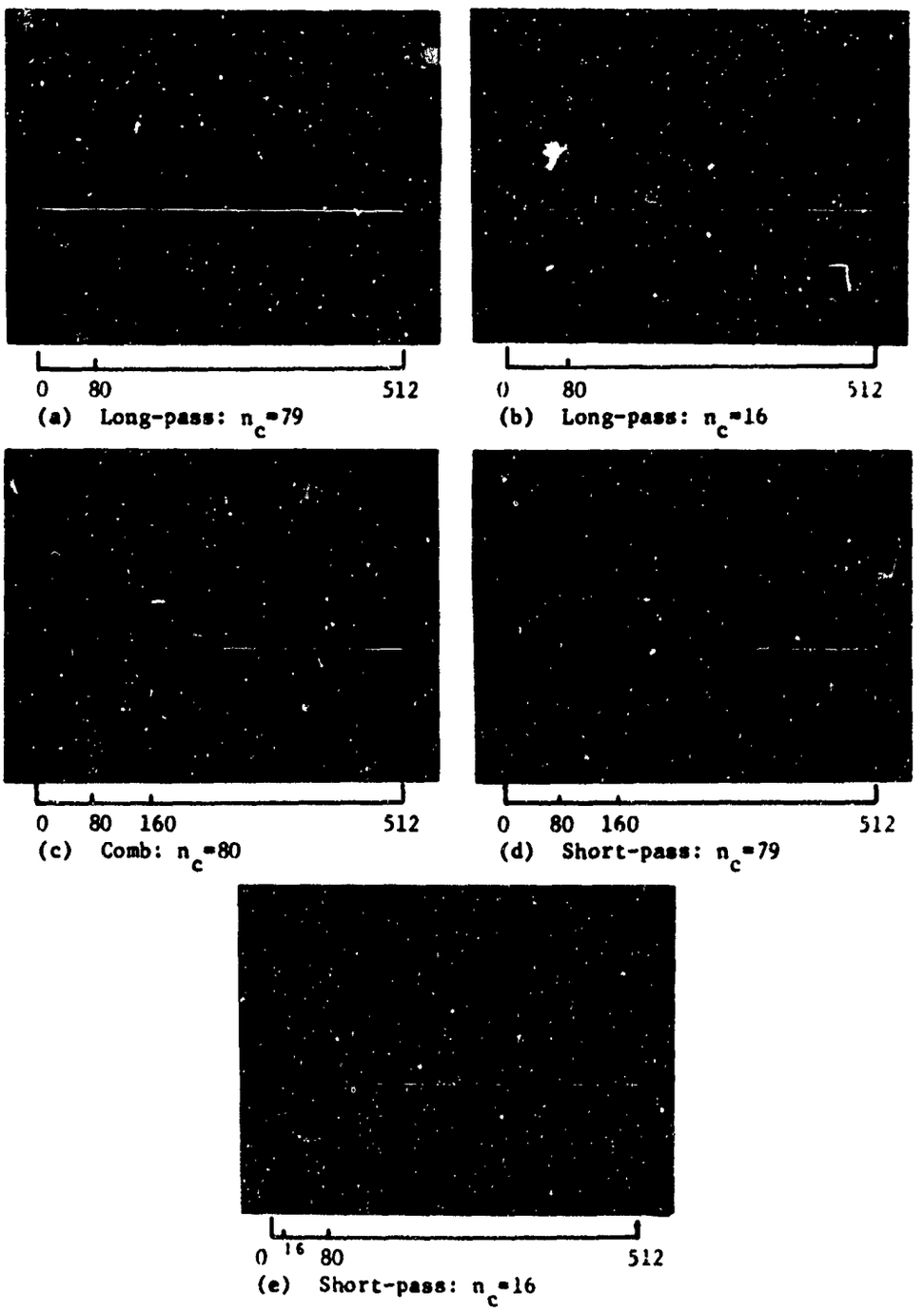


Fig. 32. Output waveforms for the example of Fig. 30.

distinguishable from Fig. 30a; however, in Fig. 32c we have shown the output when the short-pass system was used with $n_c = 16$. A look at Fig. 30f shows that much of $\hat{s}(n)$ has been deleted by the linear system, and therefore we should not expect the output to look exactly like s(n). The fact that the shape is the same is the result of the simplicity of the distribution of zeros of S(z), and the fact that s is minimum-phase.

We shall say more about the implications of minimum phase in this example, but first let us consider the problem of recovering the sequence p. From the complex cepstrum in Fig. 30f it is clear that if we choose the linear system of Fig. 31c, with $n_c < 80$, most of $\hat{s}(n)$ will be deleted, and we shall retain the part attributable to p. Such a system could be appropriately called a long-pass system. In Fig. 32a we show the output of the inverse system for a long-pass linear system with $n_c = 79$. We note that very clearly we have recovered a sequence whose values are very close to p(n). In Fig. 32b we show the output for $n_c = 16$. In this case, we have retained a significant part of $\hat{s}(n)$, and thus the impulses of p are convolved with the filtered s(n). Since most of the significant part of $\hat{s}(n)$ was removed, however, the part of the output from s(n) is very small.

We have previously noted that the input sequence was minimum-phase. Thus the recursive algorithm will give the same result as that obtained by using the integral expressions on the unit circle. The recursion formula for the inverse system is

$$y(n) = e^{\hat{y}(0)} \qquad n = 0$$

$$= y(0)\,\hat{y}(n) + \sum_{k=0}^{n-1} \frac{k}{n}\,\hat{y}(k)\,y(n-k) \qquad n > 0,$$

where $\hat{y}(n) = \ell(n)\,\hat{x}(n)$.

The recursive expression helps us to understand the appearances of each of the outputs in Fig. 32. For example, in both Fig. 32a and 32b, $\hat{y}(0) = 0$. Since

$$y(0) = e^{\hat{y}(0)},$$

we see that since y is minimum phase and $\hat{y}(0) = 0$, then $y(0) = 1$. Similarly, we see that if $\hat{y}(n) = 0$ for $0 \leq n < n_c$, then $y(n) = 0$ for $0 < n < n_c$. This is shown by both Fig. 32a, where $n_c = 79$, and 32b, where $n_c = 16$. We note that this explains why the contribution attributable to s(n) in Fig. 32b begins 16 samples from each impulse.

In Fig. 32c, 32d, and 32e, we recall that $\hat{y}(n) = \hat{x}(n)$ for $0 \leq n < n_c$. Thus from the recursion formula, we see that $y(n) = x(n)$ for $0 \leq n < n_c$ (within computational accuracy). Since $\hat{x}(n) = \hat{s}(n)$ $0 < n \leq n_c$ in each case, we see that $y(n) = s(n)$ for $0 \leq n < n_c$. Because $\hat{s}(n)$ is very small for $n \geq 80$, we see that $y(n) \approx s(n)$ for $n \geq n_c$ in Fig. 32c, and that y(n) should be a better approximation in Fig. 32d, since the comb system retains most of $\hat{s}(n)$. In Fig. 32e, we note that $y(n) = s(n)$ $0 \leq n < 16$, but for $n \geq 16$, the output decays much too fast. This can clearly be accounted for by the recursion formula.

This example has illustrated many of the important concepts in the use of

homomorphic deconvolution in processing signals containing echoes. We have seen that it is possible to remove the echo from the sequence x by using a short-pass or comb linear system. Furthermore, we have seen that it is possible to recover the impulse train by using a long-pass system. We note that this impulse train can be used in echo detection, and we shall see that it has advantages over using either the cepstrum or complex cepstrum for this purpose. It should be clear that the characteristics of the complex cepstrum of the impulse train are of primary importance in choosing the linear system. Thus, we shall next study this question in detail.

5.2 COMPLEX CEPSTRUM OF AN IMPULSE TRAIN

An impulse train is defined as a sequence in which the nonzero samples are spaced at intervals that are greater than one. An example of such a sequence is the sequence whose values are

$$p(n) = \delta(n) + \sum_{k=1}^{M-1} a_k \delta(n-n_k). \tag{113}$$

Sequences of this kind can be used to represent signals that contain echoes or reverberations. For example, the sequence whose values are

$$x(n) = s(n) + \sum_{k=1}^{M-1} a_k s(n-n_k),$$

can be represented as the convolution

$$x = s \otimes p,$$

where the values of p are given by (113), and s is a sequence whose values are spaced at unit intervals. The complex cepstrum of x has values

$$\hat{x}(n) = \hat{s}(n) + \hat{p}(n).$$

We have seen that even if the sequence is of finite length, the complex cepstrum $\hat{s}$ is in general of infinite extent. Most of the energy in $\hat{s}$ is concentrated, however, in an interval of the same order of magnitude as the duration of s. We saw in the example that the impulse train

$$p(n) = \delta(n) + a\delta(n-n_1)$$

has a complex cepstrum

$$\hat{p}(n) = \sum_{k=1}^{\infty} (-1)^{k+1} \frac{a^k}{k} \delta(n-kn_1).$$

This component may be clearly seen in Fig. 30d. Thus, in this simple case, we see that the complex cepstrum is also an impulse train with samples spaced at intervals of n_1. We have also seen that the two sequences $\hat{s}$ and $\hat{p}$ may occupy essentially different intervals so that it is possible to recover either s or p by using very simple frequency-invariant systems. Since we must have some knowledge of the complex cepstrum of the impulse train in order to choose the linear system, we shall consider in detail the properties of the complex cepstrum of an impulse train.

Let us consider an impulse train such as Eq. 113. The corresponding z-transform is

$$P(z) = 1 + \sum_{k=1}^{M-1} a_k z^{-n_k}. \tag{114}$$

In general, the analytical determination of the complex cepstrum corresponding to (114) is quite difficult because it is quite hard to determine the zeros of the right-hand side of (114). In the special case wherein the impulses are equally spaced, however, it is possible to give a general result. Consider an impulse train

$$p(n) = \sum_{k=0}^{M-1} a_k \delta(n-kn_o).$$

Let us define a sequence q having values

$$q(n) = a_n \qquad 0 \leq n \leq M-1$$

$$= 0 \qquad \text{elsewhere.}$$

Thus we see that

$$p(n) = q\left(\frac{n}{n_o}\right) \qquad n = 0, n_o, \ldots, (M-1)n_o$$

$$= 0 \qquad \text{elsewhere.}$$

The z-transform of p is

$$P(z) = \sum_{k=0}^{M-1} a_k z^{-kn_o} = \sum_{k=0}^{M-1} a_k \left(z^{n_o}\right)^{-k} = Q\left(z^{n_o}\right).$$

Thus it is clear that p will be given by

$$\hat{p}(n) = \hat{q}\left(\frac{n}{n_o}\right) \qquad n = 0, \pm n_o, \pm 2n_o, \ldots$$

$$= 0 \qquad \text{elsewhere.}$$

That is, for impulse trains with equal spacing, the complex cepstrum is also an impulse train with the same spacing. We note that even though p is a finite-length impulse train, $\hat{p}$ is in general of infinite duration.

If the spacing of the samples is not uniform, it is not possible to use the argument above. Since we are not able to give a general result, we shall illustrate the range of possibilities with several examples.

Example 1

Let $n_k = kn_o$ and $a_k = a^k$ in Eq. 114, so that P(z) can be written

$$P(z) = \sum_{k=0}^{M-1} a^k z^{-kn_o}. \tag{115}$$

By a simple manipulation of (115), we can obtain the more convenient form

$$P(z) = \frac{\left(1-a^M z^{-Mn_o}\right)}{\left(1-az^{-n_o}\right)}. \tag{116}$$

The logarithm of P(z) is

$$\hat{P}(z) = \log\left(1-a^M z^{-Mn_o}\right) - \log\left(1-az^{-n_o}\right),$$

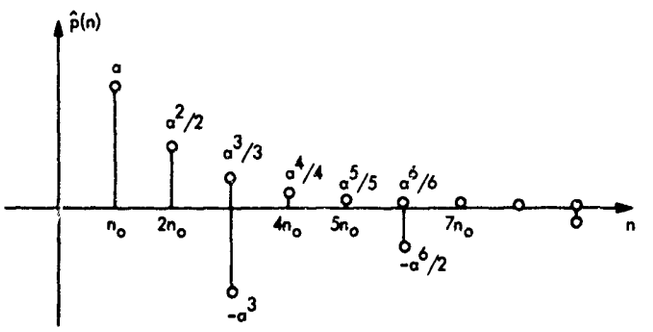
and if $|a| < 1$, we may write

$$\hat{P}(z) = \sum_{k=1}^{\infty} \frac{a^k}{k} z^{-kn_o} - \sum_{k=1}^{\infty} \frac{a^{Mk}}{k} z^{-kMn_o}.$$

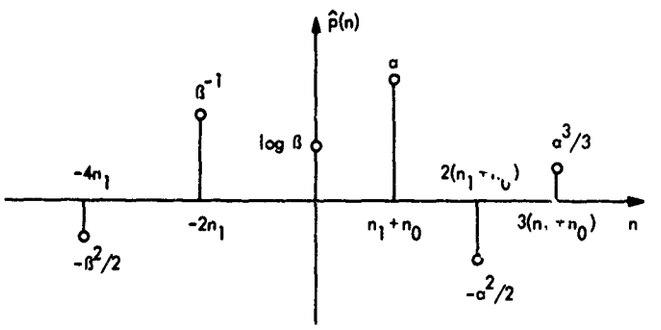
(Note that such an expansion implies that the region of convergence includes the unit circle.) Therefore, the complex cepstrum is

$$\hat{p}(n) = \sum_{k=1}^{\infty} \frac{a^k}{k} \delta(n-kn_o) - \sum_{k=1}^{\infty} \frac{a^{kM}}{k} \delta(n-kMn_o). \tag{117}$$

The two sequences making up $\hat{p}$ are shown in Fig. 33 for M = 3. Thus, we see that if the echoes are exponentially decreasing in amplitude and also equally spaced, the complex cepstrum is an impulse train with the same spacing. We also note that p is minimum-phase for $|a| < 1$, and therefore $\hat{p}(n) = 0$ for $n < 0$.



(a)



(b)

Fig. 33. (a) Complex cepstrum for Eq. 115.
(b) Complex cepstrum for Eq. 118.

Example 2

Let us suppose that

$$P(z) = 1 + az^{-n_1-n_o} + \beta z^{-2n_1} + a\beta z^{-3n_1-n_o}, \tag{118}$$

where $|\beta| > 1$ and $|a| < 1$. Equation 118 may be written

$$P(z) = \beta z^{-2n_1}\left(1+az^{-n_1-n_o}\right)\left(1+\beta^{-1}z^{2n_1}\right),$$

and the logarithm of P(z) is therefore

$$\hat{P}(z) = \log \beta + \log z^{-2n_1} + \log\left(1+az^{-n_1-n_o}\right) + \log\left(1+\beta^{-1}z^{2n_1}\right). \tag{119}$$

If we assume that the region of convergence contains the unit circle, we may expand (119) as

$$\hat{P}(z) = \log \beta + \log z^{-2n_1} + \sum_{k=1}^{\infty} (-1)^{k+1} \frac{\alpha^k}{k} z^{-(n_1+n_o)k} + \sum_{k=1}^{\infty} (-1)^{k+1} \frac{\beta^{-k}}{k} z^{2n_1 k}.$$

If we also assume that the term, $\log z^{-2n_1}$ is removed in computing the phase curve, we see that $\hat{p}(n)$ can be written

$$\hat{p}(n) = (\log \beta)\, \delta(n) + \sum_{k=1}^{\infty} (-1)^{k+1} \frac{\alpha^k}{k} \delta(n-kn_1-kn_o) + \sum_{k=1}^{\infty} (-1)^{k+1} \frac{\beta^{-k}}{k} \delta(n+k2n_1).$$

The sequence is depicted in Fig. 33b.

We note that in this case the sequence p is nonminimum-phase after a shift left of $2n_1$ samples, and therefore $\hat{p}(n) \neq 0$ for $n < 0$. We note, however, that $\hat{p}$ is again an impulse train with spacing related to the spacing of the sequence p. Clearly, it would be difficult to detect all echoes in this case by using only the impulses in the complex cepstrum. If, however, we use a long-pass system that replaces the samples in the interval $-2n_1 < n < n_1 + n_o$ with zero, and then transform the result with the inverse system, we shall obtain an output placing in evidence all of the samples of the sequence p.

Example 3

Let P(z) be

$$P(z) = 1 = \alpha_1 z^{-n_1} + \alpha_2 z^{-n_2}, \tag{120}$$

where $n_2 > n_1$. The logarithm of P(z) is

$$\hat{P}(z) = \log \left(1+\alpha_1 z^{-n_1}+\alpha_2 z^{-n_2}\right).$$

If we assume that the region of convergence of $\hat{P}(z)$ contains the unit circle and that

$$\max_{-\pi<\omega<\pi} \left| \alpha_1 e^{-j\omega n_1} + \alpha_2 e^{-j\omega n_2} \right| < 1,$$

then we may write

$$\hat{P}(z) = \sum_{k=1}^{\infty} (-1)^{k+1} \frac{\left(\alpha_1 z^{-n_1} + \alpha_2 z^{-n_2}\right)^k}{k}. \tag{121}$$

(We shall say more about this restriction on P(z) after this example.) The binomial term may be expanded as

$$\left(a_1 z^{-n_1} + a_2 z^{-n_2}\right)^k = \sum_{r=0}^{k} \binom{k}{r} a_1^{k-r} a_2^{r} z^{-kn_1+rn_1-rn_2},$$

where the quantity $\binom{k}{r}$ is the binomial coefficient

$$\binom{k}{r} = \frac{k!}{(k-r)!\,r!}.$$

Thus we can write (121)

$$\hat{P}(z) = \sum_{k=1}^{\infty} (-1)^{k+1} \frac{1}{k} \sum_{r=0}^{k} \binom{k}{r} a_1^{k-r} a_2^{r} z^{-(kn_1-rn_2+rn_1)},$$

and the complex cepstrum is therefore

$$\hat{p}(n) = \sum_{k=1}^{\infty} (-1)^{k+1} \frac{1}{k} \sum_{r=0}^{k} \binom{k}{r} a_1^{k-r} a_2^{r} \delta(n-kn_1+rn_1-rn_2).$$

This expression places in evidence several important facts. First, we note that $\hat{p}(n) = 0 \quad n < 0$. This is a result of our assumption regarding a_1 and a_2. Second, we can see that the impulses occur at

$$n = (k-r)n_1 + rn_2 \qquad k = 1, 2, \ldots \quad \text{and} \quad r = 0, 1, \ldots, k.$$

The values of $\hat{p}(n)$ are given in Table 1 for $1 \leq k \leq 6$.

The most striking thing about Examples 2 and 3 is how much more complicated the complex cepstrum becomes when the impulse train is nonminimum-phase, or when the impulses are not equally spaced.

In general, we must consider impulse trains of the form of Eq. 113, and clearly if the impulse train is both not equally spaced and nonminimum-phase, the complex cepstrum will have impulses located throughout the range $-\infty < n < \infty$. It is to our advantage if the impulses in the complex cepstrum occur only in the "long-time" region. Since the spacing of the impulses is not under our control, we can only look at the possibility of making the impulse train minimum-phase. With respect to this question we can make some definite statements. We recall that exponential weighting can be used to make a sequence minimum-phase. We also recall that if x is a convolution, its values are given by

$$x(n) = \sum_{k} s(k)\, p(n-k).$$

The exponentially weighted sequence can be written

$$\beta^n x(n) = \sum_k \beta^k s(k)\, \beta^{n-k} p(n-k).$$

That is, each member of the convolution is also exponentially weighted. Therefore, exponential weighting can be used to make the impulse train minimum-phase. This

Table 1. Locations and values of the impulses in the complex cepstrum of $p(n) = \delta(n) + \alpha_1\delta(n-n_1) + \alpha_2\delta(n-n_2)$.

k	n	$\hat{p}(n)$	k	n	$\hat{p}(n)$
1	n_1	α_1	5	$5n_1$	$\alpha_1^5/5$
1	n_2	α_2	5	$4n_1+n_2$	$\alpha_1^4\alpha_2$
2	$2n_1$	$-\alpha_1^2/2$	5	$3n_1+2n_2$	$2\alpha_1^3\alpha_2^2$
2	n_1+n_2	$-\alpha_1\alpha_2$	5	$2n_1+3n_2$	$2\alpha_1^2\alpha_2^3$
2	$2n_2$	$-\alpha_2^2/2$	5	n_1+4n_2	$\alpha_1\alpha_2^4$
3	$3n_1$	$\alpha_1^3/3$	5	$5n_2$	$\alpha_2^5/5$
3	$2n_1+n_2$	$\alpha_1^2\alpha_2$	6	$6n_1$	$-\alpha_1^6/6$
3	n_1+2n_2	$\alpha_1\alpha_2^2$	6	$5n_1+n_2$	$-\alpha_1^5\alpha_2$
3	$3n_2$	$\alpha_2^3/3$	6	$4n_1+2n_2$	$-5\alpha_1^4\alpha_2^2/2$
4	$4n_1$	$-\alpha_1^4/4$	6	$3n_1+3n_2$	$-10\alpha_1^3\alpha_2^3/3$
4	$3n_1+n_2$	$-\alpha_1^3\alpha_2$	6	$2n_1+4n_2$	$-5\alpha_1^2\alpha_1^4/2$
4	$2n_1+n_2$	$-2\alpha_1^2\alpha_2^2$	6	n_1+5n_2	$-\alpha_1\alpha_2^5$
4	n_1+3n_2	$-\alpha_1\alpha_2^3$	6	$6n_2$	$-\alpha_2^6/6$
4	$4n_2$	$-\alpha_2^4/4$			

will then insure that the part of the complex cepstrum attributable to p will occur only in the region $n \geq n_1$, where n_1 is the spacing between the first and second impulses.

In general we are concerned with

$$P(z) = 1 + \sum_{k=1}^{M-1} \alpha_k z^{-n_k}, \tag{122}$$

where $n_1 < n_2 < \ldots < n_{m-1}$. Since exponential weighting effectively replaces α_k in Eq. 122 by $\beta^{n_k}\alpha_k$, it is advantageous to have at least a sufficient condition that these quantities must satisfy in order that the sequence be minimum-phase. Such a condition is easily obtained since the requirement of minimum phase is equivalent to requiring that

$$P(z) = \log\left[1 + \sum_{k=1}^{M-1} \alpha_k z^{-n_k}\right] \tag{123}$$

have a power-series expansion converging in a region $|z| > a$, where $0 < a < 1$. A sufficient condition for this to be true is

$$\max_{-\pi<\omega<\pi} \left| \sum_{k=1}^{M-1} \alpha_k e^{-j\omega n_k} \right| < 1. \tag{124}$$

The condition of (124) is satisfied if

$$\sum_{k=1}^{M-1} |\alpha_k| < 1. \tag{125}$$

Thus (125) constitutes a sufficient condition for the sequence corresponding to (122) to be minimum-phase. If the α_k do not satisfy (125), the sequence may still be minimum-phase (as in Example 1); however, in general we can always insure that the impulse train is minimum-phase by choosing β so that

$$\sum_{k=1}^{M-1} \left|\beta^{n_k}\alpha_k\right| < 1.$$

Since we do not generally know in advance the size of the α_k, we must rely on approximate information about the shortest echo time, the spacings, and the relative sizes of the echoes in order to choose the proper value for β.

In conclusion, we note that if the sequence p is minimum-phase, then the complex cepstrum may be obtained from the following expansion of Eq. 123:

$$\hat{P}(z) = \sum_{k=1}^{\infty} (-1)^{k+1} \frac{1}{k} \left(\sum_{r=0}^{M-1} \alpha_r z^{-n_r} \right)^k .$$

In this case we must use the multinomial expansion for $\left(\sum_{r=0}^{M-1} \alpha_r z^{-n_r}\right)^k$, and clearly the result will be a rather complicated distribution of impulses in the complex cepstrum.

We can state, however, that there will be no impulses in the complex cepstrum for $n < n_1$. Furthermore, the impulse sizes will approach zero for large n and will occur at

$$n = \sum_{k=1}^{M-1} m_k n_k,$$

where the m_k take on all positive integer values.

5.3 DISTORTED ECHOES

Up to this point, we have assumed that the echoes were exact replicas of some sequence s. In practice, of course, we are interested in sampled continuous-time

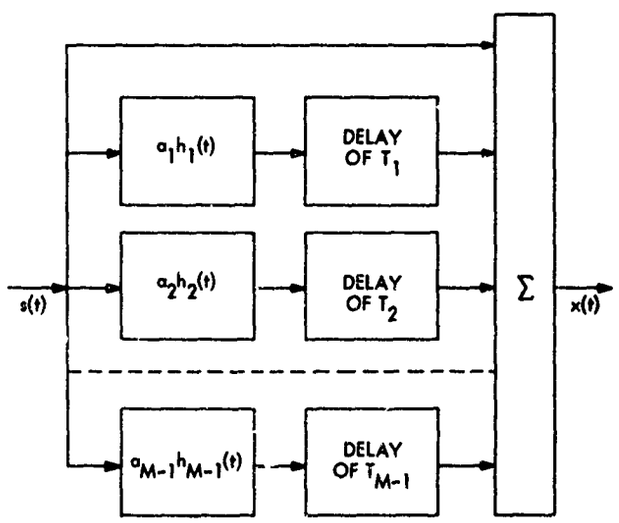


Fig. 34. Linear model for the production of echoes.

signals, and this assumption is generally not exactly correct. A more realistic model for the generation of continuous time signals containing echoes is shown in Fig. 34. The input s(t) is assumed to be bandlimited, and the $h_k(t)$ specify the continuous-time impulse responses of M − 1 linear systems corresponding to M − 1 different echo paths. Thus x(t) is bandlimited, and is given by

$$x(t) = s(t) \otimes [u_o(t) + a_1 h_1(t-T_1) + \ldots + a_{M-1} h_{M-1}(t-T_{M-1})].$$

If we sample x(t) (assuming that the Nyquist frequency is 1), it can be shown that the samples of x(t) are given by

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

where

$$p(n) = \delta(n) + a_1 h_{\Delta_1}(n-n_1) + \ldots + a_{M-1} h_{\Delta_{M-1}}(n-n_{M-1}). \tag{126}$$

In Eq. 126, the n_k are integers and

$$T_k = n_k - \Delta_k,$$

where

$$0 \leq \Delta_k < 1.$$

Thus, the $h_{\Delta_k}(n)$ are the samples of $h_k(t+\Delta_k)$; that is,

$$h_{\Delta_k}(n) = h_k(n+\Delta_k).$$

The Fourier transform of the sequence p is

$$P(e^{j\omega}) = 1 + \sum_{k=1}^{M-1} a_k H_{\Delta_k}(e^{j\omega})\, e^{-j\omega n_k}, \tag{127}$$

where $H_{\Delta_k}(e^{j\omega})$ is the Fourier transform of the samples of $h(t+\Delta_k)$, and

$$H_{\Delta_k}(e^{j\omega}) = H_k(e^{j\omega})\, e^{j\omega\Delta_k}.$$

$H_k(e^{j\omega})$ is the Fourier transform of the samples of $h_k(t)$.

This discrete model accounts for the fact that each echo may be distorted by its transmission path and also for the small shift encountered if the echo delay time is not an integer multiple of the sampling period.

We recall that the analysis for multiple echoes was quite difficult, and complicated expressions were obtained for the complex cepstrum. We also saw, however, that the simple case of a single echo illustrates most of the important concepts. Thus, for clarity, let us consider the special case

$$P(e^{j\omega}) = 1 + aH_{\Delta}(e^{j\omega})\, e^{-j\omega n_1},$$

where $H_{\Delta}(e^{j\omega})$ is the transform of the sequence of samples of a continuous-time, band-limited, impulse response $h(t+\Delta)$. If we assume that

$$\max_{-\pi<\omega<\pi} \left| aH_{\Delta}(e^{j\omega}) \right| < 1,$$

then the logarithm of $P(e^{j\omega})$ can be written

$$\hat{P}(e^{j\omega}) = \log\left[1 + aH_\Delta(e^{j\omega})\, e^{-j\omega n_1}\right]$$

$$= \sum_{k=1}^{\infty} (-1)^{k+1} \frac{a^k}{k}\left[H_\Delta(e^{j\omega})\right]^k e^{-j\omega k n_1}. \tag{128}$$

The condition on $|aH_\Delta(e^{j\omega})|$ can be satisfied for all minimum-phase systems and many nonminimum-phase systems by exponential weighting of x(n), since $\beta^n x(n)$ will have the transform

$$X(\beta^{-1}e^{j\omega}) = S(\beta^{-1}e^{j\omega})\, P(\beta^{-1}e^{j\omega}),$$

where

$$P(\beta^{-1}e^{j\omega}) = 1 + a\beta^{n_1} H_\Delta(\beta^{-1}e^{j\omega})\, e^{-j\omega n_1}.$$

Thus we must choose β so that

$$\max_{-\pi<\omega<\pi} \left| a\beta^{n_1} H_\Delta(\beta^{-1}e^{j\omega}) \right| < 1.$$

The complex cepstrum for (128) is given by

$$\hat{p}(n) = \sum_{k=1}^{\infty} (-1)^{k+1} \frac{a^k}{k}\left[{}^{(k)}h(n-kn_1)\right], \tag{129}$$

where

$${}^{(k)}h(n) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \left[H_\Delta(e^{j\omega})\right]^k e^{j\omega n}\, d\omega. \tag{130}$$

We also note from (130) that ${}^{(k)}h_\Delta(n)$ satisfies

$${}^{(k)}h_\Delta(n) = {}^{(k-1)}h_\Delta(n) \otimes h_\Delta(n)$$

where

$${}^{(0)}h_\Delta(n) = \delta(n).$$

We see that if $\Delta = 0$ and $h_\Delta(n) = \delta(n)$, Eq. 129 reduces to

$$\hat{p}(n) = \sum_{k=1}^{\infty} (-1)^{k+1} \frac{a^k}{k}\, \delta(n-kn_1),$$

as we would expect. If this is not the case, then the impulses in the complex cepstrum are dispersed by convolution with ${}^{(k)}h_{\Delta}(n)$. In general, if the systems $H_{\Delta_k}(e^{j\omega})$ are wideband, then the corresponding impulse response, $h_{\Delta_k}(n)$, will be of short duration. On the other hand, if $H_{\Delta_k}(e^{j\omega})$ is narrow-band, the impulse response will not approach zero very rapidly. In the first case, the impulses in the complex cepstrum will be convolved with relatively sharp pulses, while in the latter case the impulse will be convolved with sequences that are rather broad and dispersed.

To see how this affects our results, let us consider two examples.

Example 4

Consider a single echo path for which $h(n) = a^n \quad n \geq 0$, where $0 < a < 1$. It can be shown that

$$ {}^{(k)}h(n) = \frac{(n+k-1)!}{n!\,(k-1)!}\, a^n \qquad n \geq 0. $$

Thus we see that ${}^{(k)}h(n)$ becomes increasingly spread, because of the successive convolutions, and thus the peaks at large values of kn_1 will be considerably smeared

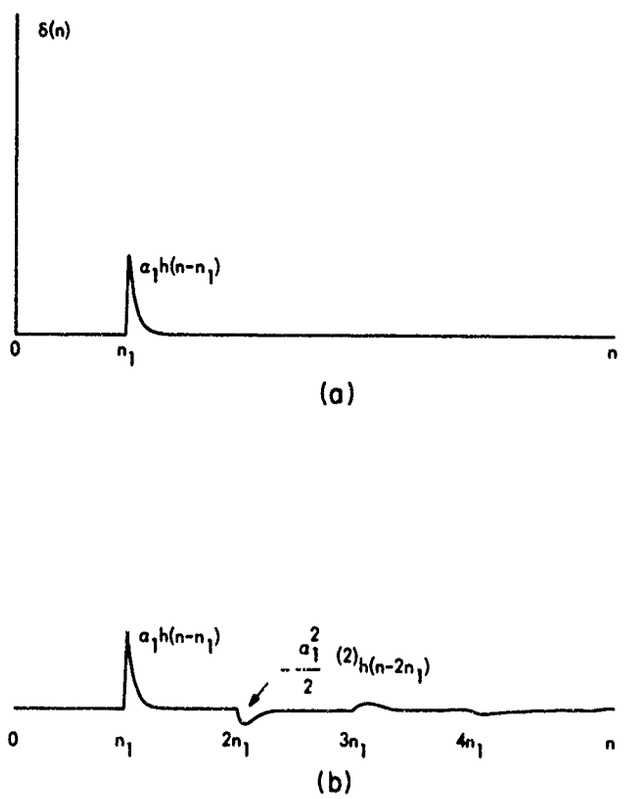


Fig. 35. Impulse train for distorted echoes. (a) Sequence $p(n) = \delta(n) + \alpha_1 h(n-n_1)$ for $h(n) = a^n \quad n \geq 0$. (b) Complex cepstrum for $p(n)$.

out. This is especially true if a is close to 1. Figure 35 shows a plot of p(n) in (a) and $\hat{p}(n)$ in (b) for this example. (The samples have been connected by straight lines for ease in plotting.)

Example 5

Let us consider a single distorted echo obtained by filtering with a system whose system function is

$$H(e^{j\omega}) = e^{j\theta_o sq(\omega)},$$

where $sq(\omega)$ is given by

$$sq(\omega) = 1 \qquad 0 < \omega < \pi$$

$$= -1 \qquad -\pi < \omega < 0,$$

and $sq(\omega) = sq(\omega + r2\pi)$ $r = 0, \pm 1, \pm 2, \ldots$. Such distortions can occur in acoustic-wave reflections.[11] Thus we can write

$$P(e^{j\omega}) = \sum_{k=1}^{\infty} (-1) \frac{a^k}{k} e^{jk\theta_o sq(\omega)} e^{-j\omega k n_1}.$$

It can be shown that the sequence whose transform is $e^{j\theta_o k sq(\omega)}$ has values

$$^{(k)}h(n) = \frac{-2 \sin \theta_o k}{\pi n} \left(\sin \frac{\pi}{2} n\right)^2 \qquad n \neq 0$$

$$= \cos \theta_o k \qquad n = 0.$$

The complex cepstrum $\hat{p}(n)$ is given by

$$\hat{p}(n) = \sum_{k=1}^{\infty} (-1)^{k+1} \frac{a^{k+1}}{k} {}^{(k)}h(n-kn_1).$$

In this case, the shape of the sequence $^{(k)}h(n)$ remains the same for all values of k, but we note that the relative size of the values of the sequence depends on the phase angle θ_o. In this case, and in the previous example wherein a is close to 1, it may be quite difficult to remove the echo, since it will require a comb system to remove the components $\frac{a^k}{k} {}^{(k)}h(n-kn_1)$ without significantly disturbing the part that is due to s(n).

5.4 LINEAR SYSTEMS FOR ECHO REMOVAL AND DETECTION

We have seen that a simple model for a signal containing echoes is the convolution $x = s \otimes p$, where p is an impulse train in which the samples are spaced at intervals greater than one. In section 5.1, we presented an example that indicated that the

components s and p may be recovered by using frequency-invariant linear systems of the form

$$\hat{y}(n) = \ell(n)\,\hat{x}(n).$$

We have shown that the essential concepts of the simple example are true in general, even though, in many cases, the complex cepstrum becomes quite complicated. We now wish to clarify our definitions and terminology and discuss some of the details of using frequency-invariant linear systems for echo removal and detection.

We recall that in section 5.1 we introduced the terms "short-pass," "long-pass," and "comb" systems. Precise definitions of these systems will now be given.

An "ideal short-pass system" is a frequency-invariant linear system for which

$$\begin{aligned} \ell(n) &= 1 \quad -n_- < n < n_+ \\ &= 0 \qquad \text{elsewhere,} \end{aligned}$$

where n_- and n_+ are integers. Such a system is shown in Fig. 36a.

An "ideal long-pass system" is defined as

$$\begin{aligned} \ell(n) &= 0 \quad -n_- < n < n_+ \\ &= 1 \qquad \text{elsewhere,} \end{aligned}$$

where n_- and n_+ are integers. This system is shown in Fig. 36b.

An "ideal comb system" is defined as

$$\begin{aligned} \ell(n) &= 0 \quad n_k - \frac{\Delta n_k}{2} < n < n_k + \frac{\Delta n_k}{2} \qquad k = 1, 2, \ldots \\ &= 1 \qquad \text{elsewhere.} \end{aligned}$$

In general, n_k can be a positive or negative integer. Such a system is shown in Fig. 36c. We note that in computation, the z-transform is replaced by the FFT, and $\hat{x}$ is replaced by the periodic sequence $\hat{\underset{\sim}{x}}$. Thus for computation, $\ell(n)$ is effectively periodic, although we only work with one period of the sequence $\hat{\underset{\sim}{x}}$.

If N [the number of samples of $X(e^{j\omega})$] is fairly large, it is faster to compute the complex cepstrum and multiply by $\ell(n)$ than to do the equivalent convolution of the transform of $\ell(n)$ with the transform of $\hat{\underset{\sim}{x}}$.[17] The ideal systems of Fig. 36 are easily realized in the computer by replacing the contents of appropriate registers with zero. It is well known that such sharp cutoff systems will produce ripples in the transform of $\hat{y}$; therefore, in many cases, it may be preferable to use approximations to these ideal systems which have smoother transitions between one and zero.

In considering the advantages of homomorphic deconvolution over linear inverse filtering, the basic consideration is the amount of information about the signals which is required to design the system. Clearly, in inverse filtering we must have a good approximation to the signal that is to be removed. If we do know this signal, a scheme

that is equivalent to inverse filtering is subtracting its complex cepstrum from that of the input. In this case we could recover the desired signal by using the system D^{-1}. One cannot do better than this. In choosing to remove the undesired component with a linear

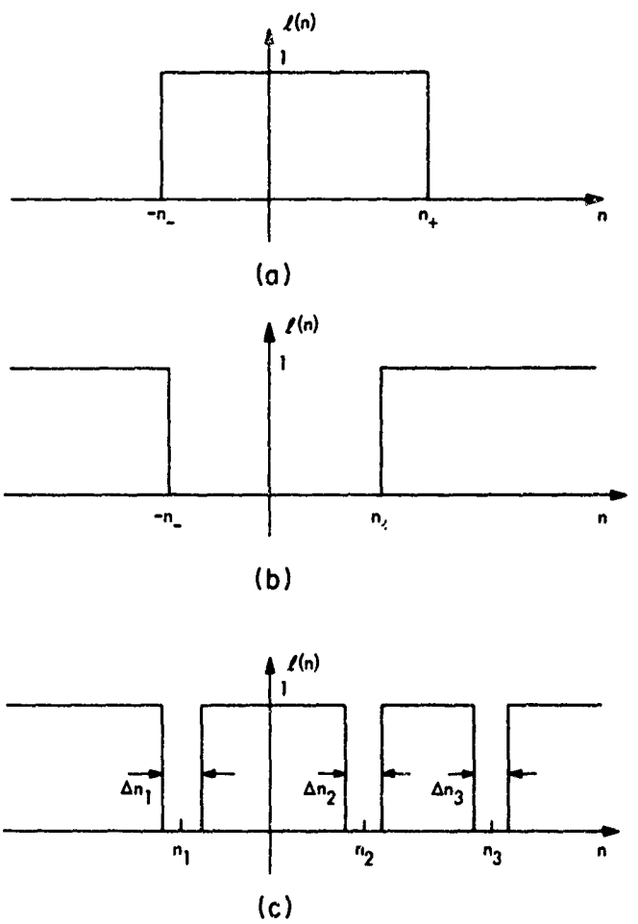


Fig. 36. Ideal frequency-invariant linear systems: (a) short-pass; (b) long-pass; (c) comb.

system, we can expect to obtain only an approximation to the desired signal. The advantage of homomorphic deconvolution is that for signals that are convolutions of an impulse train with a sequence having its samples spaced at unit intervals the complex cepstra of the two components are somewhat separated "in time." In this case, only partial information about the signals is required.

From the examples that we have given, we can see that it is possible to remove either of the two components if we are given only partial information about the impulse train. For example, if the impulse train is minimum-phase (or has been made minimum-phase by exponential weighting), and we know the smallest echo delay, we can recover the impulse train by using a long-pass system that is zero for all n less than the

smallest echo time. We note that the impulses will be convolved with the sequence

$$y_s = D^{-1}[\hat{s} \cdot \ell].$$

The values of this sequence will normally be quite small if the smallest echo time is long enough so that most of the energy of $\hat{s}(n)$ is removed by the long-pass system.

A similar method is suggested for recovering the signal that is convolved with the impulse train. If we use a short-pass system that is equal to 0 for all n greater than the smallest echo time and equal to 1 for n less than that value, we shall remove the component attributable to the impulse train completely. We shall also remove part of the complex cepstrum of s. In most practical situations, that is, when the echoes

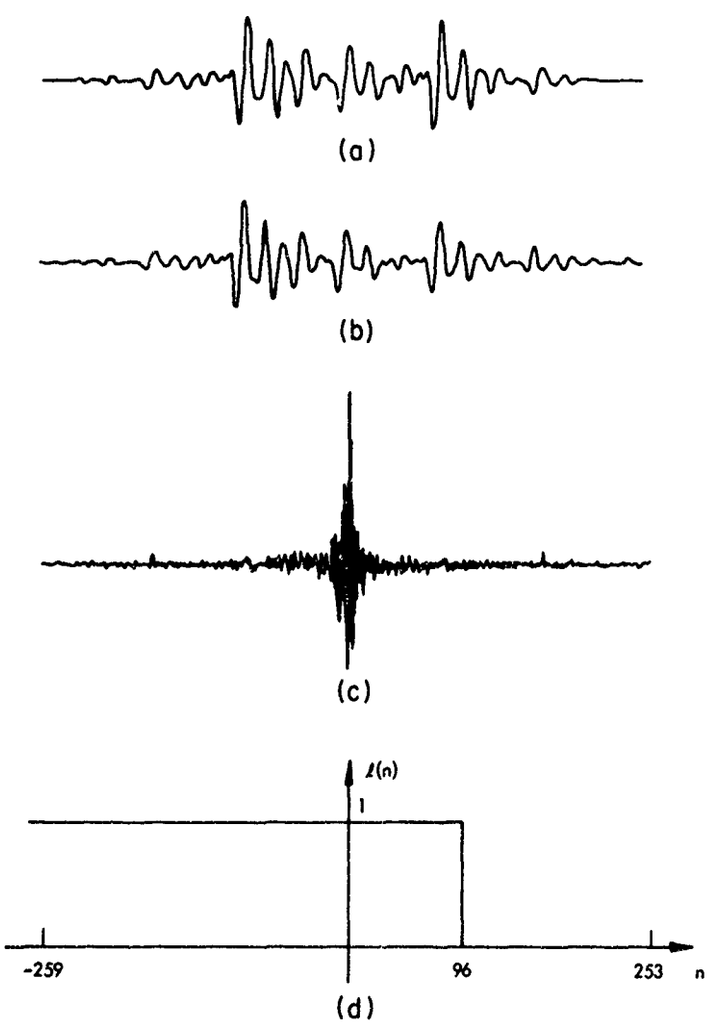


Fig. 37. Effect of a short-pass system. (a) Input sequence. (b) Output for short-pass system of (d). (c) Complex cepstrum for (a). (Note: all traces have the same time origin as $\ell(n)$.)

overlap significantly with s, this method generally requires that we remove too much of $\hat{s}$ to obtain a very good approximation to s at the output of the inverse system D^{-1}. A simple example of this is given in Fig. 32e, where a short-pass system that was equal to 1 only up to n = 16 was used to recover s. Another example is given in Fig. 37. In Fig. 37a we show a segment of a speech waveform (without echoes) which has been weighted with a Hamming window. In Fig. 37c we show the complex cepstrum of this segment, and in Fig. 37d we show a short-pass system. The output of system D^{-1} for this case is shown in Fig. 37b. (The origin of all traces is in line with that of $\ell(n)$.) By comparing Fig. 37a and 37b, we observe that the two are almost identical for values of $n < 50$. (The time axis in all plots coincides with d.) For $n > 50$ we observe that Fig. 37b differs significantly from the original waveform. That this is true in general for short-pass systems can be shown from our discussion in section 2.7.

Recall that for finite-length sequences with no linear phase component

$$X(z) = A \prod_{k=1}^{m_i} \left(1-a_k z^{-1}\right) \prod_{k=1}^{m_o} (1-b_k z).$$

Thus x(n) is zero outside the interval $-m_o \leq n \leq m_i$. Recall that x and y may be written

$$x = x_{min} \otimes x_{max}$$

$$y = y_{min} \otimes y_{max},$$

where

$$\hat{x}_{min}(n) = \hat{x}(n) \qquad n \geq 0$$

$$= 0 \qquad n < 0,$$

$$\hat{x}_{max}(n) = \hat{x}(n) \qquad n < 0$$

$$= 0 \qquad n \geq 0.$$

We also see that for short-pass systems, as in Fig. 37d,

$$y_{min}(n) = x_{min}(n) \qquad 0 \leq n \leq n_c$$

$$= 0 \qquad n < 0$$

$$\neq x_{min}(n) \qquad n_c < n$$

and

$$y_{max}(n) = x_{max}(n) \qquad -m_o \leq n \leq 0$$

$$= 0 \qquad \text{elsewhere.}$$

It is relatively easy to see from these results that

$$
\begin{aligned}
&= 0 && n < m_o \\
y(n) &= x(n) && -m_o \leq n \leq n_c - m_o \\
&\neq x(n) && n_c - m_o < n.
\end{aligned}
$$

In the example of Fig. 27, $m_o = 259$ and $m_i = 253$. Since $n_c = 96$, we see that the result above predicts that y(n) will be different from x(n) for all $n > -163$. This is not detectable visually, until approximately $n = 0$. This is so because for $n > n_c - m_o$, $y_{max}(n)$ is relatively small, and it is only for larger values of n that this error in y_{max} is reflected in y(n).

In general, the short-pass system is not as useful in echo removal as we might hope, since in cases wherein the echoes significantly overlap we are required to remove too much of the complex cepstrum of the desired output.

The alternative to the short-pass system is the comb system. In this case, we need much more information about the impulse train in order to choose the values of n_k and Δn_k. We do not need to know the sizes of the echoes but either we must know their locations or the locations of the impulses in the complex cepstrum. If the echoes have been distorted by a linear system (as discussed in section 5.3), we also need to know the approximate duration of the impulse response in order to choose the Δ_k. In general, we do not have this much information about the impulse train. Since the significant peaks that are due to the impulse train stand out, however, in the complex cepstrum, except for very short echo delays, we may detect these peaks and set the parameters of the comb system appropriately. This can be done in an on-line computation system if a display screen, or other graphical output device, is available, and if the experimenter is capable of interacting with the program. This was done in much of the experimental work which is reported here. It is also clearly possible to program relatively simple algorithms for searching appropriate regions of the complex cepstrum for peaks attributable to the impulse train. The information obtained from such algorithms can then be used to set the parameters of the comb system.

In connection with such algorithms, it is worth pointing out that we have found that the phase component of the sequence s normally contributes a larger component to the complex cepstrum than the log magnitude. Thus, in order to make the detection of impulses in the complex cepstrum easier, it is desirable to use only the even part of the complex cepstrum. If the impulse train is minimum-phase, the impulses in the even part will occur at the same locations as those in the complete complex cepstrum. Thus detection of the impulses in the even part of the complex cepstrum for $n > 0$ is equivalent to detection of the impulses in the complex cepstrum for minimum-phase sequences.

We have discussed the basic forms of frequency-invariant systems that can be used in echo detection and echo removal. Although the short-pass system seems to distort

the desired output too much, it is possible to use a sort of self-adjusting comb system for echo removal. We shall discuss examples of echo removal and echo detection, using the ideas just formulated.

5.5 SHORT-TIME ECHO REMOVAL

We have discussed the properties of the complex cepstrum of sequences of the form

$$x = s \otimes p,$$

where s is a sequence with values spaced at unit intervals, and p is an impulse train. We also showed how frequency-invariant linear systems can be used to recover either s or p. Throughout all of the previous discussion, we have assumed that the sequence x was of finite length and that we were able to compute the z-transform (or FFT) of the entire sequence. In some applications, for example, removal of echoes from speech signals, the duration of the signal and high sampling rate combine to give sequences with a great many samples. To process such sequences all at once, we are required to take the FFT of a long sequence. In turn, for efficient operation, the FFT would need a large amount of high-speed memory (that is, core storage). Thus, we are led to inquire into the possibility of processing such sequences in shorter segments, and then somehow putting these segments back together to form the complete output sequence. This can indeed be done and we shall give an analysis of this procedure.

We shall begin with some definitions. We define the <u>short-time z-transform</u> of the sequence x to be

$$X(\xi, z) = \sum_{n=0}^{L-1} x(\xi, n)\, z^{-n}, \tag{131}$$

where

$$x(\xi, n) = x(\xi+n).$$

The short-time z-transform with window w is defined to be

$$X_w(\xi, z) = \sum_{n=0}^{L-1} x(\xi, n)\, w(n)\, z^{-n}. \tag{132}$$

Clearly, $X_w(\xi, z)$ is just the z-transform of a finite-length, weighted segment of the sequence x, and the parameter ξ simply serves to specify which segment is under consideration. Therefore, if $w(n) \neq 0$ for $0 \leq n < L$, then

$$x(\xi+n) = \frac{1}{2\pi j w(n)} \oint_C X_w(\xi, z)\, z^{n-1}\, dz, \tag{133}$$

where the contour C may be the unit circle.

In particular, if w is an exponential window

$$w(n) = a^n \qquad 0 \leq n < L,$$

we define

$$X(\xi, z, a) = \sum_{n=0}^{L-1} x(\xi+n)\, a^n z^{-n}.$$

We see that

$$X(\xi, z, a) = X(\xi, z/a)$$

and

$$X(\xi, z, 1) = X(\xi, z).$$

With these definitions in mind, let us assume that the sequence x has values

$$x(n) = s(n) + \sum_{k=1}^{M-1} \alpha_k s(n-n_k), \tag{134}$$

where the sequence s is a sequence with unit spacing of samples (such as a speech waveform). The short-time z-transform with exponential window is

$$X(\xi, z, a) = \sum_{n=0}^{L-1} s(\xi+n)\, a^n z^{-n} + \sum_{k=1}^{M-1} \alpha_k \sum_{n=0}^{L-1} s(\xi+n-n_k)\, a^n z^{-n}. \tag{135}$$

If we let $q = n - n_k$, we can write

$$\alpha_k \sum_{n=0}^{L-1} s(\xi+n-n_k)\, a^n z^{-n} = \alpha_k a^{n_k} z^{-n_k} \sum_{q=-n_k}^{L-1-n_k} s(\xi+q)\, a^q z^{-q}. \tag{136}$$

Through some simple manipulations, (136) may be written

$$\alpha_k \sum_{k=0}^{L-1} s(\xi+n-n_k)\, a^n z^{-n} = \alpha_k S(\xi, z, a)\, a^{n_k} z^{-n_k} + E_k(\xi, z, a),$$

where

$$S(\xi, z, a) = \sum_{n=0}^{L-1} s(\xi+n)\, a^n z^{-n} \tag{137}$$

and

$$E_k(\xi, z, a) = \alpha_k a^{n_k} z^{-n_k} \sum_{n=-n_k}^{-1} s(\xi+n)\, a^n z^{-n}$$

$$-\alpha_k a^{n_k} z^{-n_k} a^L z^{-L} \sum_{n=-n_k}^{-1} s(\xi+L+n)\, a^n z^{-n}. \tag{138}$$

Thus Eq. 135 may be written

$$X(\xi, z, a) = S(\xi, z, a)\, P(z/a) + E(\xi, z, a), \tag{139}$$

where

$$P(z/a) = 1 + \sum_{k=1}^{M-1} \alpha_k a^{n_k} z^{-n_k}$$

and

$$E(\xi, z, a) = \sum_{k=1}^{M-1} E_k(\xi, z, a).$$

Let us pause and interpret (139). Suppose that s is finite length, and $s(n) = 0$ for $n < 0$. If $\xi = 0$, and if L is greater than the total length of the sequence x, the terms $E_k(\xi, z, a)$ will all vanish in (139). Under these conditions, we are transforming all of the sequence at once, and we should expect that

$$X(0, z, a) = X(z/a) = S(z/a)\, P(z/a), \tag{140}$$

where $X(z)$ is the z-transform of x. Thus, the term $E(\xi, z, a)$ is appropriately termed "the error in $X(\xi, z, a)$." That is, it is the amount by which Eq. 139 fails to have the form of the right-hand side of Eq. 140. The reason for these errors can be seen from Eq. 138. Let us write (138) as

$$E_k(\xi, z, a) = F_k(\xi, z, a) - a^L z^{-L} F_k(\xi+L, z, a), \tag{141}$$

where

$$F_k(\xi, z, a) = \alpha_k a^{n_k} z^{-n_k} \sum_{n=-n_k}^{-1} s(\xi+n)\, a^n z^{-n}. \tag{142}$$

The first term on the right in (141) is seen to be due to overlap of the kth echo from the previous segment into the present segment. That is, the samples $s(\xi+n-n_k)$ $0 \leq n < n_k$ do not appear in the basic segment $s(\xi+n)$ $0 \leq n < L$. The second term on the right in (141) is due to samples that appear in the segment $s(\xi+n)$ $0 \leq n < L$ and do not appear in the segment of the echo $s(\xi+n-n_k)$ $0 \leq n < L$. We shall refer to $F_k(\xi, z, a)$ as the error at the beginning of the segment, and $-a^L z^{-L} F_k(\xi+L, z, a)$ as the error at the end of the segment. We note that, except for a delay and multiplication by a constant, the error at the end of the segment corresponding to $\xi = \xi_o$ is the negative of the error at the beginning of the segment corresponding to $\xi = \xi_o + L$.

Now let us consider the logarithm of $X(\xi, z, a)$. We define

$$\hat{X}(\xi, z, a) = \log [X(\xi, z, a)] = \log [S(\xi, z, a) P(z/a) + E(\xi, z, a)], \tag{143}$$

and the <u>short-time complex cepstrum</u> as

$$\hat{x}(\xi, n, a) = \frac{1}{2\pi} \int_{-\pi}^{\pi} X(\xi, e^{j\omega}, a)\, e^{j\omega n}\, d\omega. \tag{144}$$

We can write Eq. 143

$$X(\xi, z, a) = \log \left[S(\xi, z, a) + \frac{E(\xi, z, a)}{P(z/a)} \right] + \log [P(z/a)]. \tag{145}$$

From the second term on the right in (145), we see that the short-time complex cepstrum can be thought of as containing the same impulse-train component as the complex cepstrum of the entire sequence. This assumption implies certain restrictions on the length of the window and the nature of the signal s. These restrictions will be discussed below. If this component were removed by a comb system, the short-time transform of the output segment would be simply

$$Y(\xi, z, a) = S(\xi, z, a) + \frac{E(\xi, z, a)}{P(z/a)}. \tag{146}$$

The samples of this output segment are

$$y(\xi+n)\, a^n = \frac{1}{2\pi} \int_{-\pi}^{\pi} Y(\xi, z, a)\, e^{j\omega n}\, d\omega \qquad 0 \leq n < L. \tag{147}$$

The effect of the exponential weighting may be removed by multiplying by a^{-n}, to obtain the sequence of samples which has the transform

$$Y(\xi, z, 1) = S(\xi, z, 1) + \frac{E(\xi, z, 1)}{P(z)}. \tag{148}$$

From (148) it is clear that in the interval $0 \leq n < L$, the output segment is the sum of the desired output $s(\xi+n)$ and a sequence related to the error in $X(\xi, z, 1)$. To see

the nature of this term, it is helpful to consider some special cases.

Example 6

Assume that M = 2; that is, there is only a single echo. Then from Eq. 141 we have

$$Y(\xi, z, 1) = S(\xi, z, 1) + \frac{E_1(\xi, z, 1)}{1 + \alpha_1 z^{-n_1}}.$$

Thus the output segment has values

$$y(\xi+n) = s(\xi+n) + \sum_{r=0}^{\infty} (-\alpha_1)^r e_1(\xi+n-rn_1),$$

for $n \geq 0$. This example is illustrated in Fig. 38.

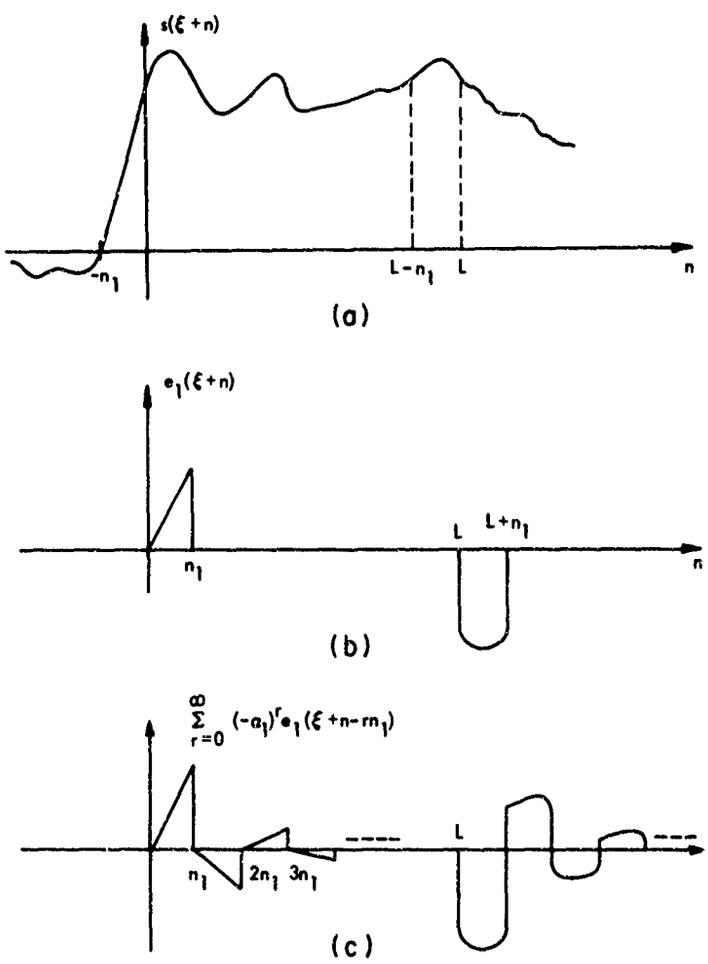


Fig. 38. Errors for short-time echo removal. (a) The sequence $s(\xi+n)$. (b) Error in the input for a single echo with $\alpha_1 = 1/2$, n_1. (c) Error in the output.

Example 7

Assume that $\alpha_k = \alpha^k$ and $n_k = kn_1$. Therefore, from Eq. 148,

$$Y(\xi, z, 1) = S(\xi, z, 1) + \frac{\sum_{k=0}^{M-1} E_k(\xi, z, 1)}{\sum_{k=0}^{M-1} \alpha^k z^{-kn_1}},$$

which can be written

$$Y(\xi, z, 1) \approx S(\xi, z, 1) + \left(1 - \alpha z^{-n_1}\right) \sum_{k=1}^{M-1} E_k(\xi, z, 1)$$

if a^M is small. Thus the output segment has values

$$y(\xi+n) \approx s(\xi+n) + \sum_{k=1}^{M-1} [e_k(\xi+n) - \alpha e_k(\xi+n-kn_1)].$$

Using Eq. 141, we can write

$$\sum_{k=1}^{M-1} E_k(\xi, z, 1) = \sum_{k=1}^{M-1} F_k(\xi, z, 1) - z^{-L} \sum_{k=1}^{M-1} F_k(\xi+L, z, 1),$$

and using Eq. 142 and some manipulation, we obtain

$$\left(1 - \alpha z^{-n_1}\right) \sum_{k=1}^{M-1} F_k(\xi, z, 1) = \alpha z^{-n_1} \sum_{n=-n_1}^{-1} s(\xi+n)\, z^{-n} + \alpha^2 z^{-2n_1} \sum_{n=-2n_1}^{-n_1-1} s(\xi+n)\, z^{-n}$$

$$+ \ldots + \alpha^{M-1} z^{-(M-1)n_1} \sum_{n=-(M-1)n_1}^{-(M-2)n_1-1} s(\xi+n)\, z^{-n}$$

$$- \alpha^M z^{-Mn_1} \sum_{n=-Mn_1}^{-1} s(\xi+n)\, z^{-n}. \tag{149}$$

Careful examination of the terms in (149) shows that all of the terms on the right except the last contribute to the error of $y(\xi+n)$ only in the interval $0 \leq n < n_1$, while the last term is quite small but does contribute over the interval $0 \leq n < Mn_1$. An identical result holds for the error at the end of the segment so that the error term

$$\sum_{k=1}^{M-1} [e_k(\xi+n)-ae_k(\xi+n-kn_1)]$$

will be approximately zero everywhere but in the intervals $0 \leq n < n_1$, and $L \leq n < L+n_1$. An example of this is shown in Fig. 42f where the first two traces correspond to $0 \leq n < L$, and the last two traces correspond to $L \leq n < 2L$.

Example 8

Let there be two echoes not necessarily equally spaced. Then, from Eq. 148,

$$Y(\xi, z, 1) = S(\xi, z, 1) + \frac{E_1(\xi, z, 1) + E_2(\xi, z, 1)}{1 + a_1 z^{-n_1} + a_2 z^{-n_2}}.$$

We could proceed as before. In this case, however, the equations become so complex as to be relatively useless. We can see from Eq. 138 that the sequence corresponding to

$$E_1(\xi, z, 1) + E_2(\xi, z, 1)$$

will be nonzero only in the intervals $0 \leq n < n_2$ and $L \leq n < L + n_2$. This sequence is then convolved with the sequence corresponding to $\left(1 + az^{-n} + a_2 z^{-n_2}\right)^{-1}$ which will be an impulse train.

The three previous examples have several things in common. We note that the term

$$E(\xi, z, 1) = \sum_{k=1}^{M-1} E_k(\xi, z, 1)$$

is the amount by which $X(\xi, z, 1)$ fails to be a product of $S(\xi, z, 1)$ and $P(z)$ the transform of the impulse train. We have called this the error in $X(\xi, z, 1)$. Similarly, we note that

$$\frac{E(\xi, z, 1)}{P(z)}$$

is the amount by which $Y(\xi, z, 1)$ fails to be the desired output $S(\xi, z, 1)$. We shall refer to this as the error in $Y(\xi, z, 1)$. Note that we use the term "error" in a slightly different sense in this case. The error in the output can be thought of as the error in the input, passed through the inverse system for the impulse train that represents the echoes. We have seen that the error in the output consists of two segments, one of which is primarily in the interval $0 \leq n < L$, and a second segment which is in the region $L \leq n$. We have also seen that in general these errors tend to become small for large n. Furthermore, the error in the interval $L \leq n$ for the segment $\xi = \xi_0$ is the negative of the error in the interval $0 \leq n < L$ for the segment $\xi = \xi_0 + L$. This fact will be used when

we discuss putting the output segments back together.

In obtaining the previous results it was assumed that the short-time complex cepstrum can be thought of as containing a component that is equal to the complex cepstrum of the impulse train that produces the echoes. In order for this to be true, the length of the segment L must be large compared with n_{M-1}, the longest echo time. Furthermore, the signal s must change character over the segment of interest. These points are clear intuitively, but quantitative results appear to be difficult to formulate. Clearly, we want the error segments in the input to be short compared with the total length L. Since the length of these error segments is equal to the longest echo delay, we require

$$L \gg n_{M-1}.$$

This means that the corresponding errors in the output will be concentrated primarily in the two intervals $0 \leq n < L$ and $L \leq n < 2L$.

With respect to the character of the signal s, let us consider an example. Suppose s is a sine wave so that

$$\begin{aligned} s(n) &= \sin(n) \qquad && n \geq 0 \\ &= 0 && n < 0. \end{aligned}$$

If the sequence x has values

$$x(n) = s(n) + \alpha s(n-n_1),$$

it is clear that all segments of x for which $\xi > n_1$ will look just like a sine wave with some phase shift. That is,

$$\begin{aligned} x(n) &= \sin n + \alpha \sin(n-n_1) \\ &= (1+\alpha\cos n_1)\sin n - \sin n_1 \cos n \qquad && n \geq n_1 \\ &= A \sin(n+\theta). \end{aligned}$$

All periodic sequences suffer the same difficulty, as well as exponential sequences of the form a^n $n \geq 0$. In all of these cases, the short-time complex cepstrum will not exhibit an impulse train because of the echo. Speech signals, for example, change character as time progresses. Clearly, the requirements on the character of the waveform and length of the segment L are interrelated. If we make the segment long enough, almost any nonperiodic signal will change sufficiently across the segment.

The previous results were based on the z-transform, whereas we shall realize such processing by using the FFT. This, of course, means that the short-time complex cepstrum that we compute will be aliased, and will have values given by

$$\hat{\underset{\sim}{x}}(\xi, n, a) = \sum_{r=-\infty}^{\infty} \hat{x}(\xi, n+rN, a) = \frac{1}{N}\sum_{k=0}^{N-1} \log\left[\underset{\sim}{X}(\xi, k, a)\right] e^{j\frac{2\pi}{N}kn},$$

where $N \geq L$, and

$$\underset{\sim}{X}(\xi, k, a) = \sum_{n=0}^{N-1} x(\xi, n)\, a^n\, e^{-j\frac{2\pi}{N}kn}.$$

The resulting output of the system D^{-1} after removing the impulse train caused by the echoes is approximately the exponentially weighted output

$$\sum_{r=-\infty}^{\infty} a^{n+rN} y(\xi+n+rN).$$

If we unweight with a^{-n} in the interval $0 \leq n < N$, we obtain approximately

$$\underset{\sim}{y}(\xi, n) \approx y(\xi, n) + a^N y(\xi, n+N) \tag{150}$$

for $0 \leq n < N$. All indices are taken modulo N.

From (150) we see that if $N = L$, the error in $y(\xi, n)$ is primarily at the beginning of the segment, since it is composed primarily of the error in the beginning of $y(\xi, n)$ plus the error at the end of $a^N y(\xi, n+N)$. If we choose $N = 2L$, the error in $\underset{\sim}{y}(\xi, n)$ is approximately the error in $y(\xi, n)$. Thus, in this case, we can apply most of the present results even though aliasing does occur.

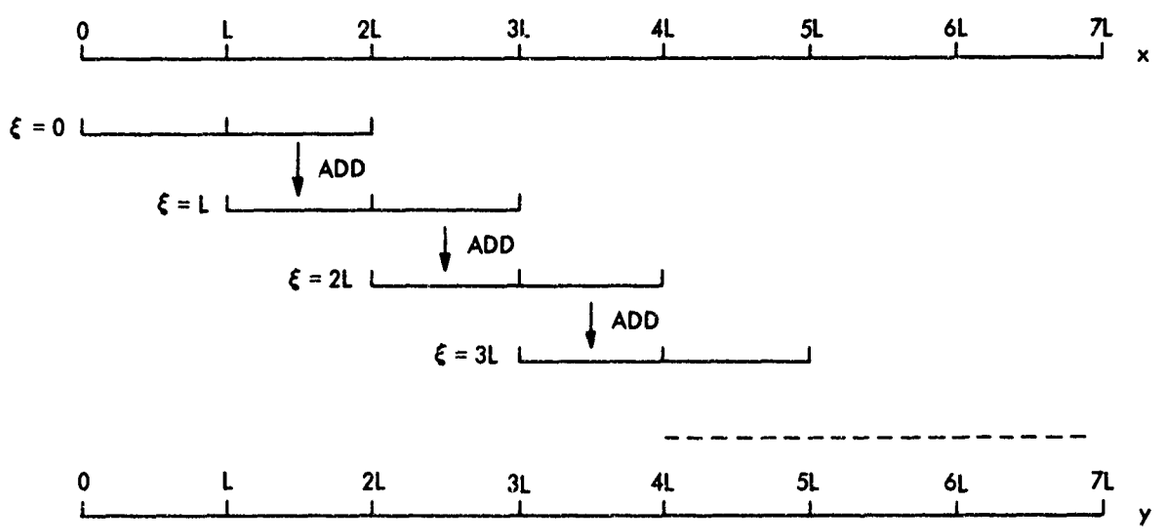


Fig. 39. Correction method for short-time echo removal.

Having discussed the actual computation of the short-time complex cepstrum and resulting output segment, we are able to indicate how the output segments may be put together to form the total output sequence. We have noted that if the errors in the

output tend to approach zero, the error in the interval $L \leq n < 2L$ for the segment $\xi = \xi_0$ is approximately the negative of the error in the interval $0 \leq n < L$ for the segment $\xi = \xi_0 + L$. This suggests that we can move the window along in jumps of L samples. Each time we save the output for $L \leq N < 2L$ from the segment $\xi = \xi_0$ so that it may be added to the values in the interval $0 \leq n < L$ of the segment $\xi = \xi_0 + L$. This method is illustrated in Fig. 39. We refer to this method as the "correction method."

As an alternative to the above scheme, we note that if the segment is quite long compared to the longest echo time, the error is generally small over some interval $L_0 \leq n < L$ so that we may retain only this part which is relatively free of error. This

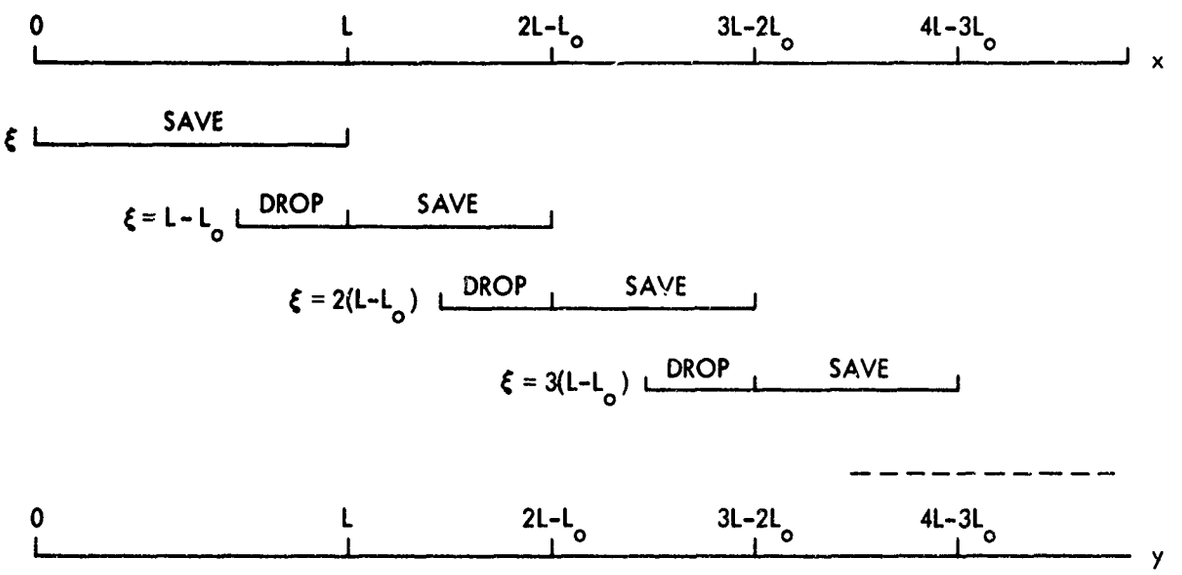


Fig. 40. Overlap method of short-time echo removal.

scheme is shown in Fig. 40, and it is referred to as the "overlap method." Examples indicating the feasibility of these schemes will be given.

5.6 REMOVAL OF ECHOES FROM SPEECH SIGNALS

We shall present some examples of the application of the previous results. We shall show several examples of the removal of computer-simulated echoes from speech signals. In all figures, each trace in a given picture represents 1024 samples unless otherwise noted. The consecutive traces represent consecutive 1024 sample segments of the speech waveform. The entire waveform corresponds to the sentence, "A pot of tea helps to pass the evening." The sampling rate was 10 kHz. In all examples, we used an exponential window with a = 0.9987 and L = 2048. The value of N for the FFT was N = 4096. The segment of speech was, therefore, augmented with 2048 zeros before transformation.

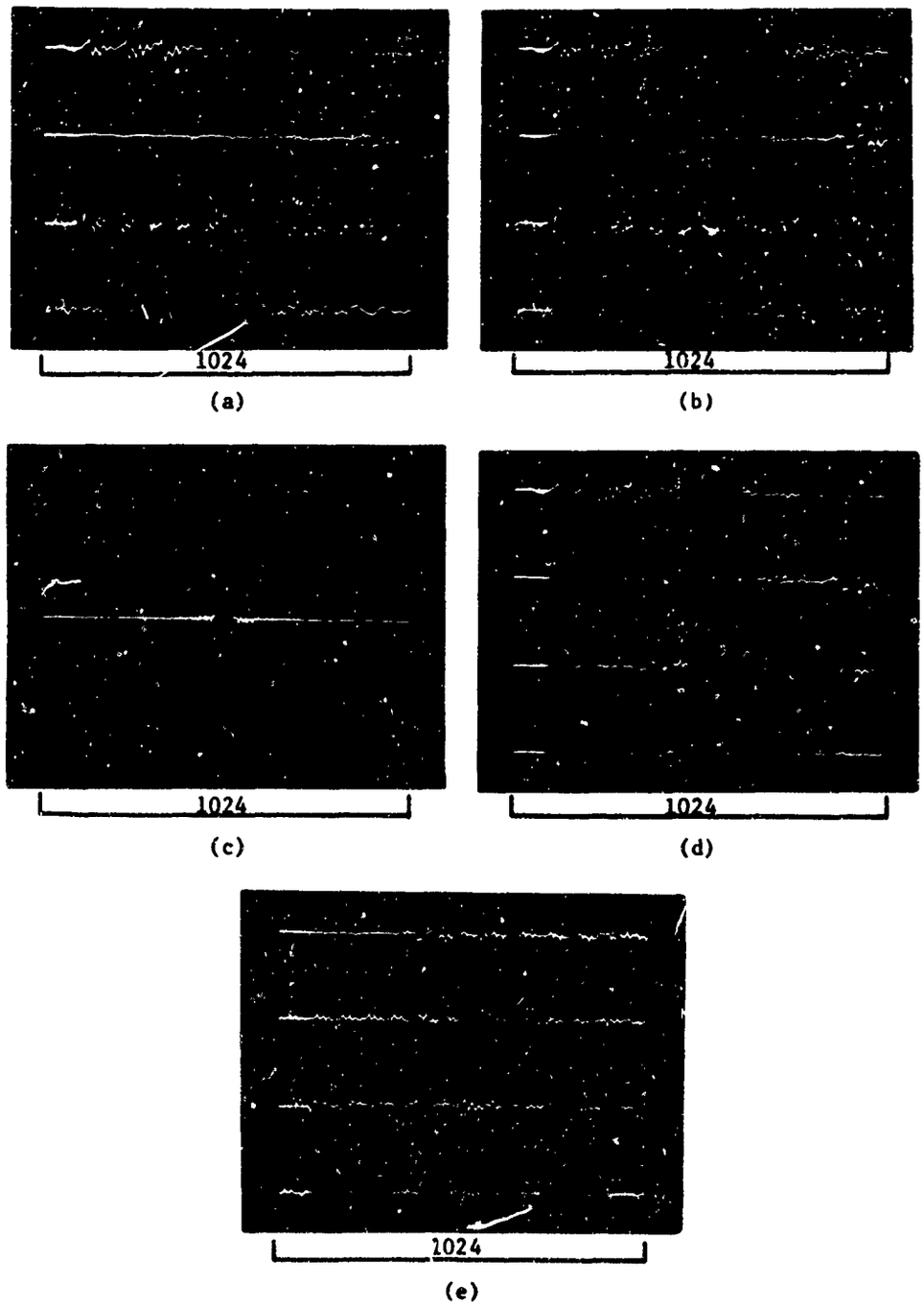


Fig. 41. Short-time echo removal for an impulse train

$p(n) = \delta(n) + 3/4\ \delta(n-500)$.

(a) Signal $s(n)$. (First 4096 samples.)
(b) Signal $x(n) = s(n) \otimes p(n)$.
(c) Complex cepstrum for the first 2048 samples of x.
(d) Output for the first 2048 samples.
(e) Error for the second 2048 samples.

Example 9

The speech signal was convolved with

$$p(n) = \delta(n) + \frac{3}{4}\delta(n-500).$$

The first 4096 samples of the signal s and the sequence x are shown in Fig. 41a and 41b. The complex cepstrum for the segment corresponding to $\xi = 0$ is shown in Fig. 41c. We note the impulse appearing at $n = 500$. The result of removing the impulses at $n = 500$, 1000, etc., using a comb system, and then transforming the result with the system D^{-1} is shown in Fig. 41d. The waveform in the interval $0 \leq n < 2048$ is indistinguishable by eye from the first two traces of Fig. 41e. We also note that the error at the end of the segment appears repeated as predicted by Example 6. There is no error at the beginning of the segment, since the speech signal was essentially zero for $n < 0$. The segment corresponding to $\xi = 2048$, that is, the second two traces in Fig. 41b, was processed in the same way, and the difference between the resulting output and the original segment augmented with zeros (the error in the output) is shown in Fig. 41e. We note that this is just the error in the output, and we see clearly that the first two traces in Fig. 41e are approximately the negative of the last two traces in Fig. 41d. If the last two traces of Fig. 41d were added to the first two traces in Fig. 41e (that is, the segment for $\xi = 2048$), the error in the interval $0 \leq n < L$ would be essentially eliminated. Similarly, we could save the last two traces of Fig. 41e, to use as a correction for the next segment.

We also can see from Fig. 41e that the error is relatively small in the interval $1024 \leq n < 2048$. This suggests that we could also use $L = N = 4096$, and disregard the first 1024 samples in each output segment, as suggested by Fig. 40. In this case, we would obtain 3072 samples per segment, rather than 2048 as in the correction method.

Example 10

In this case, the echoes were specified by

$$p(n) = \sum_{k=0}^{M-1} \left(\frac{3}{4}\right)^n \delta(n-k500).$$

The first 4096 samples of the sequences s and x are shown in Fig. 42a and 42b. The short-time complex cepstrum obtained from the first two traces of Fig. 42b augmented with 2048 zeros is shown in Fig. 42c. The impulses at $n = 500$, 1000, etc., were removed with a comb system, and the resulting output is shown in Fig. 42d. The second two traces in Fig. 42b were processed similarly, and the difference between this output and the last 2 traces of Fig. 42a augmented with 2048 zeros (the desired output) is shown in Fig. 42f. We note that the error is concentrated in the interval $0 \leq n < 500$, as predicted by Example 7. We also see that the error in the interval $0 \leq n < 2048$ in Fig. 42f is the negative of the error in the interval $2048 \leq n < 4096$ in Fig. 42d.

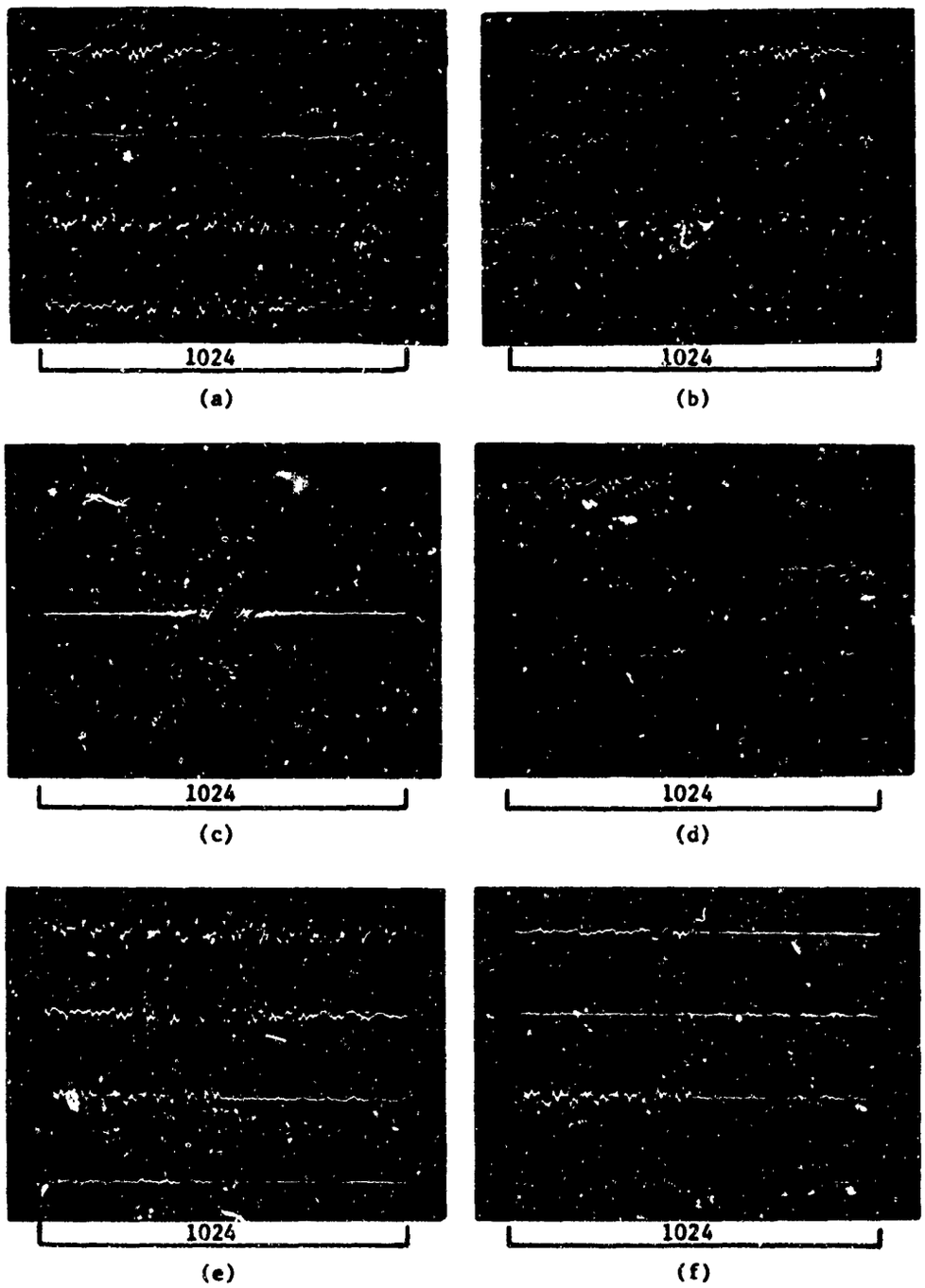


Fig. 42. Short-time echo removal for an impulse train

$$p(n) = \sum_{k=0}^{M-1} (3/4)^k \, \delta(n-k500).$$

(a) Signal s(n). (First 4096 samples.)
(b) Signal $x(n) = s(n) \otimes p(n)$.
(c) Complex cepstrum for the first 2048 samples of x.
(d) Output for the first 2048 samples.
(e) Output for the second 2048 samples.
(f) Error for the second 2048 samples.

Example 11

As a final example, the speech signal was convolved with

$$p(n) = \delta(n) + \frac{3}{4}\,\delta(n-320) + \frac{1}{2}\,\delta(n-500).$$

Again, the first 2048 samples of the resulting sequence were augmented with 2048 zeros and processed as before. The error for the first output segment is shown in Fig. 43a. The error for the second segment ($\xi = 2048$) is shown in Fig. 43b. In Fig. 43c we show the sum of the last two traces of Fig. 43a and the first two traces of Fig. 43b. We note that the errors clearly tend to cancel. Significant error would remain in the second segment, however. This error is primarily due to the fact that the window is not long enough relative to the echo time. (To see the size of the error relative to the desired output see Fig. 41a.)

All of the previous examples have indicated that it is possible to put the output segments back together in a meaningful way either by using the correction method or the overlap method. This was done, in fact, for several different variations of echo times, and the resulting output speech was converted to analog form for listening. Informal listening tests showed that if a suitable value of L is chosen, echoes can be removed from speech signals by using these techniques. The processed speech was slightly more noisy than the input speech. This noise in the output is attributable to the fact that there are small errors remaining in each segment, as in Fig. 43c.

5.7 EFFECT OF ADDITIVE NOISE

The examples that we have shown were carried out on signals with a high signal-to-noise ratio. Suppose that we have a sequence x which is of the form

$$x = s \otimes p + g,$$

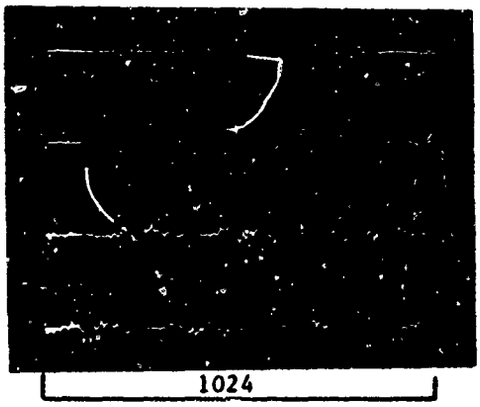
where s is the desired signal, p is an impulse train, and g is an additive noise sequence. The short-time transform of a segment of the sequence x is

$$X(\xi, z) = S(\xi, z)\,P(z) + E(\xi, z) + G(\xi, z),$$

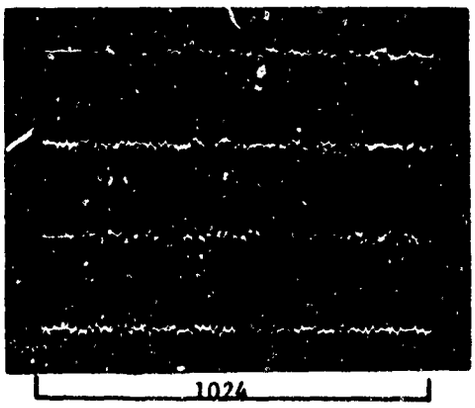
where $S(\xi, z)$ is the short-time transform of the signal, $P(z)$ is a polynomial in z^{-1} that is the transform of the impulse train, $E(\xi, z)$ is the error as defined in section 5.5, and $G(\xi, z)$ is the short-time transform of the noise. If the noise level is low, we may again assume that the short-time complex cepstrum has a component that is due to $P(z)$ which can be removed by a comb system. The output sequence will then be of the form

$$Y(\xi, z) = S(\xi, z) + \frac{E(\xi, z) + G(\xi, z)}{P(z)}. \tag{151}$$

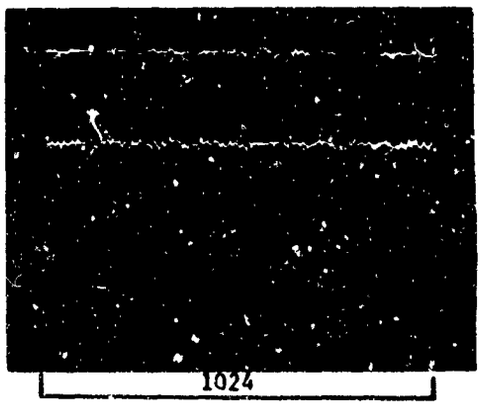
Thus, in addition to the errors previously discussed, we see that the noise in the



(a)



(b)



(c)

Fig. 43. Short-time echo removal for the impulse train

$$p(n) = \delta(n) + (3/4)\,\delta(n-320) + (1/2)\,\delta(n-500).$$

(a) Error in the output for the first 2048 samples of $x(n) = s(n) \otimes p(n)$. ($s(n)$ is the same as in Fig. 41.)
(b) Error in the output for the second 2048 samples of x.
(c) Sum of the second two traces of (a) and the first two traces of (b).

Fig. 44. Short-time echo detection

$$p(n) = \sum_{k=0}^{M-1} (3/4)^k \, \delta(n-k832).$$

(a) Complex cepstrum of the second 2048 samples of $x(n) = s(n) \otimes p(n)$. ($s(n)$ is as in Fig. 41a.)

(b) Output for a long-pass system.

$$p(n) = \delta(n) + h(n-832)$$

$$h(n) = (3/4)^n/4 \qquad n \geq 0$$

$$= 0 \qquad n < 0$$

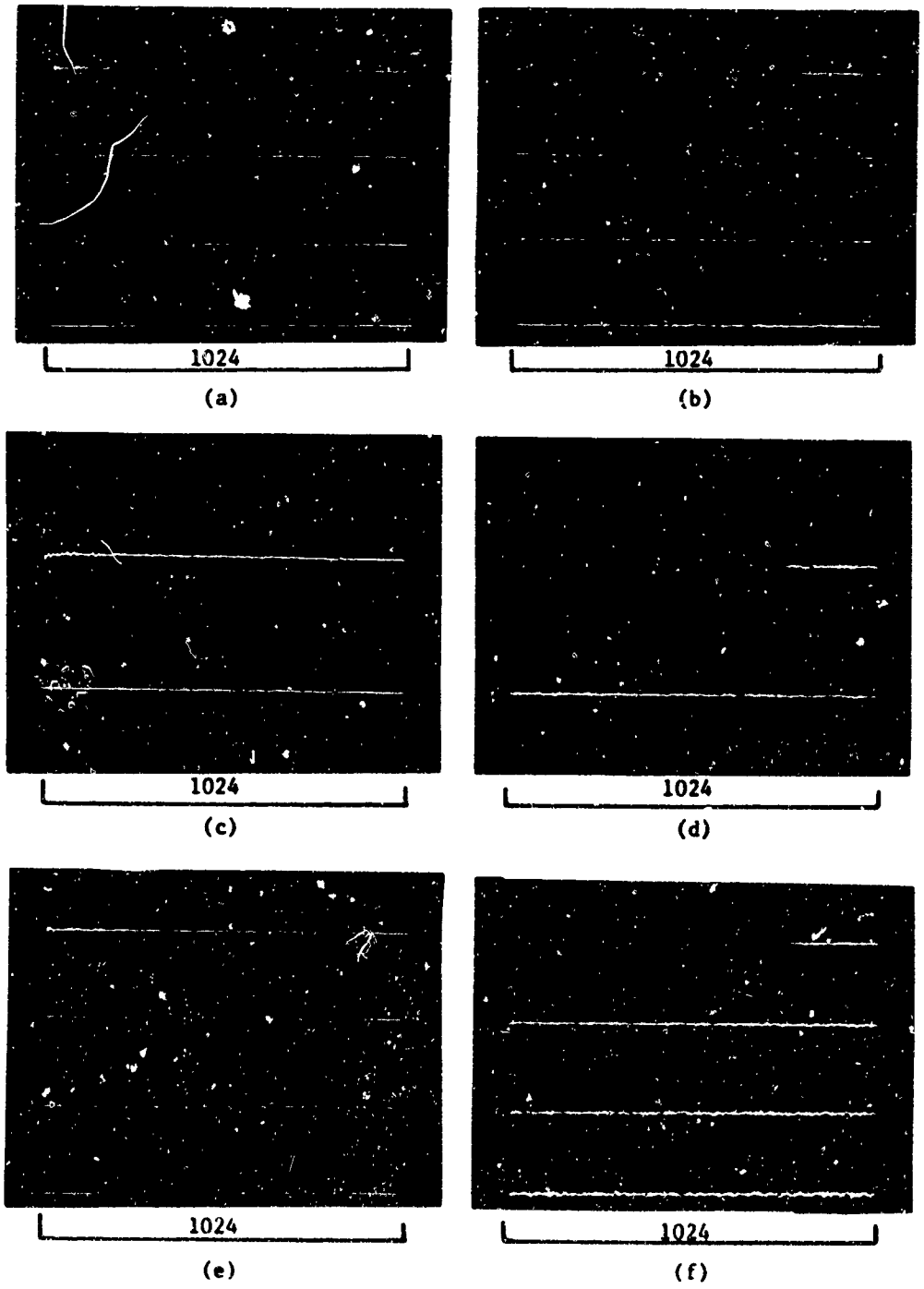
(c) Complex cepstrum for the second 2048 samples of x.

(d) Output for a long-pass system.

$$p(n) = \delta(n) + (3/4)\,\delta(n-832) + (1/2)\,\delta(n-1536).$$

(e) Complex cepstrum for the second 2048 samples of x.

(f) Output for a long-pass system.


1024
(a)
1024
(b)
1024
(c)
1024
(d)
1024
(e)
1024
(f)

input segment also appears in the output. We note that the noise is also effectively filtered by the linear system corresponding to $1/P(z)$.

To test the effectiveness of short-time echo removal at low signal-to-noise ratios, we added a white Gaussian noise sequence to the sequence $s \otimes p$, where p was an impulse train with equal spacing. Signal-to-noise ratios as low as 15 dB were used. The sequences were processed as before, and the resulting outputs were converted to analog form. Informal listening tests for the speech signals showed that in the examples, at least, the echoes were removed and the noise level of the output was about the same as that of the input. We should point out, however, that only limited experimentation was done, and a clear understanding of the effect of additive noise is yet to be obtained.

5.8 DETECTION OF ECHOES

We have considered a signal containing echoes to be represented by a convolution of a basic signal and an impulse train. We have also seen that the complex cepstrum of an impulse train is itself an impulse train. We have shown how echoes may be removed by using a linear frequency-invariant system (possibly on a short-time basis). In conclusion, let us consider the problem of detection of echoes, that is, recovery of the impulse train $p(n)$.

As might be expected, the problem of echo detection is not as difficult as the problem of extraction of the signal. We have seen that in simple cases, the complex cepstrum, or its even part, may be used for echo detection. If the impulse train is not equally spaced, however, the problem of determining the number and locations of all of the echoes from the complex cepstrum becomes quite difficult. It is therefore interesting to consider some examples of recovery of the impulse train.

If the impulse train is minimum-phase (or has been made minimum-phase by exponential weighting), then we have seen that the impulses attributable to the echoes occur only in the region $n \geq n_1$, where n_1 is the shortest echo delay. Thus a long-pass system that multiplies by zero for all $n < n_1$ can be used to recover an approximation to the impulse train. The calculations can be further simplified if the impulse train is minimum-phase, since we can use the log magnitude alone to compute the even part of the complex cepstrum. Then the long-pass system can be chosen to perform the Hilbert transform operation on the impulse train, as well as remove most of the minimum-phase part of the signal.

We have contended that the short-time complex cepstrum can be thought of as having a component caused by the impulse train. If this is so, then we also should be able to carry out short-time echo detection. To show that this is true, let us consider some examples.

Example 12

The speech signal was convolved with

$$p(n) = \sum_{k=0}^{M-1} (3/4)^k \, \delta(n-k832).$$

The input was exponentially weighted and the even part of the short-time complex cepstrum was computed by using only the log magnitude of the FFT. The length of each segment was $L = 4096$ samples and $N = 8192$. In Fig. 44a we show the cepstrum for the second segment ($\xi + 4096$). Each trace corresponds to 1024 samples. Note the impulses at 832, 1664, and so forth. After multiplying by 2 for $n \geq 800$ and by 0 for $n < 800$, the output of the inverse system appears as in Fig. 44b. We have recovered a very good approximation to $p(n)$.

Example 13

The speech signal was convolved with

$$p(n) = \delta(n) + 3/4\, \delta(n-832) + \frac{1}{2}\delta(n-1536).$$

As in the previous example, $L = 4096$ and $N = 8192$. In Fig. 44e and 44f we show the cepstrum and output for the second segment ($\xi = 4096$). Again, it is clear that the output is a very good approximation to $p(n)$.

Example 14

The speech signal was convolved with

$$p(n) = \delta(n) + h(n-832),$$

where

$$h(n) = \frac{1}{4}(3/4)^n \qquad n \geq 0$$

$$= 0 \qquad n < 0.$$

In Fig. 44c we show the cepstrum, and in Fig. 44d we show the output after processing as in the other examples. We see that the impulses in the cepstrum are convolved with the sequences ${}^{(k)}h(n)$, as discussed in section 5.3. The output also shows that we have recovered a good approximation to $p(n)$. Note how the impulses are dispersed in the cepstrum. This, of course, makes it even more difficult to detect the echoes in the cepstrum.

VI. CONCLUSION

6.1 SUMMARY

We have presented a new approach to separating convolved signals. A detailed analysis of homomorphic systems for deconvolution has been given, and we have shown how such signal transformations may be realized by using a digital computer. As an application we have considered the class of signals in which one of the members of the convolution is an impulse train. Although our examples have dealt primarily with speech analysis and the removal of echoes from speech signals, it should be emphasized that almost all of our results, particularly in Section II, apply to more general situations. Therefore, it is felt that the point of view that is reflected in this work is important and our results have demonstrated that homomorphic deconvolution may be a useful approach in many interesting problems.

6.2 SUGGESTIONS FOR FUTURE RESEARCH

Although some interesting results have been obtained, there are still significant questions that would be worthy of further investigation. For example, it is quite possible that other computational realizations can be obtained. This comment is prompted by our observation in section 2.7 that for sequences of length M, a total of M values of the complex cepstrum suffice to completely determine the original sequence. For nonminimum phase sequences, a direct method of computation of the necessary values of the complex cepstrum which would avoid aliasing would be a worthwhile result.

Another issue is the question of appropriate window functions to use in short-time analysis. In speech analysis, this is an important consideration. It is also possible that other weighting sequences besides the exponential can be found which would tend to minimize the errors that we have discussed with respect to short-time echo removal. In both speech analysis and echo removal, it would also be useful to have more general results on the complex cepstrum of an impulse train with nonuniform spacing.

One of the issues that we have only touched upon is the effect of additive noise. Limited experimental results have been obtained, but adequate understanding of this issue is a challenging problem.

As well as the issues relating to carrying out homomorphic deconvolution, it is of interest to consider situations in which our techniques might be successfully applied. The present work has shown that there are clearly advantages to homomorphic deconvolution when one of the signals is an impulse train. Given the ease with which one can think of signals of this class, it seems clear at this point that the techniques presented here should find application in many diverse areas. For example, it is possible that homomorphic deconvolution may be used to obtain very accurate synchronous pitch detection. Also, it appears that there are possibilities of application to seismic signals, for both dereverberation and detection. Still another area may be in processing underwater acoustical signals.

APPENDIX

Vector Space for Convolution

A.1 DEFINITIONS

A.1.1 Field

A field consists of a set of objects called scalars (in our case numbers), together with two operations called addition and multiplication which satisfy the following conditions.[25]

1. To every pair of scalars a and b, there corresponds a scalar a + b called the sum of a and b such that

(a) $a + b = b + a$

(b) $a + (b+c) = (a+b) + c$

(c) there is a unique zero scalar, such that $a + 0 = a$

(d) to every scalar a there corresponds a unique scalar $-a$ such that $a + (-a) = 0$.

2. To every pair of scalars a and b there corresponds a scalar ab called the product of a and b such that

(a) $ab = ba$

(b) $a(bc) = (ab)c$

(c) there exists a unique scalar 1 called one such that $a1 = a$,

(d) to every nonzero scalar a, there corresponds a unique scalar a^{-1} such that $aa^{-1} = 1$

(e) $a(b+c) = ab + ac$.

Examples of a field are the sets of real numbers, of rational numbers, and of complex numbers, where addition and multiplication have their usual meaning.

A.1.2 Vector Space

A vector space consists of a field of scalars, together with a collection of elements called vectors having the following properties.[25]

1. To every pair of vectors x and y there corresponds a vector x + y called the sum of x and y such that

(a) addition is commutative, $x + y = y + x$

(b) addition is associative, $w + (x+y) = (w+x) + y$

(c) there is a unique zero vector such that $x + 0 = x$

(d) there is a unique inverse vector $x + (-x) = 0$.

2. To every scalar a and vector x there corresponds a vector ax such that

(a) $a(bx) = (ab)x$

(b) $1x = x$ for every vector x

(c) $a(x+y) = ax + ay$

(d) $(a+b)x = ax + bx$.

We shall verify that convolution is an appropriate operation for vector addition, and attempt to clarify the meaning of scalar multiplication for convolutional vector spaces.

A.2 CONVOLUTION AS VECTOR ADDITION

Our set of vectors is the set of all sequences whose z transforms exist and have overlapping regions of convergence. The operation of convolution is

$$x \otimes y = \sum_{k=-\infty}^{\infty} x(k)\, y(n-k). \tag{A.1}$$

By a simple change of summation index, we see that convolution is commutative, that is,

$$x \otimes y = \sum_{k=-\infty}^{\infty} y(k)\, x(n-k) = y \otimes x.$$

Similarly,

$$(w \otimes x) \otimes y = \sum_{m=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} w(k)\, x(m-k)\, y(n-m) = \sum_{k=-\infty}^{\infty} w(k) \sum_{m=-\infty}^{\infty} x(m-k)\, y(n-m)$$

$$= \sum_{k=-\infty}^{\infty} w(k) \sum_{m=-\infty}^{\infty} x(m)\, y(n-k-m) = w \otimes (x \otimes y),$$

so that convolution is also associative. The zero vector is clearly the sequence δ such that

$$\delta(n) = 0 \qquad n \neq 0$$
$$\quad\;\; = 1 \qquad n = 0.$$

The inverse sequence for a sequence x is simply the inverse z transform of $1/X(z)$, where $X(z)$ is the z transform of x.

A.3 SCALAR MULTIPLICATION

We shall denote scalar multiplication by ${}^{(b)}x$. To begin to see what scalar multiplication means for convolutional vector spaces, let us assume that b is an integer. Conventionally we say that multiplication of a vector x by an integer is equivalent to adding the vector to itself b times. This is a direct consequence of the postulates because, for example,

$$2x = (1+1)x = x + x.$$

Thus in the case of convolution we say that ${}^{(b)}x$ corresponds to the convolution of x with itself b times. That is,

$${}^{(b)}x = {}^{(b-1)}x \otimes x,$$

where

$$^{(1)}x = x$$

$$^{(0)}x = \delta$$

and

$$^{(-1)}x \otimes x = \delta.$$

The set of all positive and negative integers does not constitute a field; however, the set of all rational numbers does. In this case, it is only slightly more difficult to interpret scalar multiplication. For example, suppose

$$y = {}^{(1/2)}x.$$

A reasonable interpretation in this case is given by the expression

$$x = {}^{(2)}y = y \otimes y.$$

In general for rational scalars we can give the interpretation

$$y = {}^{\left(\frac{a}{b}\right)}x \qquad {}^{(b)}y = {}^{(a)}x,$$

where a and b are integers.

An alternative interpretation of the scalar multiplication results from consideration of the z transform. For example, the z transform of $^{(b)}x$, where b is an integer, is $[X(z)]^b$. That is, X(z) raised to the bth power. In the second case, we note that for

$$y = {}^{\left(\frac{a}{b}\right)}x,$$

we obtain

$$[Y(z)]^b = [X(z)]^a.$$

Alternatively, the function $[X(z)]^{a/b}$ is normally defined by

$$[X(z)]^{a/b} = \exp\left\{\frac{a}{b} \log [X(z)]\right\},$$

where $[X(z)]^{a/b}$ is clearly a multivalued function of z. In fact, we may define

$$[X(z)]^b = \exp\{b \log [X(z)]\} \tag{A. 2}$$

for b a real or complex number.

It is evident from Eq. A. 2 that the definition of scalar multiplication is intimately related to the proper definition of log [X(z)]. In order that $[X(z)]^a$ have a Laurent

expansion and thus be thought of as a z transform, we require that $[X(z)]^a$ be single-valued. This can be accomplished through the concept of the Riemann surface.[22] Under the assumption that $[X(z)]^a$ is uniquely defined, the verification of the conditions regarding scalar multiplication involves only straightforward manipulations of powers of the z transform.

Acknowledgment

This report represents only a part of an extremely enjoyable and stimulating educational experience. For this experience, I am indebted to many people.

In particular, I wish to thank Professor Alan V. Oppenheim. I was fortunate to have been able to work closely with him during the development of the present research, and our association has been one of the highlights of my experience at the Massachusetts Institute of Technology. I am also grateful to Dr. T. G. Stockham, Jr., who assisted in the supervision of my thesis, but, more importantly, contributed his enthusiastic support when it was needed most. Thanks are also due to Professor Amar G. Bose, who also assisted in the supervision of my thesis, and whose comments and personal example have shown me the value of clarity of thought and expression.

I would like to express my appreciation to Professor Y. W. Lee and Professor Henry J. Zimmermann, of the Research Laboratory of Electronics, M. I. T., for their generous support of my thesis research. I would also like to thank Dr. I. L. Lebow of Lincoln Laboratory, M. I. T., for making it possible for me to use the TX-2 computer.

References

Homomorphic Systems

1. A. V. Oppenheim, "Superposition in a Class of Nonlinear Systems," Technical Report 432, Research Laboratory of Electronics, M. I. T., Cambridge, Mass., March 31, 1965.
2. A. V. Oppenheim, "Optimum Homomorphic Filters," Quarterly Progress Report No. 77, Research Laboratory of Electronics, M. I. T., Cambridge, Mass., April 15, 1965, pp. 248-260.
3. A. V. Oppenheim, "Nonlinear Filtering of Convolved Signals," Quarterly Progress Report No. 80, Research Laboratory of Electronics, M. I. T., January 15, 1966, pp. 168-175.

Deconvolution

4. T. G. Kincaid, "Time Domain Analysis of Impulse Response Trains," Technical Report 445, Research Laboratory of Electronics, M. I. T., Cambridge, Mass., May 31, 1967.
5. James R. Reeder and S. S. Wolff, "Fourier Techniques for the Resolution of Overlapping Waveforms," Proceedings of the Third Annual Allerton Conference on Circuit and System Theory, University of Illinois, Urbana, Illinois, 1965, pp. 310-319.
6. E. A. Robinson, "Predictive Decomposition of Time Series with Application to Seismic Exploration," Geophys. Vol. XXXII, No. 3, pp. 418-484, June 1967.
7. M. M. Sondhi, "An Adaptive Echo Canceller," Bell System Tech. J., Vol. XLVI, pp. 497-511, March 1967.
8. Tzay Y. Young, "Epoch Detection – A Method for Resolving Overlapping Signals," Bell System Tech. J., Vol. XLIV, pp. 401-426, March 1965.
9. R. W. Schafer, "Echo Removal by Generalized Linear Filtering," 1967 NEREM Record, Boston, 1967.

The Cepstrum

10. B. P. Bogert, M. J. R. Healy, and J. W. Tukey, "The Quefrency Alanysis of Time Series for Echoes: Cepstrum, Pseudo-Autocovariance, Cross-Cepstrum and Saphe Cracking," Proceedings of the Symposium on Time Series Analysis, edited by M. Rosenblatt (John Wiley and Sons, Inc., New York, 1963), Chap. 15, pp. 209-243.
11. B. P. Bogert and J. F. Ossanna, "The Heuristics of Cepstrum Analysis of a Stationary Complex Echoed Gaussian Signal in Stationary Gaussian Noise," IEEE Transactions on Information Theory, Vol. IT-12, No. 3, pp. 273-380, July 1966.
12. A. M. Noll, "Short-Time Spectrum and 'Cepstrum' Techniques for Vocal-Pitch Detection," J. Acoust. Soc. Am. 36, 296-302 (February 1964).
13. A. M. Noll, "Cepstrum Pitch Determination," J. Acoust. Soc. Am. 41, 293-309 (February 1967).

Fast Fourier Transform

14. J. W. Cooley and J. W. Tukey, "An Algorithm for the Machine Calculation of Complex Fourier Series," Math. Computation, Vol. 19, No. 90, pp. 297-301, April 1965.

15. W. M. Gentleman and G. Sande, "Fast Fourier Transforms – for Fun and Profit," 1966 Fall Joint Computer Conference, AFIPS Proc., Vol. 29, pp. 563-578 (Spartan Books, Washington, D.C., 1966).

16. "Special Issue on Fast Fourier Transform," IEEE Trans. on Audio and Electroacoustics, Vol. AU-15, June 1967.

17. T. G. Stockham, "High-Speed Convolution and Correlation," 1966 Spring Joint Computer Conference, AFIPS Proc., Vol. 28, pp. 229-233 (Spartan Books, Washington, D.C., 1966).

Speech Analysis

18. J. L. Flanagan, Speech Analysis Synthesis and Perception (Academic Press, Inc., New York, 1965).

19. A. V. Oppenheim and R. W. Schafer, "Homomorphic Analysis of Speech," IEEE Trans. on Audio and Electroacoustics, Vol. AU-16, pp. 221-226, June 1968.

20. A. V. Oppenheim, "A Speech Analysis-Synthesis System Based on Homomorphic Filtering," J. Acoust. Soc. Am. 45, 458-465 (February 1969).

Linear Systems

21. H. S. Carslaw, Introduction to the Theory of Fourier's Series and Integrals (Dover Publications, Inc., New York, 3rd edition, 1952).

22. R. V. Churchill, Complex Variables and Applications (McGraw-Hill Book Co., Inc., New York, 1960).

23. H. Freeman, Discrete-Time Systems (John Wiley and Sons, Inc., New York, 1965).

24. E. A. Guillemin, Theory of Linear Physical Systems (John Wiley and Sons, Inc., New York, 1963).

25. P. R. Halmos, Finite-Dimensional Vector Spaces (D. Van Nostrand Co., Inc., Princeton, N. J., 1958).

26. E. I. Jury, Theory and Application of the z-Transform Method (John Wiley and Sons, Inc., New York, 1964).

27. W. M. Siebert, "Notes for M.I.T. Course 6.05," 1967 (unpublished).

UNCLASSIFIED

Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author) Research Laboratory of Electronics Massachusetts Institute of Technology Cambridge, Massachusetts 02139	2a. REPORT SECURITY CLASSIFICATION Unclassified 2b. GROUP

3. REPORT TITLE

Echo Removal by Discrete Generalized Linear Filtering

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)

Technical Report

5. AUTHOR(S) (First name, middle initial, last name)

Ronald W. Schafer

6. REPORT DATE February 28, 1969	7a. TOTAL NO. OF PAGES 136	7b. NO. OF REFS 27
8a. CONTRACT OR GRANT NO. DA 28-043-AMC-02536(E) b. PROJECT NO. 200-145) 1-B31F c. d.	9a. ORIGINATOR'S REPORT NUMBER(S) Technical Report 466 9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report) None	

10. DISTRIBUTION STATEMENT

Distribution of this report is unlimited.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY Joint Services Electronics Program thru USAECOM, Fort Monmouth, N.J.

13. ABSTRACT

A new approach to separating convolved signals, referred to as homomorphic deconvolution, is presented. The class of systems considered in this report is a member of a larger class called homomorphic systems, which are characterized by a generalized principle of superposition that is analogous to the principle of superposition for linear systems.

A detailed analysis based on the z-transform is given for discrete-time systems of this class. The realization of such systems using a digital computer is also discussed in detail. Such computational realizations are made possible through the application of high-speed Fourier analysis techniques.

As a particular example, the method is applied to the separation of the components of a convolution in which one of the components is an impulse train. This class of signals is representative of many interesting signal-analysis and signal-processing problems such as speech analysis and echo removal and detection. It is shown that homomorphic deconvolution is a useful approach to either removal or detection of echoes.

DD FORM 1 NOV 65 1473 (PAGE 1)

UNCLASSIFIED

Security Classification

UNCLASSIFIED
Security Classification

14 KEY WORDS	LINK A ROLE	LINK A WT	LINK B ROLE	LINK B WT	LINK C ROLE	LINK C WT
homomorphic systems						
deconvolution						
echo removal						
speech analysis						
complex cepstrum						
Fast Fourier Transform						

UNCLASSIFIED
Security Classification