DIGITAL FILTERS AND APPLICATIONS TO SEISMIC DETECTION AND DISCRIMINATION

by

Jon F. Claerbout

S.B. Massachusetts Institute of Technology
(1960)

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
February, 1963

Signature of Author _____
Department of Geology & Geophysics
January 7, 1963

Certified by _____
Thesis Supervisor

Accepted by _____
Chairman, Departmental Committee on Graduate Students

ABSTRACT

Title: Digital Filters and Applications to Seismic Detection and Discrimination

Author: Jon F. Claerbout

Submitted to the Department of Geology and Geophysics on January 14, 1963 in partial fulfillment of the requirements for the degree of Master of Science at the Massachusetts Institute of Technology

The first part of this thesis is concerned with the mathematics of filtering in discrete time. Filters are defined for the purposes of 1) condensing waveforms into impulsive functions 2) wave shaping 3) noise suppression 4) signal detection according to the criterion of maximum signal-to-noise output at an instant and 5) the same over an interval. The behavior of the complex Fourier transforms of some of these filters is considered and connection is made with the theory of orthogonal polynomials. This leads to the possibility of a feed back representation of these filters.

In the second part, computational experiments are described in which digital filters are applied to seismic body waves to 1) try to determine whether the first arrival is up or down on a seismogram corrupted with microseismic noise, 2) increase signal-to-noise ratio on seismograms where noise has almost obliterated signal 3) assign polarity to each of two seismic first motion wavelets so they can be termed "same" or "opposite," 4) remove spectrum of seismometer from data, 5) investigate the time varying spectral structure of underground nuclear shot seismograms.

Thesis Supervisor: Stephen M. Simpson, Jr.

Title: Associate Professor of Geophysics

TABLE OF CONTENTS

THEORY

APPLICATIONS

INTRODUCTION

Although time is a continuous parameter it often happens that observations are made at discrete time intervals. Even when continuous observations are made, it is often desirable to digitize them for computer processing. This is strong reason to do some mathematics in discrete time. An even stronger reason as we will see is that things which are conceptually quite hard in continuous time have analogues in discrete time which are easier to understand.

Fortunately, in discrete time many general principles can be observed with wavelets of very short time duration. This enables us to consider some very simple examples before launching off into the general theory. These simple examples, however, will not forshadow the way in which we will connect the theory of least squares filters with the general theory of orthogonal polynomials.

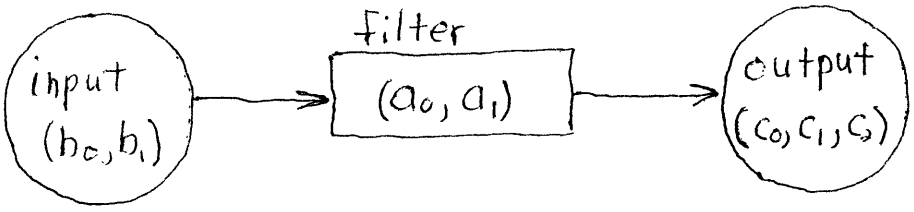
We denote time functions b_t with a subscript as the time parameter. When the time function has finite time duration, we may denote it as $\vec{b}$ or $(b_0, b_1, \ldots, b_n)$. Any time function which has finite energy is called a wavelet. The memory functions of filters, too, are sometimes called wavelets.

I. Introductory Examples

We introduce the main topics by means of some examples. One is given in discrete time an input series (b_0, b_1) of length (time duration) two, a filter (a_0, a_1) with an impulse response of length two, and the output resulting from convolution to be (c_0, c_1, c_2) of length necessarily three. The c is determined from the a and the b by convolution as is the usual procedure for linear filters, i.e.

$$\vec{c} = \begin{cases} c_0 = a_0 b_0 \\ c_1 = a_0 b_1 + a_1 b_0 \\ c_2 = a_1 b_1 \end{cases}$$

or



1) Spiking filter

To design a spiking filter one would choose (a_0, a_1) so that c comes out as closely as possible to a spike, i.e. either (1,0,0) or (0,1,0) or (0,0,1).

2) Wave shaper

To design a wave shaping filter one would choose (a_0, a_1) so that c comes out as closely as possible to some prescribed waveform (d_0, d_1, d_2).

3) A matched filter

To design a matched filter one would choose (a_0, a_1) so that c_1 comes out as large as possible while making the unit energy constraint ($a_0^2 + a_1^2 = 1$) on the filter. In this problem one doesn't care what c_0 and c_2 turn out to be.

4) Maximum energy sum filter

To design a maximum energy sum filter one would choose (a_0, a_1) so that the energy output ($c_0^2 + c_1^2 + c_2^2$) comes out as large as possible while making the unit energy constraint ($a_0^2 + a_1^2 = 1$) on the filter.

A quick sketch of the solutions to these problems is as follows: Since the spiking filter is a special case of the wave shaper it will be sufficient to work out the solution for the wave shaper. Requiring $\vec{c}$ to be as close as possible to $\vec{d}$ is equivalent to minimizing the squared distance between them

$$|\vec{c} - \vec{d}|^2 = (c_0 - d_0)^2 + (c_1 - d_1)^2 + (c_2 - d_2)^2$$
$$= (a_0 b_0 - d_0)^2 + (a_0 b_1 + a_1 b_0 - d_1)^2 + (a_1 b_1 - d_2)^2$$

Setting the partial derivatives with respect to a_0 and a_1 equal to zero we get the simultaneous set for a.

$$\begin{aligned}(b_0^2+b_1^2)\,a_0 \;+(b_1b_0)\,a_1 &= b_0d_0+b_1d_1\\ (b_1b_0)\,a_0 \;+(b_0^2+b_1^2)\,a_1 &= b_0d_1+b_1d_2\end{aligned}$$

We mention the particular case $\vec{d}=(1,0,0)$ called the zero delay spiking filter. The solution of the simultaneous set is

$$(a_0,a_1) \sim \left(b_0^2+b_1^2,\; -b_0b_1\right)$$

Recalling that subscripts are the time variable we now consider the Fourier transform of the solution

$$F_a(\omega) = a_0+a_1e^{-i\omega} = \left(b_0^2+b_1^2\right) - b_0b_1e^{-i\omega}$$

The only zero of this complex function is in the upper half of the complex frequency plane, a fact which will be shown true for all zero delay spiking filters. This has considerable importance in feedback systems and in some other connections to be discussed.

The solution to the <u>matched filter</u> problem posed in 3) above is most easily done by means of Lagrange multipliers. We wish to maximize c_1 under the constraint $a_0^2+a_1^2=1$. Lagrange's method is then to maximize

$$\max\left[c_1-\lambda\left(a_0^2+a_1^2\right)\right]$$

$$\max\left[a_0b_1+a_1b_0-\lambda\left(a_0^2+a_1^2\right)\right]$$

Setting the derivatives with respect to a_0 and a_1 equal to zero, one gets

$$(a_0, a_1) = (b_1, b_0)/2\lambda = \frac{(b_1, b_0)}{\sqrt{b_0^2 + b_1^2}}$$

Thus the filter (a_0, a_1) is simply the signal input time-reversed and multiplied by a scale factor.

The solution to the maximum energy sum problem 4) is somewhat like the matched filter. Again one uses Lagranges method and maximizes

$$c_0^2 + c_1^2 + c_2^2 - \lambda(a_0^2 + a_1^2)$$

by setting derivatives with respect to the components of a equal zero. This results in the equations

$$\left\{\begin{bmatrix} b_0^2 + b_1^2 & b_0 b_1 \\ b_0 b_1 & b_0^2 + b_1^2 \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right\} \begin{Bmatrix} a_0 \\ a_1 \end{Bmatrix} = \begin{Bmatrix} 0 \\ 0 \end{Bmatrix}$$

which is the standard eigenvector ($\vec{a}$), eigenvalue (λ) problem. The two solutions to this problem are

$$\vec{a}_1 = \frac{(1, 1)}{\sqrt{2}} \qquad \text{and} \qquad \vec{a}_2 = \frac{(1, -1)}{\sqrt{2}}$$

It is notable that the fourier transform of these functions

$$F_{1a}(\omega) \sim 1 + e^{-i\omega} \qquad F_{2a} \sim 1 - e^{-i\omega}$$

have zeros on the real frequency axis. This will also happen with longer filters.

II. Spiking Filters

A. Normal Equations

In the first introductory example we considered the problem of building a two term filter which would condense

a two term input into a spike function. Now we would like to build an m+1 term filter to condense an n+1 term input into a spike.

A data wavelet is given by $\vec{b}=(b_0,b_1,\ldots,b_n)$. We plan to construct a filter $\vec{a} = (a_0,a_1,\ldots,a_m)$. Filtering is defined in this way: When data $\vec{b}$ goes into a filter $\vec{a}$, an output wavelet $\vec{c}$ is produced according to the following matrix multiplication.

$$\begin{bmatrix} c_0 \\ c_1 \\ c_2 \\ \vdots \\ c_{m+n} \end{bmatrix} = \begin{bmatrix} b_0 & \text{zeros} \\ b_1 & b_0 \\ b_2 & b_1 & \ddots & b_0 \\ \vdots & & & b_1 \\ b_n & & & \vdots \\ \text{zeros} & & & b_n \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_m \end{bmatrix} \qquad \text{(II-1)}$$

This operation is often called <u>complete transient convolution</u>. This is more loosely written as

$$c_i = \sum_j b_{i-j}\, a_j \qquad \text{(II-2)}$$

Here a small amount of confusion can arise about the limits of the summation because negative subscripts may appear within the summation. What is meant is that one should consider the terms "off-the-ends" of the wavelets to be zero. With this consideration we might write the limits of the summation as minus to plus infinity. The artifice of using infinite limits on the sums turns out to avoid some needlessly cumbersome notation.

Now we introduce another wavelet $\vec{d}$ which will have the same number of components as $\vec{c}$. We call $\vec{d}$ the desired output of the filter. We saw that $\vec{c}$ is the actual output. The actual output $\vec{c}$ was seen to be a function of the input $\vec{b}$ and the filter $\vec{a}$. The problem now is to determine a so

that $\vec{c}$ and $\vec{d}$ are very much alike. Specifically we will choose $\vec{a}$ so that the difference vector $|\vec{c}-\vec{d}|$ has minimum length squared (in n+m+1 dimensional space). In other words we are minimizing

$$\sum_{i=0}^{m+n} (c_i - d_i)^2 \qquad \text{(II-3)}$$

by varying the components of $\vec{a}$. Inserting the expression for $\vec{c}$ in terms of $\vec{a}$ and $\vec{b}$ we get

$$\sum_{i=0}^{m+n} \left(\sum_j b_{i-j} a_j - d_i \right)^2 \qquad \text{(II-4)}$$

This function of m+1 variables will be minimized if its partial derivative with respect to each of $(a_0, a_1, \ldots, a_m)$ equals zero. Setting derivatives with respect to a_ℓ equal zero we get an expression for m+1 equations

$$0 = \sum 2\, b_{i-\ell} \left(\sum_j b_{i-j} a_j - d_i \right) \qquad \text{(II-5)}$$

where one equation is implied for each value of ℓ ($0 \le \ell \le m$).

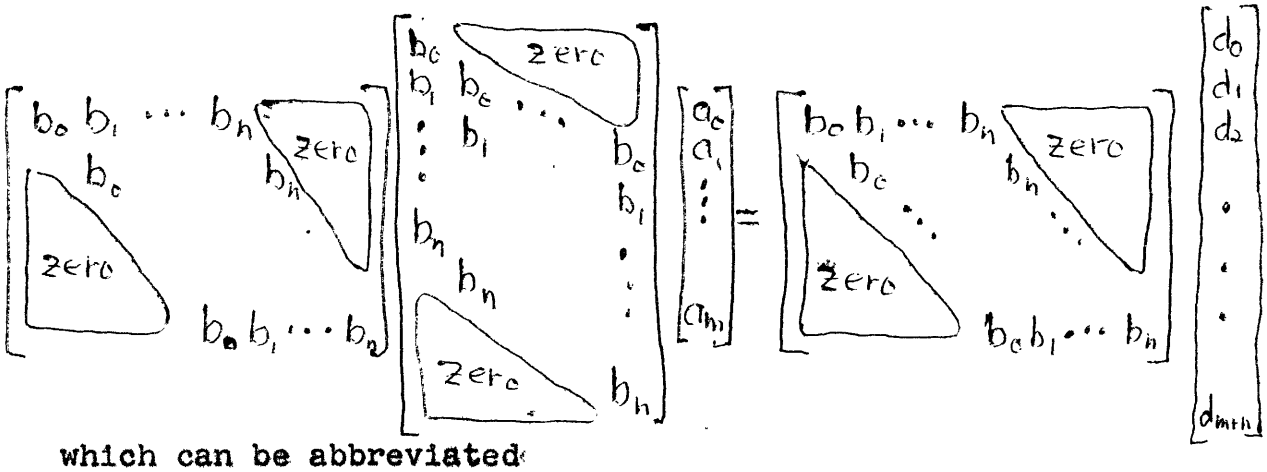
These are called normal equations because they say that the error vector, the quantity in brackets, will be normal or perpendicular to the space spanned by the vector set $b_{i-\ell}$ (column vectors in the matrix of equation II-1).

We bring the equations into standard form by bringing the homogeneous part (the part depending on $\vec{a}$) to the left side and the inhomogeneous part to the right

$$\sum_{i=0}^{m+n} b_{i-\ell} \left(\sum_j b_{i-j} a_j \right) = \sum_{i=0}^{m+n} b_{i-\ell}\, d_i \qquad \text{(II-6)}$$

In matrix form the normal equations become

(II-7)



which can be abbreviated

$$B^T(Ba) = B^T d \qquad \text{(II-8)}$$

and which is identically equal to

$$(B^T B)a = B^T d \qquad \text{(II-9)}$$

The matrix $B^T B$ can be written as

$$B^T B = \begin{bmatrix} r_0 & r_1 & r_2 & \cdots & r_m \\ r_1 & r_0 & r_1 & & \\ r_2 & & r_0 & & \\ \vdots & & & & \\ r_m & & & & r_0 \end{bmatrix}$$

where

$$r_j = \begin{cases} \sum_{i=0}^{n-j} b_i b_{i+j} & \text{if } j \leq n \\ 0 & \text{if } j > n \end{cases}$$

This r_j is called the unnormalized transient autocorrelation of b.

We list three special cases of these equations.

1. Zero delay inverse filter - This is when $\vec{d} = (1,0,0,\ldots,0)$.

2. Spiking filter - This is when the impulse is chosen any where in $\vec{d}$. It has been frequently observed in practice that putting the impulse near the middle of $\vec{d}$ results in an actual output $\vec{c}$ which resembles $\vec{d}$ more closely than if $\vec{d}$ had been chosen as in the zero delay case.

3. Waveshaping filter - This is when $\vec{d}$ is not chosen to be an impulse at all, but is chosen to be some arbitrary wavelet. The filter a then tries to convert the wavelet $\vec{b}$ into the wavelet $\vec{d}$.

It is worth noticing that the homogeneous part of the normal equations (II-9) depends only upon the autocorrelation of the input $\vec{b}$ and not on $\vec{b}$ itself. If the desired output of the filter is an impulse with no delay ($\vec{d} = (1,0,0,\ldots,0)$) then the inhomogeneous part becomes the column vector

$$B^T d = \begin{bmatrix} b_0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

Now in this case we see that the waveform $\vec{b}$ does not enter the inhomogeneous part either, except for the magnitude of b_0. Inspection of the normal equations shows that this magnitude will not affect the waveform of the filter a except as a scale factor.

Thus the normal equations in this special case (zero delay inverse filter) depend upon the signal waveform, but only through its autocorrelation. Since autocorrelations contain no phase information it would be a curious point as to what the phase spectrum will be of the solution $\vec{a}$. We

will study this later and come to the curious conclusion that the phase spectrum is such that as much as possible of the energy in the waveform a is cramped up as close as possible to a_0. This is called the property of minimum phase delay of the waveform $\vec{a}$.

To fix ideas we now give an example of the determination of a zero delay inverse wavelet. Suppose that the signal we are dealing with is the waveform $\vec{b}=(2,1)$. We want to design a three-term filter $\vec{a}=(a_0,a_1,a_2)$. The desired output must then be n+m+1 = 1+2+1 terms long and is $\vec{d}$ = (1,0,0,0). From (II-10) $r_0 = 5$, $r_1 = 2$, $r_2 = 0$. The normal equations are

$$\begin{bmatrix} 5 & 2 & 0 \\ 2 & 5 & 2 \\ 0 & 2 & 5 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$$

and the solution is $\vec{a}$ = (42,-20,8)/85. To see how good the filter is we compare:

actual output $\vec{c}$ = (84,2,-4,8)/85
desired output $\vec{d}$ = (1,0,0,0)

B. Minimum Phase

Discussions of minimum phase in the literature are mostly in terms of continuous time. Here we wish to develop its properties from the point of view of digital filters which are not so well known. We begin by considering an autocorrelation function of the type of equation (II-11) where

$$\vec{r} = (r_{-n}, r_{-n+1}, \ldots, r_{-1}, r_0, r_1, \ldots r_{n-1}, r_n) \qquad \text{(II-13)}$$

We wonder what functions $\vec{b}$ might have this autocorrelation. After we have found the class of functions $\vec{b}$ that have this autocorrelation we can enquire which one has its energy as

close as possible to b_o and is, therefore, the minimum phase delay wavelet. One thing which we know to begin with is that more than one wavelet $\vec{b}$ may have autocorrelation $\vec{r}$ (for example; the time reversed waveform, the negative waveform, and the time reverse of the negative waveform).

We begin by spectral considerations. Let F denote Fourier transform. It is commonly known that the energy density spectrum of the wavelet may be expressed in two equal ways:

$$F_r(\omega) = F_b(\omega)\,\overline{F_b(\omega)} \tag{II-14}$$

Thus the problem is to factor $F_r(\omega)$ into $F_b(\omega)$ and $\overline{F_b(\omega)}$. Then we can simply take the inverse transform of $F_b(\omega)$ to get the waveform $\vec{b}$. The Fourier transform of $\vec{r}$ is simply

$$F_r(\omega) = r_0 + 2\sum_{k=1}^{n} r_k \cos \omega k \tag{II-15}$$

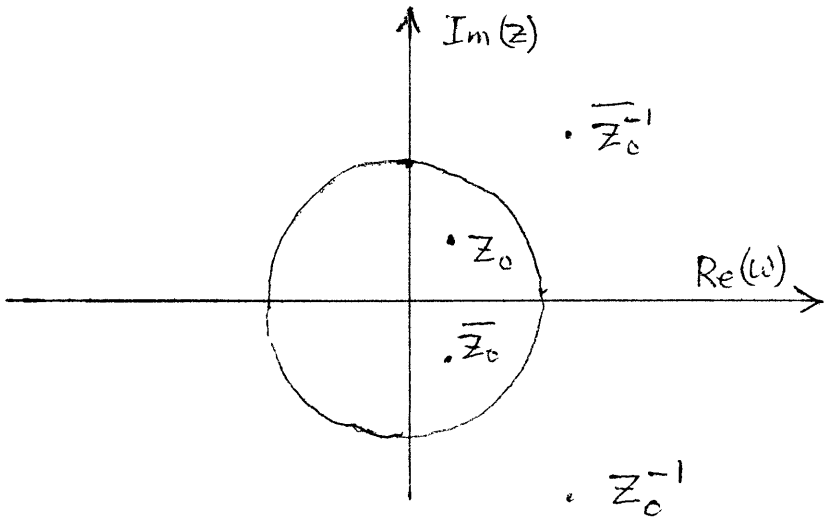
and letting $z = e^{-i\omega}$ we get

$$F_r(Z) = Z^{-n}r_n + Z^{-n+1}r_{n-1} + \cdots + Z^{-1}r_1 + r_0 + Zr_1 + \cdots Z^n r_n$$

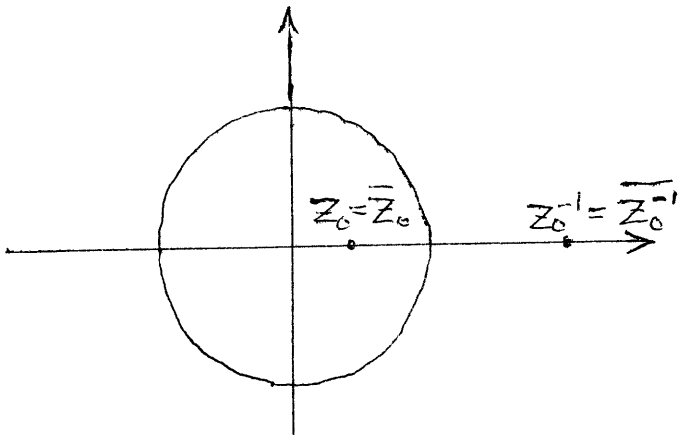
$$= Z^{-n}\left(r_n + r_{n-1}Z + \cdots + r_0 Z^n + \cdots r_n Z^{2n}\right) \tag{II-16}$$

We notice that the spectrum has been represented as a polynomial in z. The usual procedure in factoring a polynomial is to find its zeros. Since $r_k = r_{-k}$, we notice that $F_r(Z)$ is unchanged if we replace z by z^{-1}. Thus if $F_r(Z_o)$ is zero then $F_r(1/Z_o)$ will also be zero. Thus for every zero z_o, z_o^{-1} is also a zero. Also since the coefficients of the polynomial are real the zeros are either real or they occur in conjugate pairs. Thus if z_o is a zero then $\overline{Z_o}$ is a zero. Most of the zeros will probably occur then in groups of

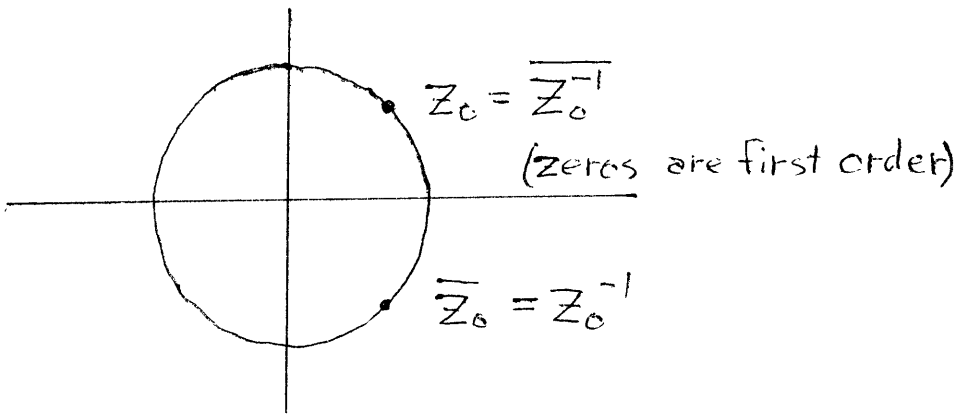
four such as



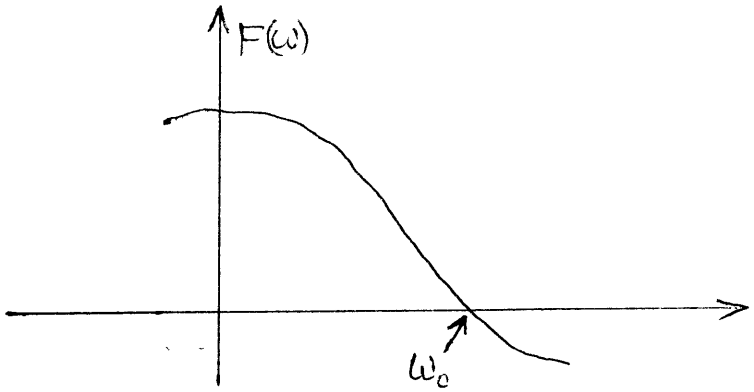
Some of the zeros may occur in groups of two such as



One might wonder about the case



where there are two single zeros on the unit circle. It turns out that this can't happen. What we are plotting here is possible locations of zeros of energy density spectra like equation (II-16). When z_o is on the unit circle ω_o is real by the relation $Z = e^{-i\omega}$. Thus we are talking about the spectrum at some real frequency. A function like the following



which has a single zero at ω_o is not an energy density spectrum because it is not positive for all frequencies. More generally, energy density spectra cannot have zeros of odd multiplicity on the unit circle in the z-plane.

We now know that for every zero $Z_o = e^{-i\omega_o}$ of the energy spectral polynomial that $\bar{Z}_o^{-1} = e^{i\omega_o}$ is another zero. After we factor the spectral polynomial we will be able to write the spectrum as

$$F_r(z) = z^{-h} r_h \left[(z-z_1)(z-z_2)\cdots\right]\left[(z-\bar{z}_1^{-1})(z-\bar{z}_2^{-1})\cdots\right] \qquad \text{(II-18)}$$

or in terms of ω

$$F_r(\omega) = e^{ih\omega} r_h \underbrace{\left[(e^{-i\omega}-e^{-i\omega_1})(\quad)\cdots\right]}_{A(\omega)}\underbrace{\left[(e^{-i\omega}-e^{-i\omega_1})(\quad)\cdots\right]}_{B(\omega)} \qquad \text{(II-19)}$$

Now if we show $\overline{A(\omega)} \sim B(\omega)$ then we have factored the spectrum $F_r(\omega)$ into the desired conjugate parts $F_b(\omega)\overline{F_b(\omega)}$.

But both are polynomials in $e^{-i\omega}$ of order n and both $\overline{A(\omega)}$ and $B(\omega)$ have the same zeros. Thus they must be the same function except for a constant multiplicative factor. This can be absorbed from the factor $r_n e^{-in\omega}$. This is called factoring the spectrum.

We notice that the factorization could have been done in many ways depending on which of the pair of zeros is put into $F_b(\omega)$ (the other one then going into $\overline{F_b(\omega)}$). Normally, there would be 2^n different ways of doing this, the exception being the degenerate case when zeros occur with multiplicity greater than unity. Then there would be fewer than 2^n wavelets with the same energy density.

One of these possible factorizations is of particular signifigance. The factoring is done so that all of the zeros which are outside[1] of the unit circle are put into $F_b(\omega)$ and the opposite member of each pair which is inside the circle then goes into $\overline{F_b(\omega)}$. In this case the wavelet must be real because each root is either real or it occurs with its complex conjugate.

Combining all complex roots z with their complex conjugates $(\alpha_k + i\beta_k)(\alpha_k - i\beta_k)$ we write for the wavelet's transform

$$F_b(z) = \text{const.}\left[\underbrace{(z-z_0)(z-z_1)}_{\text{on real axis}}\cdots\underbrace{(z^2-2\alpha_k+(\alpha^2+\beta^2))(\quad)}_{\text{other zeros outside unit circle}}\cdots\right]$$

Taking the inverse transform and letting "*" denote convolution

1 The case with zeros exactly on the unit circle corresponds to a spectrum which is exactly zero for some real ω. In any physical case one can usually perturb the spectrum slightly to avoid this difficulty.

$$b_t = \text{const}\, F^{-1}[(\quad)(\quad)\cdots(\qquad)(\qquad)\cdots]$$

(II-19.1)

$$= (-1)^n \text{const}[(z_{0,}-1)*(z_{1,}-1)*\cdots*(\alpha_k^2+\beta_k^2,-\alpha_k 2_g 1)*(\qquad)\cdots$$

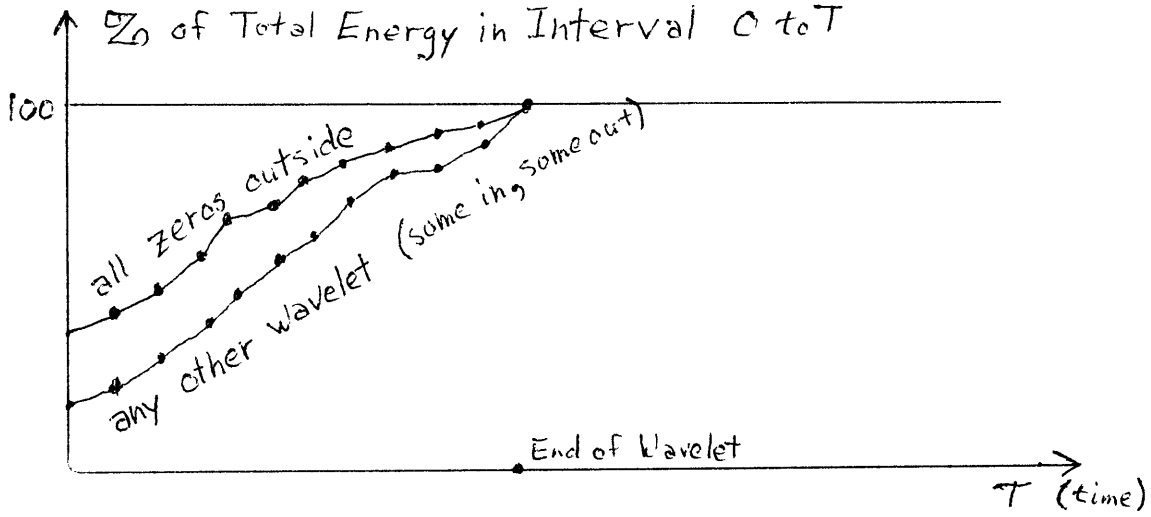
(II-19.2)

Thus we have a string of convolutions of many wavelets each of either 2 or 3 instants duration. Since all of the roots were chosen outside the unit circle we have

$$|Z_k| > 1 \qquad \text{and} \qquad \alpha_k^2+\beta_k^2 > 1$$

This means that in each of the wavelets the first term is larger in absolute value than the last. Thus in the convolution of all terms, the energy will be compacted toward the beginning. If any one of the zeros had been chosen instead, from inside the circle, then the energy would be spread further out on the time axis.

We will now prove that



the summed energy from 0 to any time t for the minimum phase wavelet is greater or equal to that of anyother wavelet with the same spectrum.

Consider a two term wavelet (b,s) "bigger," "smaller," with its zero outside the circle. Convolve it into an arbitrary wavelet $\vec{p} = (p_0, p_1, \ldots, p_k)$. The result is

$$\vec{p}_{out} = (b,s) * \vec{p} = (p_{out_1}, p_{out_2}, \ldots, p_{out_{k+1}})$$
$$= (bp_0, bp_1+sp_0, bp_2+sp_1, \ldots, sp_k)$$

If instead we had chosen the reversed wavelet (s,b) with its zero inside the circle, we would get

$$\vec{p}_{in} = (s,b) * \vec{p} = (sp_0, sp_1+bp_0, sp_2+bp_1, \ldots bp_k)$$

Then we consider the partial energy from time = 0 up to time = T and tabulate the difference between $\vec{p}_{in}$ and $\vec{p}_{out}$

	$\sum_{t=0}^{T} (p_{out})^2$	$- \sum_{t=0}^{T} (p_{in})^2$	$=$
$T=0$	$(bp_0)^2$	$-(sp_0)^2$	$= (b^2-s^2)p_0^2$
$T=1$	$(bp_0)^2 + (bp_1+sp_0)^2$	$-(sp_0)^2 - (sp_1+bp_0)^2$	$= (b^2-s^2)p_1^2$
	$(bp_1)^2$	$-(sp_1)^2$	$=$
$T=2$	$(bp_1)^2 + (bp_2+sp_1)^2$	$-(sp_1)^2 - (sp_2+bp_1)^2$	$=$
	$(bp_2)^2$	$-(sp_1)^2$	$= (b^2-s^2)p_2^2$
		etc	
$T=k$	$(bp_k)^2$	$-(sp_k)^2$	$= (b^2-s^2)p_k^2$
$T=k+1$	$(bp_k)^2 + (sp_k)^2$	$-(sp_k)^2 - (bp_k)^2$	$= 0$

Thus we conclude that for any time T the wavelet $\vec{p}_{out}$ with the zero outside the unit circle contains $(b^2-s^2)p_t^2$ more energy in the interval $0 \leq t \leq T$ than the wavelet $\vec{p}_{in}$ with the zero inside. The exception is at the last lag when they have both put out the same total energy. It is not difficult to show that the above statements would still be true if components of vectors were complex and squaring were replaced with multiplying by conjugates.

To prove the minimum phase wavelet delays energy the least, one imagines that the convolution (II-19.2) had been done so that k zeros were outside the circle and n-k were inside. We have just shown that if one of the zeros from inside were replaced with an outside zero, that the new convolution would have less energy delay. This argument is repeated until all zeros are outside.

Finally, we show that <u>zero delay spiking wavelets determined by least squares will have all their zeros outside the unit circle</u>.

We recall the following from previous portions of this thesis:

1) The least-squares spiking wavelet is a wavelet $\vec{a}$ which when convolved with a given wavelet $\vec{b}$ tries to give an output equal in the least squares sense to $\vec{d} = (d_0,0,0,\ldots,0)$. Specifically, $\vec{a}$ is chosen to minimize

$$\sum_{i=0}^{n+m} \left((a*b)_i - d_i \right)^2 = \left(a_0 b_0 - d_0 \right) + \sum_{i=1}^{n+m} (a*b)_i^2$$

2) We recall that the choice of size of d_0 affects the solution vector $\vec{a}$ only as a scale factor. Thus d_0 could always be chosen so that $a_0 = 1$. We note that a scale factor has no effect on a per cent total energy graph.

3) We recall that if a zero of a wavelet is removed from inside the circle and replaced by the conjugate inverse

zero outside, that the modified wavelet has a per cent total energy curve which lies above that of the original wavelet. The per cent total energy curves may touch one another at points except for at time t=0 where the curve with fewest zeros inside the circle is definitely above.

We can view the normal equations as minimizing the energy in $\vec{a}$ convolve $\vec{b}$ after time t=0 subject to the constraint that the energy at t=0 be equal to $(a_o b_o)^2 = (b_o)^2$. That is, we could view the normal equations as minimizing the per cent energy after t=0. But this is the same as maximizing the per cent energy at t=0. But if the percentage energy at t=0 is to be maximized for the wavelet $\vec{a}$ convolve $\vec{b}$, then there must be as few as possible zeros inside the circle. This happens if $\vec{a}$ has none inside and hence is minimum phase.

C. Connection of Least Squares Inverse Filter with Orthogonal Polynomials

Given an energy density function

$$\Psi(z) = \cdots z^{-k} r_k + \cdots z^{-1} r_1 + r_0 + z r_1 + \cdots + z^k r_k + \cdots$$

one could take that function and use it as a weighting function to define a set of orthogonal polynomials. We choose the interval of orthogonality to be the unit circle in the z plane which corresponds to the real frequency axis from $-\pi$ to $+\pi$ in the ω plane. Thus we would construct a set of polynomials f_k

$$f_0 = \alpha_{00}$$
$$f_1 = \alpha_{10} + \alpha_{11} z$$
$$f_2 = \alpha_{20} + \alpha_{21} z + \alpha_{22} z^2$$

so that

$$\delta_{mn} = \frac{1}{2\pi} \int_{-\pi}^{+\pi} \Psi(\omega) f_m(\omega) \overline{f_n(\omega)} \, d\omega = \begin{cases} 0 & m \neq n \\ 1 & m = n \end{cases} \qquad (II\text{-}20)$$

on the real axis. Expressing the same thing with complex polynomials on the unit circle one gets

$$\delta_{mh} = \frac{i}{2\pi} \oint_{|z|=1} \Psi(z)\, f_m(z)\, \overline{f_h(z)}\, \frac{dz}{z} \qquad \text{(II-21)}$$

We illustrate the construction of these polynomials in such a way that it will be seen to be equivalent to the least squares normal equations. Consider the construction of f_2. Let $[f_n, f_m]$ denote the dot product defined by equation (II-20). The vector f_2 is of order two say

$$f_2 = c_0 + c_1 z + c_2 z^2$$

and must satisfy the orthogonality conditions

$$\begin{aligned} [f_2, f_0] &= 0 \\ [f_2, f_1] &= 0 \\ [f_2, f_2] &= 1 \end{aligned} \qquad \text{(II-22)}$$

Since f_2 perpendicular to any linear combination of f_0 and f_1, it is perpendicular to any polynomial of order less than 2. Thus the orthogonality relations could be written

$$\begin{aligned} [f_2, 1] &= 0 \\ [f_2, z] &= 0 \\ [f_2, z^2] &= \text{const} \end{aligned} \qquad \text{(II-23)}$$

This set of orthogonality requirements (II-23) can be written out in full as

$$\begin{aligned} [1,1]\, c_0 + [z,1]\, c_1 + [z^2,1]\, c_2 &= 0 \\ [1,z]\, c_0 + [z,z]\, c_1 + [z^2,z]\, c_2 &= 0 \\ [1,z^2]\, c_0 + [z,z^2]\, c_1 + [z^2,z^2]\, c_2 &= \text{const.} \end{aligned} \qquad \text{(II-24)}$$

We now examine the coefficients in this simultaneous set. Consider $\left[Z^n, Z^m\right]$

$$\left[Z^n, Z^m\right] = \frac{1}{2\pi}\int_{-\pi}^{+\pi}(r_0 + 2r_1\cos\omega + 2r_2\cos 2\omega + \cdots)e^{ih\omega}e^{-im\omega}d\omega$$

$$= \frac{1}{2\pi}\int_{-\pi}^{+\pi}(r_0 + 2r_1\cos\omega + 2r_2\cos 2\omega + \cdots)(\cos(h-m)\omega + i\sin(h-m)\omega)d\omega$$

$$= r_{h-m}$$

Hence the orthogonality relations (II-24) can be written

$$\begin{bmatrix} r_0 & r_1 & r_2 \\ r_1 & r_0 & r_1 \\ r_2 & r_1 & r_0 \end{bmatrix}\begin{bmatrix} c_2 \\ c_1 \\ c_0 \end{bmatrix} = \begin{bmatrix} \text{const.} \\ 0 \\ 0 \end{bmatrix} \qquad \text{(II-25)}$$

This is almost exactly the same as the normal equations (II-7) for the least squares inverse filter. The only differences are a scale factor in the inhomogeneous part and "time reversal" of the solution. But this will not affect the waveform $\vec{c}$ except by a scale factor and time reversal.

Thus we have shown the important result that following two problems are equivalent:

1) Find polynomials which are orthogonal on the unit circle with weight $\Psi_r(\omega)$

2) Find least squares zero delay spiking filters of different lengths for the spectrum $\Psi_r(\omega)$

This result is important because it allows us to apply many results in the classic field of orthogonal

polynomials to least squares filter theory.

One application is to use the recurrence relation between successive orthogonal polynomials to generate the filter of length n+1 from the filter of length n. This trick greatly facilitates computing the solution of the normal equations. The relationship for getting $f_{m+1}(Z)$ from $f_m(Z)$ is the recurrence relation (Geronimus 1960)

$$\alpha_{m,0} f_{m+1}(z) = \alpha_{m+1,0} f_m(z) + \alpha_{m+1,m+1} z^m f_m\left(\frac{1}{z}\right) \qquad \text{(II-26)}$$

where the two side conditions used to get $\alpha_{m+1,0}$ and $\alpha_{m+1,m+1}$ are first

$$\left(\alpha_{m,m} r_1 + \alpha_{m,m-1} r_2 + \cdots \alpha_{m,0} r_{m+1}\right) = \frac{\alpha_{m+1,m+1}}{\alpha_{m+1,0}\ \alpha_{m,0}} \qquad \text{(II-27)}$$

and second

$$\left(\alpha_{m+1,0}\right)^2 - \left(\alpha_{m,0}\right)^2 = \left(\alpha_{m+1,m+1}\right)^2 \qquad \text{(II-28)}$$

The choice of sign for the square root is immaterial as far as polynomial orthogonality is concerned. It is customary to choose it so that $\alpha_{m+1,m+1}$ the first term in the spiking wavelet is positive. The recurrence relation can be started off by choosing any value whatever for α_{00}. The result is just a scale factor in the inverse wavelet. From equation (II-20) it is evident that the recurrence formula can be started off at $\alpha_{0,0} = (r_0)^{-1/2}$.

These relations appear to have first been derived by Szego (1939).

Another readable account is Geronimus (1960). Levinson (1939) also derived similar, but not identical relations for the filter problem, although he does not mention any connection with orthogonal polynomials. Levinson's scheme is even more useful than the polynomial recurrence relations because it allows solving the normal equations for arbitrary inhomogeneous part.

Another valuable result of the connection of filter and polynomial theory is the following. All the zeros of all the polynomials generated by the recurrence relation above are known to lie inside the unit circle (Geronimus 1960). This means that the time reverse of the associated filter is minimum phase. Because of this we can invert the wavelet, i.e. take the inverse of its spectrum.

$$\frac{1}{a_m + a_{m-1} z + \cdots + a_0 z^m} = b_0 + b_1 z + b_2 z^2 + \cdots$$

Since the polynomial a(z) has no zeros inside the unit circle, the infinite series b(z) converges at least up to and including the unit circle. This means that the wavelet b_k has finite energy and is minimum phase. The wavelet $\vec{a}$ has a spectrum which is in a least squares sense* equal to $1/\Psi(\omega)$. The spectrum of the infinitely long wavelet $\vec{b}$ is exactly the inverse of the spectrum of $\vec{a}$. Hence we conclude that the spectrum of $\vec{b}$ is equal in a least squares sense to $\Psi(\omega)$. Thus we have found a way to compute in a least squares sense the minimum phase wavelet of a given autocorrelation function. Futhermore, the

*Least squares in the sense that $\int_{-\pi}^{\pi} |b(\omega)a(\omega) - 1|^2 d\omega$ is minimized where $b(\omega)$ has power spectrum $\Psi(\omega)$.

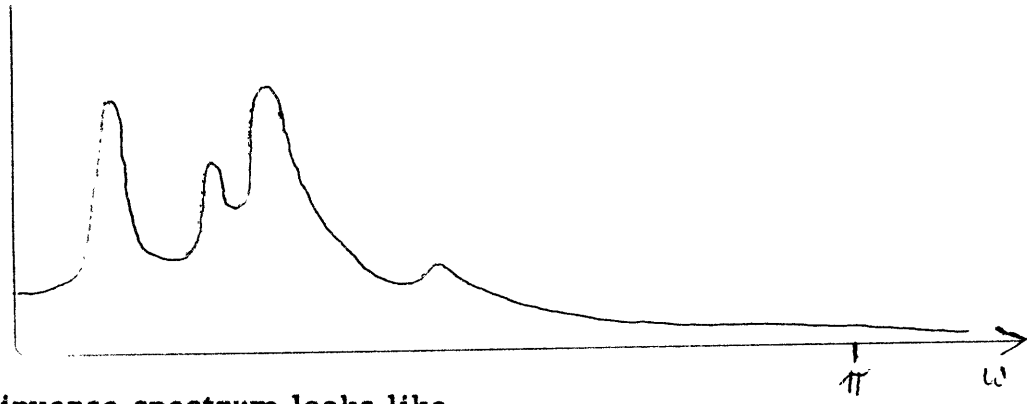
computation is quite easy because of the recurrence relation. It is the most efficient method known to the author who has computed 500 terms of the minimum phase wavelet in about a minute on an IBM 7090 computing machine.

D. A Comment on Autoregressive vs Moving Average Representations

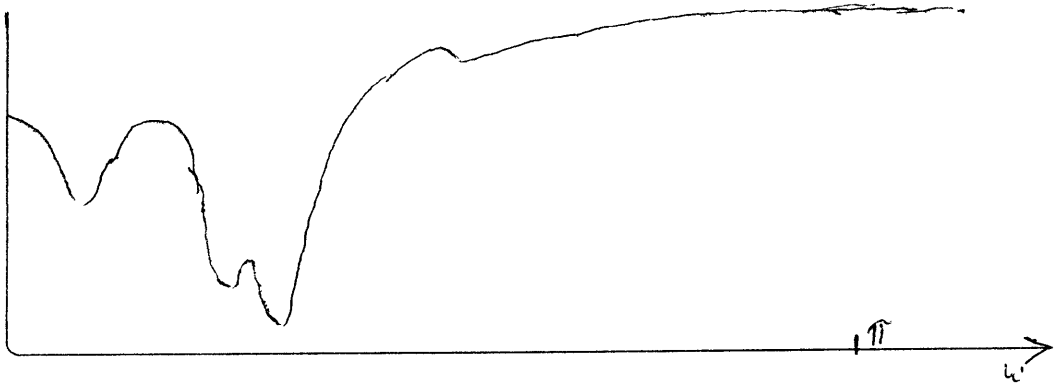
A question arises whether it is more efficient to characterize a stochastic process by the first n terms in its "autoregressive operator" or by the first n terms of its "moving average operator." What is meant by this is the following: Usually filtering is thought of in terms of convolving filter coefficients $\vec{b}$ with a data series. This might be called "moving weighted averages" or more commonly, "moving averages." This is equivalent to multiplying the Fourier transform of the data by that of the filter. Substituting $z = e^{-i\omega}$ it is equivalent to multiplying z-transform polynomials which convolves their coefficients. Filtering could be done in another way called "autoregression." Instead of multiplying the data polynomial by b(z) one divides it by the polynomial a(z). This is called "feedback" filtering for reasons which should be apparent to anyone who has ever divided polynomials by the method of synthetic division (see Lanczos 1956).

By "efficiency" we mean the following: suppose we want a filter to represent $\Psi(\omega)$ and it is easy for us to compute both $\vec{a}$ and $\vec{b}$ quite accurately; in fact, we wish to use many fewer terms than we can compute. Which characterizes $\Psi(\omega)$ more accurately for small p, the moving average approximation $b_0 + b_1 z + b_2 z^2 + \cdots b_p z^p$ or the

autoregression approximation $1/(a_0+a_1z+\ldots+a_pz^p)$? Whittle (in press) observes that the autoregressive coefficients seem more efficient and suggests that the reason is that for the series he deals with (economic), autoregression is a more realistic physical model. The author has also observed that the autoregressive coefficients seem more efficient in geophysical time series, but suggests a different reason. When we digitize continuous functions we usually digitize at a rate high enough to avoid appreciable frequency fold over. A typical spectrum looks like



The inverse spectrum looks like



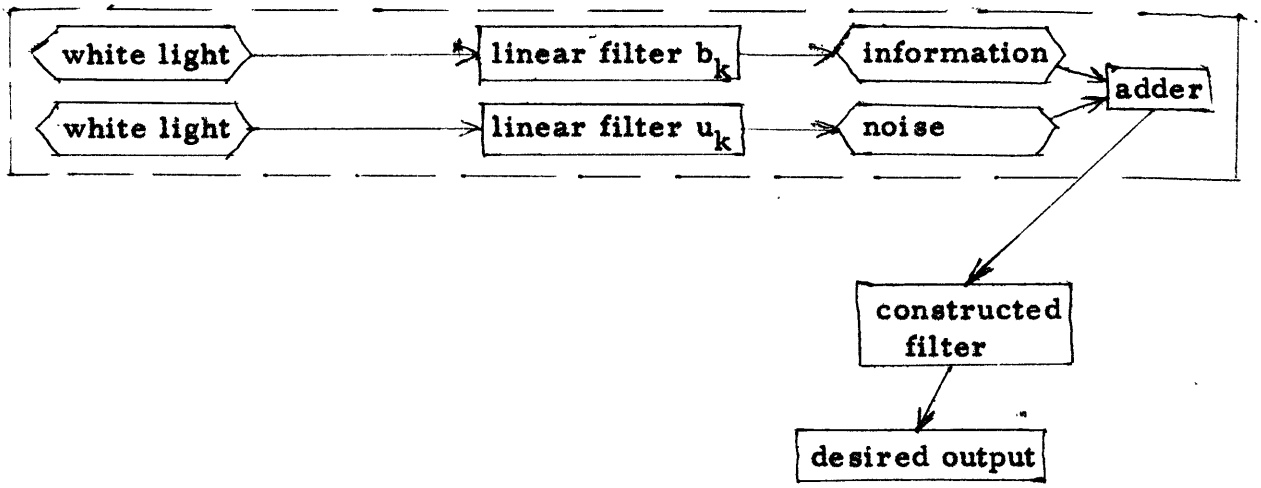
Since the inverse spectrum tends to have much more bandwidth, its wavelet tends to be shorter. This would indicate that when these conditions apply a filter using feedback can do a better job for the number of components than a filter which doesn't.

III Generalized Wave Shaper with Noise

A. Derivation of Normal Equations

Here we imagine the following model of a physical system to apply

Physical System



We want to design a filter to operate on the output of the physical system to give us some preferred output. One set of formulas will enable us to handle the following problems.

Problem 1. Given the information wavelet b_k, the power of the information, and the power spectrum of the noise, convert each information wavelet b_k which comes out of the system to some other waveform d_k. For example we may be converting a long drawn out function b_k into a nice short one like a spike or a minimum phase wavelet. Of course, we do not want the filter to respond very much to noise.

Problem 2. Given the information power spectrum and the noise power

spectrum design a filter so that just the information comes out as uncorrupted as possible. The information might be allowed to come out with some time delay. On the other hand we might want to predict the information before it comes out of the physical system. To see that prediction is a reasonable thing to do, consider an extreme case where noise is absent, the linear filter b_k "rings" for a long while, and the information white light series consists of impulses widely spaced in time. Of course we cannot predict the onset of a ring, but once a ring starts we can easily predict the rest of it.

Problem 2 was treated by Levinson and Problem 1 was solved by the author in connection with some geophysical problems. They are very little different. It will be seen that Problem 2 is a special case of Problem 1 so we begin with Problem 1 and specialize the results later.

Let b be the signal wavelet of length n+1.

Let a be the optimum filter of length m+1.

Let d be the desired convolution of a and b of length n+m+1.

Let u be any noise wavelet.

Let ξ be a white light series which is convolved with u to give a statistical model of the noise process.

Let μ be a white light series of signal wavelet (b) arrival times.

Let $*$ denote convolution.

The input to the filter is the signal plus noise, i.e., $(b*\mu + u*\xi)$. The actual output will be this convolved with the filter, i.e. $(b*\mu + u*\xi)*a$.

The desired output is the wavelet d, occurring every time a signal wavelet arrives, i.e., $d*\mu$. The expected sum squared error is defined as: expected sum squared error = expectation of (actual output-desired output)^2

$$= E\left[\left(b*\mu + u*\xi\right)*a - d*\mu\right]^2 \qquad \text{(III-1)}$$

Since convolution is associative and commutative it is valid and will be convenient to drop all asterisks in the expansion of the above square.

$$= E\left[(b\mu a)^2 + (\mu\xi a)^2 + (d\mu)^2 - 2bda\mu^2 - 2uad\xi\mu + 2bua^2\xi\mu\right] \qquad \text{(III-2)}$$

By taking the expectation inside, it is seen that the last two terms depend on $E(\xi\mu)$. We will assume this to be zero. This means that the signal wavelets arrive at times which are uncorrelated with the noise wavelets.

We recollect the remaining terms.

$$= (ba-d)^2 E(\mu^2) + (ua)^2 E(\xi^2) \qquad \text{(III-3)}$$

From here on the derivation will algebraically resemble that of the spiking filter. It is convenient to rewrite these convolutions in subscript summation notation, i.e.

$$\begin{aligned} ba-d &\rightarrow b_{k-i}a_i - d_k \\ (ba-d)^2 &\rightarrow (b_{k-i}a_i - d_k)(b_{k-j}a_j - d_k) \\ (ua)^2 &\rightarrow (u_{k-i}a_i)(u_{k-j}a_j) \end{aligned} \qquad \text{(III-4)}$$

Since we hope to minimize the expected sum squared error we will take its derivative with respect to each of the independent variables a_i and set each one equal to zero. Hence

$$0 = \frac{\partial}{\partial a_\ell}\left\{ E(\mu^2)\left[b_{k-i} b_{k-j} a_i a_j - d_k\left(b_{k-i} a_i + b_{k-j} a_j \right) + d_k d_k \right] + E(\xi^2)\left[u_{k-i} u_{k-j} a_i a_j \right] \right\} \tag{III-5}$$

This can be expressed in more compact form

$$R_{i-j} = R_{j-i} = b_{k-i} b_{k-j}$$
$$W_{i-j} = W_{j-i} = u_{k-i} u_{k-j} \tag{III-6}$$

having noticed that R and W thus defined are autocorrelation matrices or Toeplitz matrices. The expression simplifies to:

$$0 = \left[E(\mu^2) R_{i-j} + E(\xi^2) W_{i-j} \right] \frac{\partial (a_i a_j)}{\partial a_\ell} - \left[2 E(\mu^2) d_k b_{k-i} \right] \frac{\partial a_i}{\partial a_\ell}$$

$$= \left[\quad " \quad \right]\left(a_i \frac{\partial a_j}{\partial a_\ell} + a_j \frac{\partial a_i}{\partial a_i} \right) - \left[\quad " \quad \right] \frac{\partial a_i}{\partial a_\ell} \tag{III-7}$$

$$= \left[\quad " \quad \right]\left(a_i \delta_{j\ell} + a_j \delta_{i\ell} \right) - \left[\quad " \quad \right] \delta_{i\ell}$$

where δ_{ij} is the Kronecker delta.

Utilizing the symmetry of the quantities in the left hand square brackets we can write:

$$\left[E(\mu^2) R_{i-\ell} + E(\xi^2) W_{i-\ell} \right] a_i = E(\mu^2) b_{k-\ell} d_k \tag{III-8}$$

These equations can easily be rewritten as a matrix equation in the same way as with the spiking filters.

If the desired output d_k were just the signal b_k possibly with some lag or some negative lag (prediction) then the right hand side no longer contains the waveform b_k but only its autocorrelation. This would be the specialization to Levinson's problem.

We now give some examples writing equation (III-8) in matrix form.

Example 1 The signal waveform $b_k = (2,1)$. The signal arrives with a frequency which gives it an average power σ_b^2. The noise is white and has unit power. The filter should have 3 terms. The desired output is a spike after unit delay. $d = (0,1,0,0)$. The normal equations become

$$\left\{ \sigma_b^2 \begin{bmatrix} 5 & 2 & 0 \\ 2 & 5 & 2 \\ 0 & 2 & 5 \end{bmatrix} + 1 \cdot \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \right\} \begin{Bmatrix} a_0 \\ a_1 \\ a_2 \end{Bmatrix} = \begin{Bmatrix} 1 \\ 2 \\ 0 \end{Bmatrix}$$

Example 2 Like example 1 except the desired output is the same as the signal input with no delay. The normal equations are like example 1 except the right hand side becomes the column vector $(5,2,0)^T$.

Example 3 Like example 2 except that the signal should be predicted by one time unit. The normal equations are like example 1 except the right hand side because the column vector $(2,0,0)^T$.

IV Matched Filter

Suppose one is given an autocorrelation function of a noise process and also a signal wavelet. It is desired to detect the arrival of the signal wavelets in the presence of the noise. The method to be used is to filter the incoming mixture of signal and noise and then say that signals arrive where there are maximums in the output.

How should the filter be designed? If the noise were white and the filter memory wavelet had unit energy, then the power output of the filter with noise as input would be unaffected by the frequency characteristics of the filter. Then the filter need concern itself only with the signal. Thus the introductory example (Section I, no. 4) gives the whole story when the noise is white. The result is simply that the signal filter coefficients are just the time reverse of the wavelet and the actual filtering operation then amounts to crosscorrelation of the signal wavelet with the incoming data. If the noise is not white we must do something a bit more complicated.

Using the same notation as the previous section, the power output of the filter with noise input will be the quadratic form $E(\xi^2)\, U_{i-j}\, a_i a_j$. We can choose the magnification constant of the filter to be such that this power is unity. This leads to the constraint

$$E(\xi^2)\, U_{i-j}\, a_i a_j = 1 \qquad \text{(IV-1)}$$

For simplicity we choose to make the filter have the same length as the signal wavelet and we choose to have the maximum output come when the wavelet is exactly in the middle of the filter, i.e., the nth lag of the convolution where both $\vec{a}$ and $\vec{b}$ have length n. Thus we maximize

$$a_i\, b_{n-i} \qquad \text{(sum on i)}$$

subject to the constraint equation (IV-1). Using Lagrange multipliers one maximizes

$$\max\left[\; a_i b_{n-i} + \lambda E(\xi^2)\, U_{i-j}\, a_i a_j \right] \qquad \text{(IV-2)}$$

We have differentiated terms exactly like this in previous sections. Letting $\vec{b}_r$ represent the time reverse of the signal wavelet and U represent the noise autocorrelation matrix, we write the result

$$0 = \vec{b}_r + 2\lambda E(\xi^2) U \vec{a} \qquad \text{(IV-3)}$$

solving for $\vec{a}$ we get

$$\vec{a} = U^{-1} \vec{b}_r \,/\, 2\lambda E(\xi^2) \qquad \text{(IV-4)}$$

We can usually ignore $2\lambda E(\xi^2)$ since it just amounts to a magnification factor in the filter.

In practice one may prefer not to invert the matrix in (IV-4) or solve the simultaneous set (IV-3) since there is an easy way around it. One might simply prefilter the data to whiten the noise and then filter with $\vec{b}_r$. The results would be similar, the difference arising from end effects.

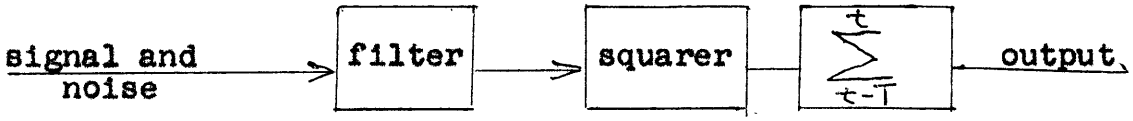
More is known about the matched filter. Suppose one wants to choose a threshold value for the output and announce "signal" whenever the threshold is exceeded and "no-signal" when it is not. Then one would like to maximize the probability of guessing correctly. It can be shown that if the noise is gaussian, then the matched filter and proper choice of threshold will maximize this probability.

V. Maximum Energy Sum Filter

Consider the following physical problem. A transient signal waveform is sent through a dispersive media. The media is such that it may badly disperse the wave without altering its spectral content a great deal. We know what spectrum to expect of the signal and we know the spectrum of the ambient noise. We would like to design an

apparatus or procedure to enable us to make a best guess as to when the signal arrives. The matched filter is not the answer because we do not know the exact signal shape, only its spectrum. The spiking filter is not appropriate for the same reason. The Wiener-Levinson filter tries to make the output look like the signal input. In this case we don't even know what the input waveform should be, we would just like to try to decide approximately when it arrives.

A solution to this problem is to design a filter which puts out lots of energy when the signal comes in and minimum power when only noise comes in. Thus our decision would be based on a system like the following:



We would search for the time t_m when the output was maximum and then we would say that signal arrived between time t_m and time t_m-T.

Taking this model then, we seek to maximize

$$\lambda = \frac{\text{energy output of filter due to signal in interval T}}{\text{expected power output due to noise}} \qquad \text{(V-1)}$$

Notice the similarity of this problem to introductory example 4. It will be seen that it turns out to be exactly the same if the noise is white.

Since we are interested in a computer application, we again specialize ourselves to filters and signals which are discrete in time, and spectra which whose autocorrelations are of finite time duration.

Using the finite autocorrelations of the signal and noise we define two wavelets b_i, a signal wavelet, and u_i, a noise wavelet. This can be done be the procedures described earlier. These two wavelets may have different phase spectra than those of our physical problem, but they will have the correct autocorrelation. Thus we begin with the definitions used earlier:

a_i = "ideal" filter coefficients ($a_i = 0$ if $i<0$ or $i>M$)
b_i = signal wavelet ($b_i = 0$ if $i<0$ or $i>N$)
u_i = noise wavelet ($u_i = 0$ if $i<0$ or $i>N$)
ξ_i = white light series associated with noise process
ξ_i has variance 1.

We use subscript summation notation; the expression

$$a_j b_{k-j}$$

has an implied summation over all values of the repeated index j, j goes from minus to plus infinity. Thus the given expression is a vector with free index k and is the complete transient convolution of a and b.

Expression (V-1) for λ with this convention now becomes

$$\lambda = \frac{a_j b_{k-j}\ a_i b_{k-i}}{a_m u_{l-m}\ a_n u_{l-n}} \qquad \text{(V-2)}$$

We notice that a quantity like $b_{k-j}b_{k-i}$ is the autocorrelation matrix B_{i-j} of the signal b_i and denoting likewise $U_{i-j} = U_{j-i}$ as the autocorrelation matrix of u_i, the expression (V-2) becomes

$$\lambda = \frac{B_{i-j}\, a_i a_j}{U_{m-n}\, a_m a_n} \qquad \text{(V-3)}$$

To try to maximize this ratio, we take its partial derivatives with respect to each of the independent variables a_ℓ and set them equal to zero.

$$0 = \frac{\partial \lambda}{\partial a_\ell} = \frac{(U_{m-n} a_m a_n)\frac{\partial}{\partial a_\ell}(B_{i-j} a_i a_j) - B_{i-j} a_i a_j}{(U_{m-n} a_m a_n)^2} \qquad \text{(V-4)}$$

Multiplying by $(U_{m-n} a_m a_n)$ we get

$$0 = \frac{\partial}{\partial a_\ell}(B_{i-j} a_i a_j) - \lambda \frac{\partial}{\partial a_\ell}(U_{m-n} a_m a_n) \qquad \text{(V-5)}$$

The derivative operations are the same in each term, working only with the first we get

$$\frac{\partial}{\partial a_\ell}(B_{i-j} a_i a_j) = B_{i-j}\left(a_i \frac{\partial a_j}{\partial a_\ell} + a_j \frac{\partial a_i}{\partial a_\ell}\right)$$

$$= B_{i-j}(a_i \delta_{j\ell} + a_j \delta_{i\ell})$$

where δ_{ij} is the Kronecker delta. Now utilizing the symmetry of B_{i-j} and the fact that i and j are dummy variables, this becomes

$$= B_{i-\ell} a_i + B_{\ell-j} a_j$$

$$= 2 B_{i-\ell} a_i \qquad \text{(V-6)}$$

Applying this result in equation (V-5), we obtain

$$0 = B_{i-\ell} a_i - \lambda U_{i-\ell} a_i \qquad \text{(V-7)}$$

This is the generalized eigenvalue problem. Furthermore, since B and U are positive definite*, this problem is known to have M distinct eigenvector solutions for the a_i associated with M eigenvalues λ_m. The eigenvalues must be real and positive. Assuming that the eigenvalues are distinct we select for our solution a_i that eigenvector which is associated with the maximum eigenvalue. We note that eigenvectors are determined only to within a scale factor. This corresponds to the physical fact that the energy power ratio (V-1) will not depend on the amplification of the filter.

Looking back to equation (V-2), we see that the numerator is the energy in the complete transient convolution of a_i and b_i, and denominator is likewise for a_i and u_i. The energy in the convolution of two transients is well known to be the integral of the product of their energy density spectra. Therefore, if we were able to find another wavelet a_i' which had the same amplitude spectrum as a_i, we would have another solution to our maximization problem.

From the z-transform analysis described in Section II, we know that many finite wavelets may have the same spectra. These different wavelets are obtained (by a method due to Wold and also Fejer) in the following way: 1) Compute the autocorrelation of the given wavelet. 2) Factor its z-transform. 3) Its zeros must occur in pairs, specifically if Z_i is a zero, then $1/\bar{Z}_i$ is a zero. Select either one from each pair and form $(Z-Z_1)(Z-Z_2)(Z-Z_3)$. This is the z-transform of a wavelet with the same autocorrelation as the given wavelet. 4) Normally there are 2^n possible different wavelets. By the reasoning of the preceding paragraph, these should all be solutions of our maximization problem.

This is an apparent contradiction to the fact that the eigenvalue problem (V-7) is known to possess a unique

*To see that B is positive definite recall that $\vec{a}B\vec{a}$ is a quadratic form representing the energy of output when the wavelet $\vec{b}$ goes into the filter $\vec{a}$. Clearly this energy is positive for any real values of a. This means that B is positive definite.

eigenvector solution a_1 for the maximum eigenvalue λ_{max}. The contradiction is resolved if and only if all of the zeros of the z-transform of each solution eigenvector lie on the unit circle. Then the zeros z_1 equal their inverse conjugates i.e.

$$z_i = 1 / \overline{z}_i$$

and the 2^n different selections of one from each of the n pairs of zeros all generate the same wavelet.

There is a curious consequence of the fact that the zeros of the z-transform of this filter must be on the unit circle. It is that the eigenvectors must be either symmetric or antisymmetric (for example (2,3,2) or (4,0,-4) respectively). Whether it is symmetric or antisymmetric depends upon whether there are an even or an odd number of zeros at the point Z=1.

This is a simple consequence of the fact that the eigenvectors are real, and any roots of the z-transform which are complex must occur in conjugate pairs. By the main theorem, they must also lie on the unit circle. For the root $z_0 = \alpha + i\beta$ we may then state

$$\alpha^2 + \beta^2 = 1$$

and

$$(z - z_0)(z - \overline{z}_0) = (z - \alpha - i\beta)(z - \alpha + i\beta)$$

$$= z^2 - 2\alpha z + (\alpha^2 + \beta^2)$$

$$= z^2 - 2\alpha z + 1$$

Hence, the coefficients of the second order and the zero

order terms in z are identical for all α and β and the wavelet is symmetric. The same is evidently true for all the complex roots. The net convolution of all these symmetric wavelets is symmetric. Hence, the eigenvector would have to be symmetric if all the zeros were complex. However, we also have the possibility of zeros at two places on the real axis, -1, and +1. The -1 corresponds to symmetric wavelet (1,1), and the +1 corresponds to the antisymmetric wavelet (-1,1). Convolution by the first leaves the eigenvector symmetric, but an odd number of convolutions by the second leave the eigenvector antisymmetric.

Numerical Example

Let

$$b_i = (3, -1, 2)$$

$$u_i = (4, 2, 1)$$

and

$$U = \begin{bmatrix} 21 & 10 & 4 \\ 10 & 21 & 10 \\ 4 & 10 & 21 \end{bmatrix} \qquad B = \begin{bmatrix} 14 & -5 & 6 \\ -5 & 14 & -5 \\ 6 & -5 & 14 \end{bmatrix}$$

solving

$$[B - \lambda U]a = 0$$

we get

$\lambda_1 = 2.7242$ $\quad {}^1a = (1, -1.4924, 1)$

$\lambda_2 = .4706$ $\quad {}^2a = (1, 0, -1)$

$\lambda_3 = .2596$ $\quad {}^3a = (1, 1.7824, 1)$

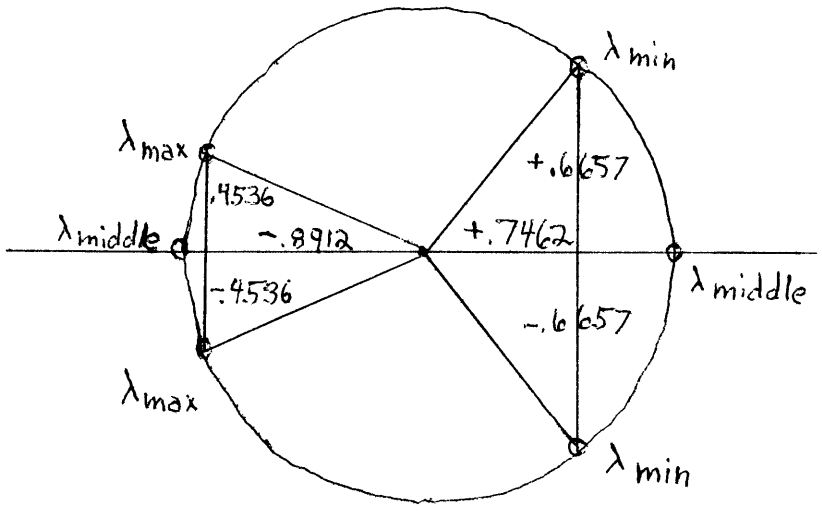
The eigen-values are distinct. The eigenvector solutions for the maximum and minimum eigen-values are seen to be symmetric, and the remaining eigen-value has an anti-symmetric eigen-vector. The zeros of the z-transform of the eigen-vectors are then computed and plotted:

$$F_{1_{CI}}(z) = 1 - 1.4924z + z^2 = (.7462 + .6657i - z)(.7462 - .6657i - z)$$

$$F_{2_{CI}}(z) = 1 - z^2 = (1 - z)(1 + z)$$

$$F_{3_{CI}}(z) = 1 + 1.7824z + z^2 = (-.8912 + .4536i - z)(-.8912 - .4536i - z)$$

The magnitudes of all the zeros are seen to be equal to 1



B. Maximum Energy Sum Filter from Spectral Considerations

We consider the same problem of determining a filter a_i of finite length in discreet time which is optimum in the sense that it maximizes the ratio:

$$I = \frac{\text{(energy output of filter due to signal)}}{\text{(expected power output due to noise)}}$$

This time we solve it in the frequency domain rather than the time domain. Define the filter energy spectrum as $A(\omega)$, the signal energy spectrum as $B(\omega)$ and the noise power spectrum as $G(\omega)$. Then the above ratio may be written:

$$I = \frac{\int_{-\pi}^{+\pi} A(\omega)\, B(\omega)\, d\omega}{\int_{-\pi}^{+\pi} A(\omega)\, G(\omega)\, d\omega} \qquad \text{(V.B.1)}$$

If the maximum of this ratio is finite then it is necessary that for perturbations in $A(\ )$ we will have $I = 0$.

Since $A(\ )$ appears in both numerator and denominator it is clear that a multiplicative scale factor in $A(\ )$ will be unimportant, in other words we can choose the scale factor as we wish. In fact, we can choose it so that the integral in the demoninator is some constant, i.e.

$$1 = \int_{-\pi}^{+\pi} A(\omega)\, G(\omega)\, d\omega \qquad \text{(V.B.2)}$$

Then the problem can be restated as maximizing the numerator

$$0 = \delta \int_{-\pi}^{+\pi} A(\omega)\, B(\omega)\, d\omega \qquad \text{(V.B.3)}$$

subject to the constraint equation (2).

This is a classic problem in the calculus of variations (see for example Hildebrand, Methods of Applied Math, section 2.6). The prodedure is to maximize the quantity

$$I' = \int_{-\pi}^{+\pi} A(\omega) B(\omega) d\omega - \lambda \int_{-\pi}^{+\pi} A(\omega) G(\omega) d\omega \qquad \text{(V.B.4)}$$

subject to no constraint. And then later λ can be determined by (2). λ is called the Lagrange multiplier. Thus we solve the problem:

$$0 = \delta \int_{-\pi}^{+\pi} \left(B(\omega) - \lambda G(\omega) \right) A(\omega) d\omega \qquad \text{(V.B.5)}$$

Since we are dealing with functions in discrete time, the spectra in equation (5) will all be periodic with period 2 (Nyquists). The spectra are also even functions of ω. Therefore, A, B, and G can always be written as

$$A(\omega) = \alpha_0 + \sum_{n=1}^{N} \alpha_n \cos n\omega$$

$$B(\omega) = \beta_0 + 2 \sum_{n=1}^{\infty} \beta_n \cos n\omega$$

$$G(\omega) = \gamma_0 + 2 \sum_{n=1}^{\infty} \gamma_n \cos n\omega \qquad \text{(V.B.6)}$$

Fourier cosine series whose coefficients, the Greek letters, can be recognized as the autocorrelation functions of the respective time functions. The limit on the summation for $A(\omega)$ is finite because the filter a_i was chosen to have finite length and hence so must its autocorrelation. We apply these forms to equation (5).

$$0 = \delta \int_{-\pi}^{+\pi} \left[\beta_0 - \lambda \gamma_0 + 2 \sum_{n=1}^{\infty} (\beta_n - \lambda \gamma_n) \cos n\omega \right] \cdot \left[\alpha_0 + 2 \sum_{n=1}^{N} \alpha_n \cos n\omega \right] d\omega \qquad \text{(V.B.7)}$$

The variation is intended to be over the correlation function α_i of the filter impulse response a_i. The α_i are not, however, allowed to be varied arbitrarily, they must only be varied in such a way as to keep the energy density $A(\omega)$ positive for all ω. In other words an arbitrary selection of the numbers α_i may not really be an autocorrelation function. Therefore, we will express the α_i in terms of the impulse response a_i and do the variation in terms of the a_i instead, because any set of numbers a_i is a valid impulse response. The expressions relating α_i and a_i are:

$$\begin{aligned}
\alpha_0 &= a_0^2 + a_1^2 + a_2^2 + \ldots\ldots + a_N^2 \\
\alpha_1 &= a_0a_1 + a_1a_2 + a_2a_3 \ldots\ldots + a_{N-1}a_N \\
\alpha_2 &= a_0a_2 + a_1a_3 + a_2a_4 + \ldots + a_{N-2}a_N \\
&\vdots \\
\alpha_N &= a_0a_N
\end{aligned} \qquad \text{(V.B.8)}$$

Performing the variation merely amounts to writing the Euler equations in terms of the a_i, the a_i being completely independent variables. Our integral is of a particularly simple form, therefore, we can obtain greater insight by performing the integration directly. Then we can set the variations (derivatives) with respect to the a_i equal to zero.

The integrand is the product of two cosine series. Using the orthonormality of these cosines over the interval $+\pi$ to $-\pi$ equation (7) becomes on integration

$$\begin{aligned}
0 &= 2\pi \frac{\partial}{\partial a_j}\left[(\beta_0 - \lambda\gamma_0)\alpha_0 + 2\sum_{h=1}^{N}(\beta_h - \lambda\gamma_h)\alpha_h\right] \\
&= (\beta_0 - \lambda\gamma_0)\frac{\partial \alpha_0}{\partial a_j} + 2\sum_{h=1}^{N}(\beta_h - \lambda\gamma_h)\frac{\partial \alpha_h}{\partial a_j}
\end{aligned} \qquad \text{(V.B.9)}$$

It is noteable that the formulas (9) no longer contain the infinite sum which is in formula (7). This important result will be referred to later. It means that only N lags of the signal and noise autocorrelations are needed for the solution, N+1 being the length of the impulse response of the filter a_i which we are constructing.

We now differentiate the α_h in equation (8) with respect to the independent variables a_j. This may be written:

$$\frac{\partial \alpha_h}{\partial a_j} = a_{j-n} + a_{j+n}$$

where $0 \leq j \leq N$

$0 \leq n \leq N$

and $a_i \equiv 0$ if $i < 0$

$a_i \equiv 0$ if $i > N$

We now insert this into formula (9) and reorder terms according to increasing subscripts of a_j. This step, although it is complicated amounts to straightforward symbol manipulation. The final result can be written as the following matrix equation:

$$\left\{ \begin{bmatrix} \beta_0 & \beta_1 & \cdots & \beta_n \\ \beta_1 & \beta_0 & & \\ \vdots & & \ddots & \\ \beta_n & & & \beta_0 \end{bmatrix} - \lambda \begin{bmatrix} \gamma_0 & \gamma_1 & \cdots & \gamma_n \\ \gamma_1 & \gamma_0 & & \\ \vdots & & \ddots & \\ \gamma_n & & & \gamma_0 \end{bmatrix} \right\} \begin{Bmatrix} a_0 \\ a_1 \\ \vdots \\ a_N \end{Bmatrix} = 0 \qquad \text{(V.B.11)}$$

Thus we are led to the same result as the time domain considerations (V-7). One wonders whether there might

be a useful connection here with the general theory of eigenfunctions as there were useful results of connecting least squares filters with the theory of orthogonal polynomials.

It is possible and seems likely that some of the statements about decision rules, maximum likelihood, etc. which are made about matched filters in Gaussian noise could also apply to the maximum energy sum filter*. This is a topic which does not appear to have been investigated.

* This possibility was suggested to the author by both Professor E. M. Hofstetter and Professor T. R. Madden.

SECTION VI First Motion Spiking

A. Object and Motivation

The direction of first motion of the ground at a seismic station has received considerable attention in nuclear detection. The essential idea is that the first motion resulting from an explosive blast should always be upward and away from the epicenter while this would probably not be true for more than half of the time for naturally occurring seismic events. This criterion has been shown to be a reasonable one for the Logan and Blanca test shots for distances less than about 700 km (Romney, 1959). The primary difficulty in considering seismograms taken at greater distances was the reduced signal-to-noise ratio further aggravated by the fact that the first motion was in almost all cases smaller than the immediately following oscillations. On some of the seismograms taken at greater distances the first motion appeared to be in the wrong direction despite a fairly strong signal-to-noise ratio. The motivation of the experiment to be discussed is that perhaps the oscillations immediately following the first motion also contain information about the polarity of the first motion, but contain this information in some latent way. This idea is not new, but no effective method has yet been applied to extract this information.

A mathematical technique for extracting this type of information is the spiking filter.

B. Method and Philosophy of the Experiment

First a wavelet, the first motion and several subsequent wiggles, is selected from a relatively near-shot, low-noise, seismogram. Then a filter is designed such that with the wavelet as input, it will produce little or no output before and while the wavelet is entering the filter, a large positive spike when the wavelet has fully

entered the filter, and little or no output thereafter. The filter is also designed to have little output when naturally occurring microseisms are its only input. In practically all cases, a filter cannot be designed to do these simultaneous tasks exactly, but the one designed does them in the least-squared-error sense. That the ultimate error will be sufficiently small for practical purposes must be tested computationally.

The filter is then applied to a seismogram with a poorer signal-to-noise ratio which may be at a different orientation to the seismic event and at a greater distance. If the filtered seismogram consists of low level noise preceding the abrupt arrival of a spike of positive polarity we might then infer that the direction of first motion is the same at the second station as it was at the first. If the impulse had negative polarity we would infer that the second signal had undergone a 180° phase shift with respect to the first signal. If no impulse showed clearly through the background noise, we would infer that this experiment was not successful.

To be more precise, in least-squares fitting to a positive impulse we are assigning a polarity to a clear first arrival wavelet; then we produce a filter which can be applied to wavelets from other seismograms of the same event which assigns a polarity to each of these.

Finally, we are in a position to examine the possibility that the polarity is the same at all orientations from the source. If it is, we infer that the source has rotational symmetry and is probably not of natural origin. If the polarity on the first clear arrival wavelet is assigned according to the direction of first motion, and if wiggles subsequent to the first motion really do contain latent information about the first motion, then the hypothesis tested by this experiment is very similar to, although not exactly the same

as, the hypothesis that the first motion caused by a nuclear explosion must be up and away at all source orientations. To point out this difference more clearly, consider the seismograms mentioned earlier on which the first motions appeared to be in the wrong direction. Possibly the first motion was in the right direction and obscured by the noise, but it might actually have been in the wrong direction. Even if it was, its polarity as determined by the first few wiggles might have been the same as that of other seismograms of the same nuclear event.

C. Choice of Parameters

Several of the seismic records from the Logan underground nuclear explosion were picked by eye, that is, the first motions were identified approximately and the first 3.5 to 4.0 seconds of the seismic trace were considered to be the essence of the signal wavelet. The section was then tapered smoothly to zero on each end. The exact way in which this was done is depicted in Figure 1. Only the shorter of the two wavelets shown (the bottom in each frame) was used. The wavelet length, about 3.75 seconds, was selected because it is long enough to include the requisite "first few wiggles" but not so long as to make the solution of the simultaneous equations excessively time consuming. A sixty point inverse wavelet which is three seconds in length at our standard digitization rate requires about one minute of IBM 709 time to compute.

The choice of a method of tapering the ends of the wavelet was rather arbitrary. It was motivated by two considerations: 1) The time of the first motion arrival could not be determined exactly, and to be sure that the first motion arrival was included, about 3/4 second of the seismic trace before the apparent arrival was included in the wavelet. Since it was also felt that the wiggles nearest the first motion probably contain the most information, wiggles further away were also tapered in amplitude.

2) If the wavelet were just extracted from the seismogram, it would be likely that there would be strong discontinuities in both the function and its derivatives at these ends. It would be undesirable if the spiking filter turned out to be particularly sensitive to these artificially caused discontinuities; hence, they, too, were removed by tapering.

To select the coefficients of the spiking filter, the following quantity was minimized:

sum of square error =

(delta function minus the convolution of the filter with the wavelet)^2

+ 2(the convolution of the filter and the noise)^2

The noise referred to in this expression is the microseismic noise which just preceded the arrival of the signal wavelet. The 2 in the second term on the right in the above expression was selected on the basis of results of earlier crude computational experiments. The choice of the delay in the delta function in the first term on the right in the above expression was made such that the filter would be acting on all of the terms in the wavelet at the time of the filter's spike output.

The length of the spiking filter was chosen to be equal to the length of the wavelet, not because of theoretical necessity, but because it was thought, for various reasons, to be a reasonable choice.

The choice of practically all of the parameters in the above discussion is somewhat arbitrary. They were all selected initially on intuitive grounds. Some have been more or less justified by simple couputational tests, others remain to be investigated.

D. Results

As a check on the computations and a check that the

sum-square-error would be small enough to make the scheme useful, the spiking filter was applied to the seismogram from which it was derived. This is presented in the upper left and lower right frames on Figure 2 and Figure 3. It is seen that the noise preceding the first motion is in all cases reduced and that the first motion is condensed to a neat spike, just as it should be. The "hash" which is near the beginning and end of some of the convolution traces is the result of applying a filter onto the ends of a finite segment of data.

The conclusion to be drawn from the first part of the experiment is that a least squares error filter can be determined with the resulting error small enough that it will be useful in simultaneously reducing noise energy and condensing a particular waveform into a spike.

The next part of the experiment was to apply these spiking filters to other seismograms. The spikes still seem to be present although they are almost down to the level of the noise. This is shown in the lower left and upper right frames in Figure 2 and Figure 3. In some cases the noise before the first motion appears to have increased after filtering. This is because all of the traces on the figures were scaled to have a certain maximum amplitude amenable to scope display. Since the spike was always smaller in cases when the spiking filter was applied to other records, the resulting displays were amplified. The spikes generated from the application of spiking filters upon other seismograms are not clearly distinguishable from the noise in all cases. The conclusion to be drawn from this is that the first motion wavelet loses much, but not all, of its character in going from the station at 1800 km to the station at 1900 km. This must be qualified, however, for the loss of character might not be quite as great as it would first appear; it should be remembered that the wavelet as determined at one

station also includes the noise at that station, hence even if there were no change in the wavelet at all during transmission from one station to the next, there will be a double corrupting effect in this computation due to the different noise at the two stations which cannot be completely eliminated.

F. Possible Modification to and Experimentation on this Mathematical Technique

The operation of the spiking filter in this experiment had the undesirable effect of increasing the high frequency noise. As a result of this, the filtered data looks much more spiky than the unfiltered data making it more difficult to observe a true spike in the filtered data. Heuristically the reasons for this are as follows. The energy in the spectra of the signal wavelet and the noise tends to be primarily at low frequencies. If we were ignoring noise and considering an infinitely long inverse wavelet, its spectrum would be just the inverse of the spectrum of the signal wavelet and in this case would contain very high frequencies. Since the filter is also expected to reject noise of low frequency, the result is a filter which is very sensitive to high frequencies and hence high frequency noise. An important conclusion of this experiment is that something should also be done about high frequency noise. The analysis suggests how to make the filter insensitive to any type of noise of known autocorrelation. Another approach is not to require an impulse to be the output of the filter, but instead, some wider burst. Reasoning again from the limiting case of filters and signals of infinite extent, this would be advantageous because the product of the spectra of the filter with that of the wavelet must equal the spectra of the desired response. By desiring a response of a wide burst instead of a spike we may expect to get a filter less sensitive to high frequency noise.

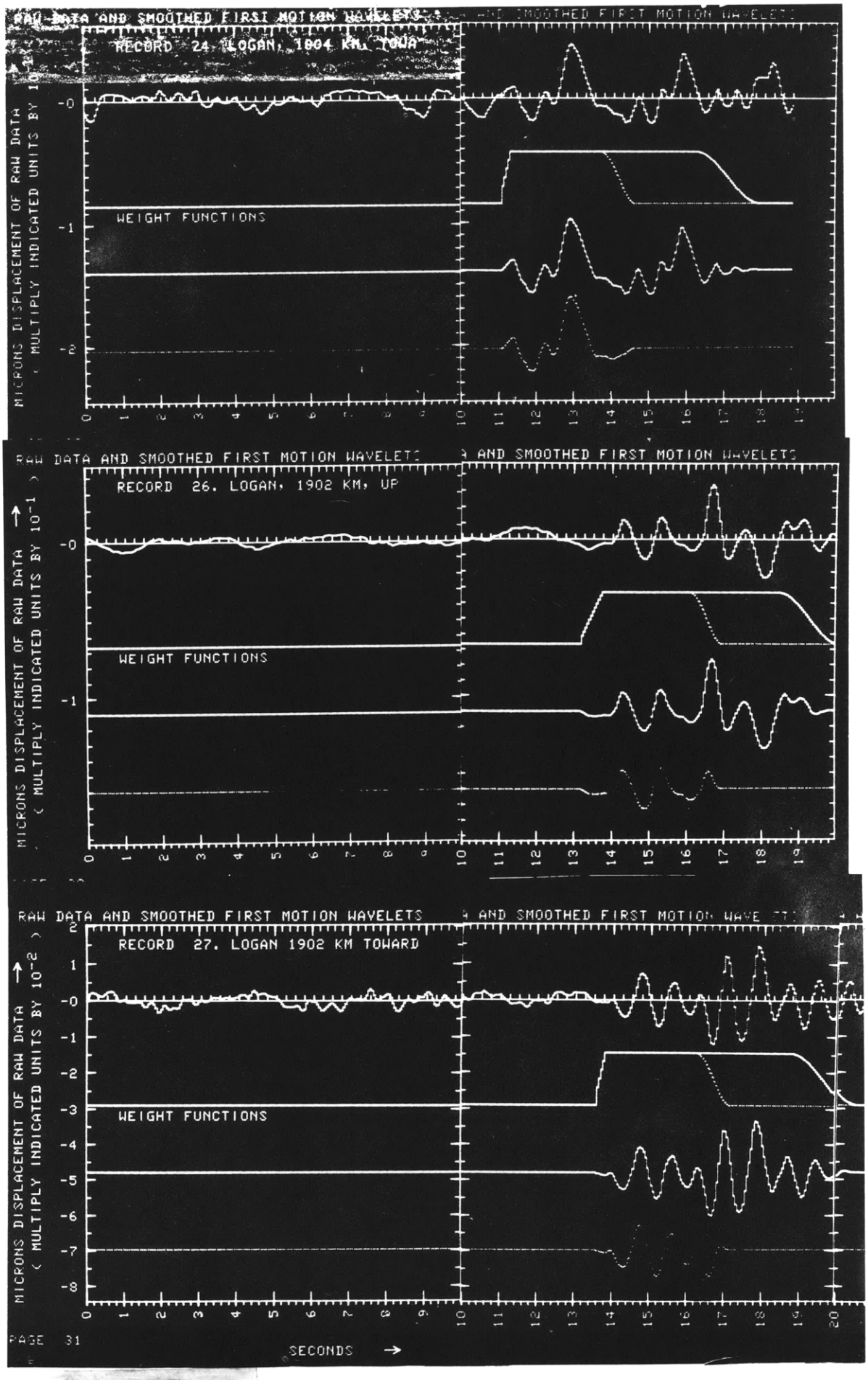

RAW DATA AND SMOOTHED FIRST MOTION WAVELETS
RECORD 24. LOGAN, 1804 KM, TOWA
WEIGHT FUNCTIONS
MICRONS DISPLACEMENT OF RAW DATA
(MULTIPLY INDICATED UNITS BY 10
RAW DATA AND SMOOTHED FIRST MOTION WAVELETS
RECORD 26. LOGAN, 1902 KM, UP
WEIGHT FUNCTIONS
MICRONS DISPLACEMENT OF RAW DATA →
(MULTIPLY INDICATED UNITS BY 10⁻¹)
RAW DATA AND SMOOTHED FIRST MOTION WAVELETS
RECORD 27. LOGAN 1902 KM TOWARD
WEIGHT FUNCTIONS
MICRONS DISPLACEMENT OF RAW DATA →
(MULTIPLY INDICATED UNITS BY 10⁻²)
PAGE 31
SECONDS →

FIG. VI-1

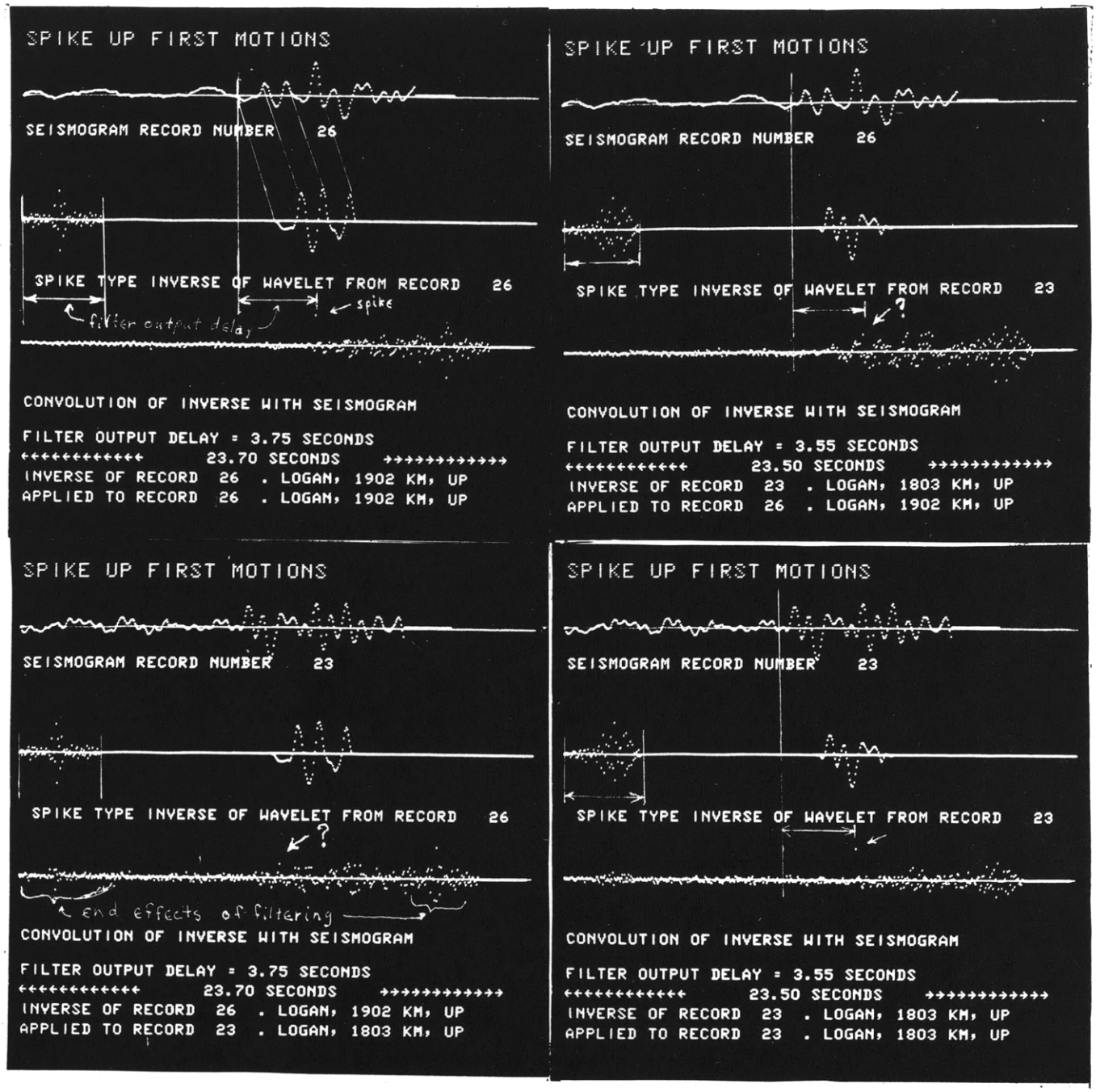
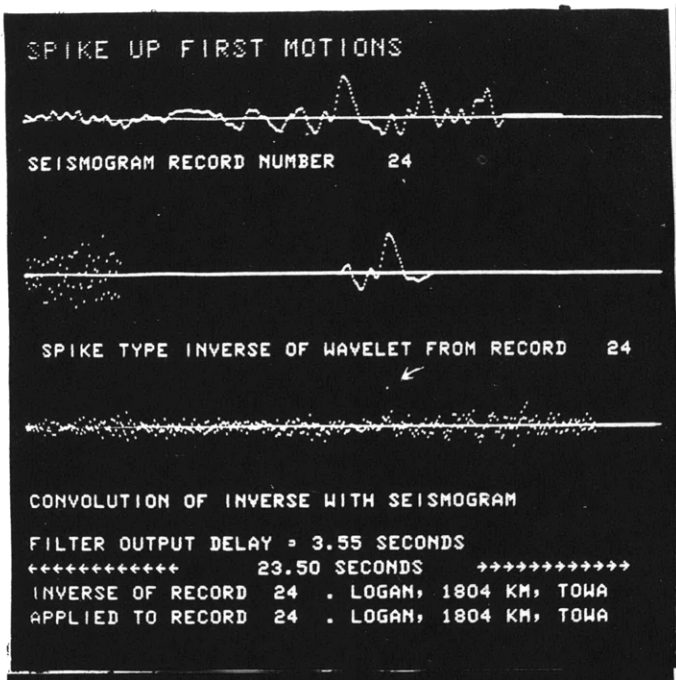
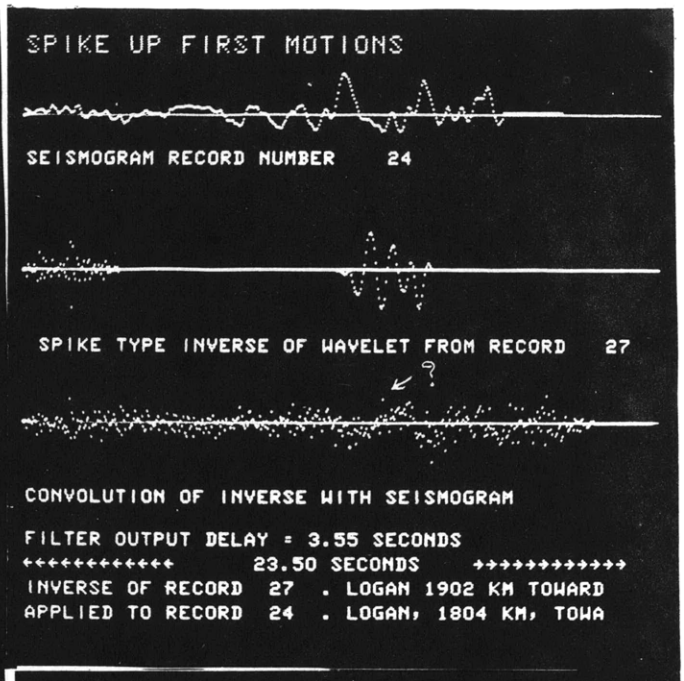
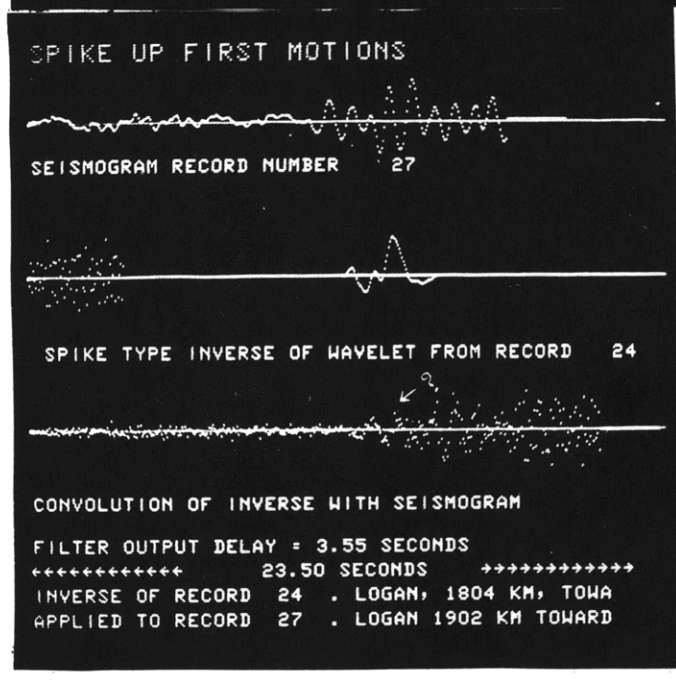

SPIKE UP FIRST MOTIONS
SEISMOGRAM RECORD NUMBER 26
SPIKE TYPE INVERSE OF WAVELET FROM RECORD 26
filter output delay
spike
CONVOLUTION OF INVERSE WITH SEISMOGRAM
FILTER OUTPUT DELAY = 3.75 SECONDS
23.70 SECONDS
INVERSE OF RECORD 26 . LOGAN, 1902 KM, UP
APPLIED TO RECORD 26 . LOGAN, 1902 KM, UP
SPIKE UP FIRST MOTIONS
SEISMOGRAM RECORD NUMBER 26
SPIKE TYPE INVERSE OF WAVELET FROM RECORD 23
?
CONVOLUTION OF INVERSE WITH SEISMOGRAM
FILTER OUTPUT DELAY = 3.55 SECONDS
23.50 SECONDS
INVERSE OF RECORD 23 . LOGAN, 1803 KM, UP
APPLIED TO RECORD 26 . LOGAN, 1902 KM, UP
SPIKE UP FIRST MOTIONS
SEISMOGRAM RECORD NUMBER 23
SPIKE TYPE INVERSE OF WAVELET FROM RECORD 26
?
end effects of filtering
CONVOLUTION OF INVERSE WITH SEISMOGRAM
FILTER OUTPUT DELAY = 3.75 SECONDS
23.70 SECONDS
INVERSE OF RECORD 26 . LOGAN, 1902 KM, UP
APPLIED TO RECORD 23 . LOGAN, 1803 KM, UP
SPIKE UP FIRST MOTIONS
SEISMOGRAM RECORD NUMBER 23
SPIKE TYPE INVERSE OF WAVELET FROM RECORD 23
CONVOLUTION OF INVERSE WITH SEISMOGRAM
FILTER OUTPUT DELAY = 3.55 SECONDS
23.50 SECONDS
INVERSE OF RECORD 23 . LOGAN, 1803 KM, UP
APPLIED TO RECORD 23 . LOGAN, 1803 KM, UP

FIG. VI-2


SPIKE UP FIRST MOTIONS
SEISMOGRAM RECORD NUMBER 24
SPIKE TYPE INVERSE OF WAVELET FROM RECORD 24
CONVOLUTION OF INVERSE WITH SEISMOGRAM
FILTER OUTPUT DELAY = 3.55 SECONDS
23.50 SECONDS
INVERSE OF RECORD 24 . LOGAN, 1804 KM, TOWA
APPLIED TO RECORD 24 . LOGAN, 1804 KM, TOWA


SPIKE UP FIRST MOTIONS
SEISMOGRAM RECORD NUMBER 24
SPIKE TYPE INVERSE OF WAVELET FROM RECORD 27
CONVOLUTION OF INVERSE WITH SEISMOGRAM
FILTER OUTPUT DELAY = 3.55 SECONDS
23.50 SECONDS
INVERSE OF RECORD 27 . LOGAN 1902 KM TOWARD
APPLIED TO RECORD 24 . LOGAN, 1804 KM, TOWA


SPIKE UP FIRST MOTIONS
SEISMOGRAM RECORD NUMBER 27
SPIKE TYPE INVERSE OF WAVELET FROM RECORD 24
CONVOLUTION OF INVERSE WITH SEISMOGRAM
FILTER OUTPUT DELAY = 3.55 SECONDS
23.50 SECONDS
INVERSE OF RECORD 24 . LOGAN, 1804 KM, TOWA
APPLIED TO RECORD 27 . LOGAN 1902 KM TOWARD

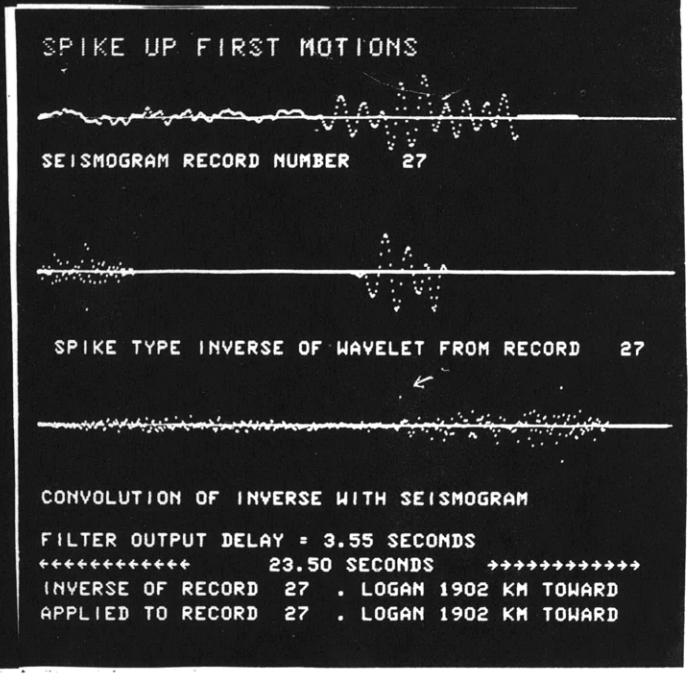

SPIKE UP FIRST MOTIONS
SEISMOGRAM RECORD NUMBER 27
SPIKE TYPE INVERSE OF WAVELET FROM RECORD 27
CONVOLUTION OF INVERSE WITH SEISMOGRAM
FILTER OUTPUT DELAY = 3.55 SECONDS
23.50 SECONDS
INVERSE OF RECORD 27 . LOGAN 1902 KM TOWARD
APPLIED TO RECORD 27 . LOGAN 1902 KM TOWARD

FIG. VI-3

G. More Possible Applications to Nuclear Detection

One could try the following different, though similar experiments:

1) On records taken at the same distance and at the same station try filters generated from a wavelet from one nuclear event on a seismogram from another nuclear event.

2) On a record with a clear first motion, compute the spiking filter and then convolve the whole record with it, in search for later arrivals of the same waveform. (If later arrivals are detected their time delays can be determined to the accuracy of the digitalization sampling. Since this is 1/20 of a second, it may lead to improvements in depth determination accuracies.)

SECTION VII Prediction Error Experiment

I. Philosophy

Microseismic noise can be predicted. For example, it was found that given past values on our noise seismograms, one can easily predict 1/10 of a second into the future with an error in power of less than 5%. Suppose we form a new signal by subtracting the predicted seismogram from the actual seismogram. This new signal is called the prediction error signal. The amplitude of the prediction error signal is expected to be small. If, however, at some time in the microseismic trace a real signal arrives, it cannot, of course, be predicted from the noise. Hence, at that time the prediction error signal should suddenly attain a large amplitude. For example, during the digitization of our seismograms one of the timing marks was accidentally traced. Naturally the timing mark could not be predicted on the basis of the noise which preceded it. The result was a large prediction error at that time. This is depicted in Figure 1.

The mathematical theory of predicting stationary time series at unit prediction distance also shows that the prediction error of a pure noise signal will be a white-light

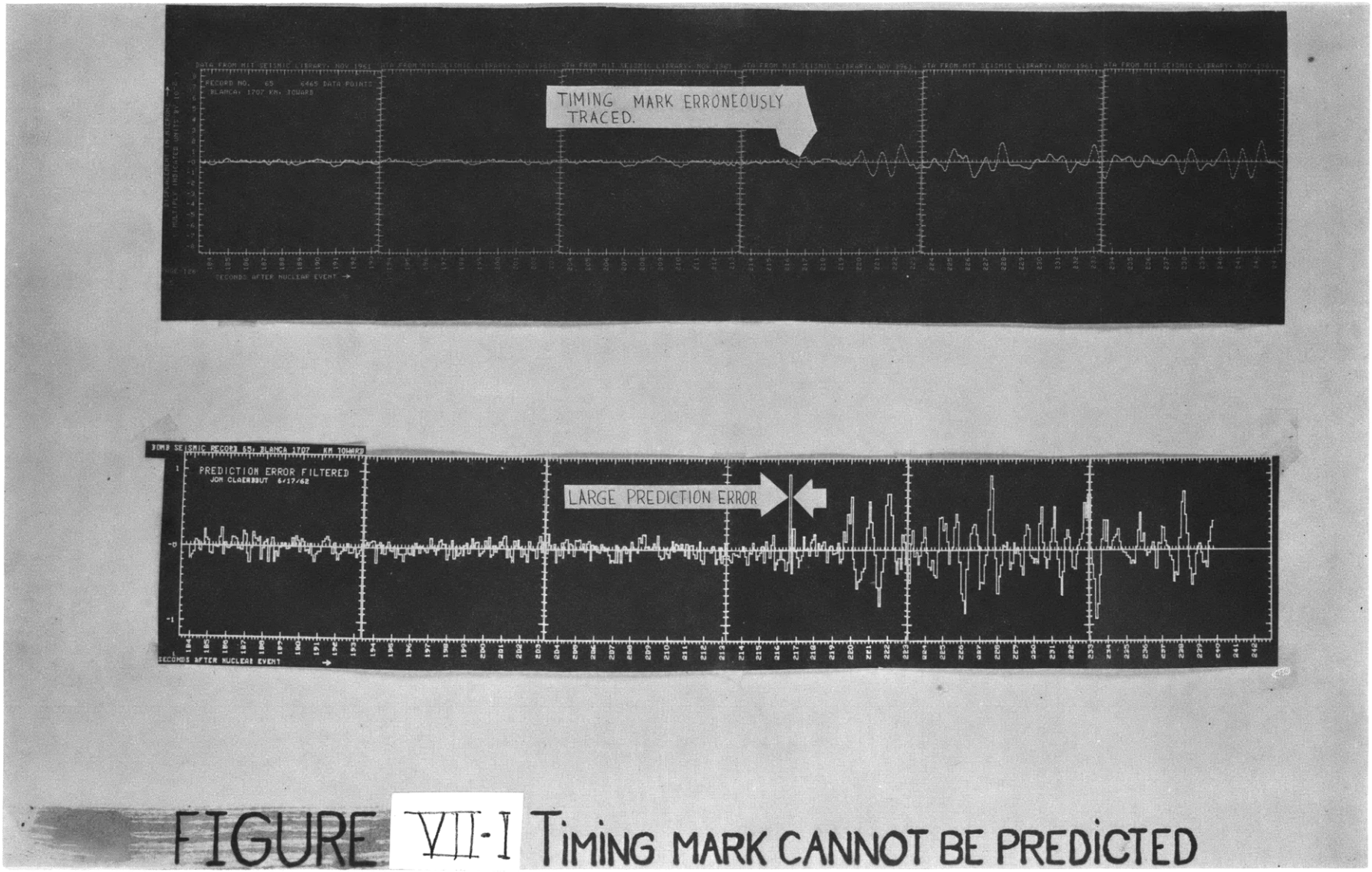


FIGURE VII-1 TIMING MARK CANNOT BE PREDICTED

series. The arrival of a signal, if it has a different spectrum than the noise, will result in non-white series.

Thus a person attempting to find a seismic signal arrival by examining the prediction error will look for:

1) large increase in amplitude

2) change in white character of trace.

There is another peculiarity of the prediction error signal. Its power spectrum is independent of the seismometer and recording system. This is true both before and after P wave arrival. Before, the spectrum is simply white. After, it is a function only of earth motion power spectra.

Another property of the prediction error trace is that the ratio of power after to power before signal arrival must be an improvement from the original seismogram.

All of these properties will now be derived.

II. Mathematical Derivation

The concept of prediction is treated in greater detail elsewhere (Robinson 1954). The formulas are briefly derived here in an heuristic manner.

First we make the following definitions. Let

- s be the given stationary series
- w be the one sided wavelet with the same spectrum as s, of length n
- x be the white light series which when convolved with w gives s
- v be the wavelet which is inverse to w of length m
- d be the predicted s at d time intervals in the future

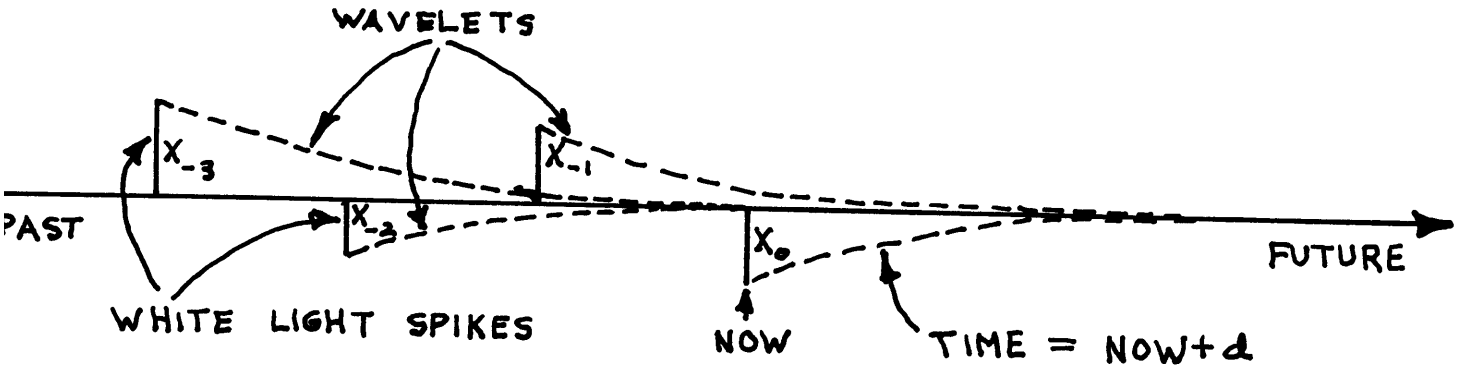
m or n or both may be infinite.

Let negative subscripts refer to the past, the zero subscript to the present, and positive subscripts refer to the future. Let "*" denote convolution. The white light series x can be generated for all past time, up to and

including the present instant by the convolution of s with v; i.e.,

$$X_i = \sum_{j=0}^{m} V_j S_{i-j} = V * S = S * V \qquad (1)$$

The white light series corresponds to arrival times of the wavelet w. The situation is depicted in the sketch below.



To find the predicted value of the series at the time (now+d) we sum up the effect of all wavelets arriving in the past. Those wavelets which may arrive between now and the time we are predicting will contribute to the error of the prediction.

Referring to the sketch above, our prediction now p_0, for the value of s_1 at the future time i=d is thus written:

$$P_0(d) = X_0 W_d + X_{-1} W_{d+1} + X_{-2} W_{d+2} + \cdots\cdots$$

$$P_0(d) = \sum_{i=0}^{n-d} X_{-i} W_{d+i}$$

More generally, the prediction $p_j(d)$ for the value

of the series s_{j+d} at time j+d is written:

$$P_j(d) = \sum_{i=0}^{n-d} x_{j-i} w_{d+i}$$

We write this symbolically as

$$P(d) = X * w_T$$

where w_T is the wavelet w, truncated of its first d terms. Utilizing (1) and the commutivity of convolutions we get

$$P(d) = s * (v * w_T)$$

and we can identify $v*w_t$ as the predicting filter.

The prediction error e_j is defined as

$$e_j(d) = \text{actual series} - \text{prediction of series.}$$

It is a function of the prediction distance d. We will now show that if d=1, the operator which generates $e_j(1)$ from s_j takes on a particularly simple form, and $e_j(1)$ must be a white light series.

Denoting z-transforms by capital letters, the z-transform of the truncation of w_t corresponding to d=1 is

$$W(z) - W_0$$

The z-transform of the predicting filter is then

$$V(z)\left(W(z) - W_0\right)$$

The z-transform of the prediction error filter is just

$$1 - V(z)\left(W(z) - W_0\right)$$

$$= 1 - V(z)\,W(z) + W_0\,V(z)$$

But w and v are inverses and also, $w_o = 1/v_o$, hence the z-transform of the prediction error filter is

$$= 1 - (1 + 1/v_o) V(z)$$
$$= 1/v_o \, V(z)$$

Hence, the prediction error filter is just the inverse wavelet, scaled so that the first term of the filter is +1. Since the prediction error filter is the inverse to the wavelet of the stationary series, it must whiten the series.

What happens to the spectrum if a signal arrives somewhere on the noise record? Letting S denote spectrum, the condition that the noise be whitened is:

S (earth noise) S (seism. system) S (prediction error filter) = 1

The spectrum of our final graph is then:

S(graph) = S (earth signal + noise) S (seism. system) S (prediction error filter)

Combining the above two expressions we get:

$$S \text{ (graph)} = \frac{S \text{ (earth signal + noise)}}{S \text{ (earth noise)}}$$

which is independent of the transfer function of the seismograph.

This derivation contains some hidden mathematical assumptions which should be valid in any real case. (Seismograph system is linear and dissipative. Ground motion satisfies Paley-Wiener criterion.)

The proof that the prediction error filter must improve the signal-to-noise ratio is omitted. It is based

on the fact that the prediction filter can be derived from the point of view of minimizing the variance of the difference between the predicted and actual noise, and that this variance must be higher for any signal with autocorrelation different from that of the noise.

III. Computational Method

One knows approximately the signal first arrival time on all of our seismograms. In some cases it is directly observable, in others one needs to use travel-time curves. The autocorrelation of the noise before the first motion is first computed. From this the inverse wavelet is computed by the method described in our previous report Appendix F part III. This is a least squares method. The length of this filter was chosen to be 70 points. This is near the limit of computational feasability* of least square procedures at the present time. A method for computing longer prediction operators was programmed but not used because in most cases we did not have a very great amount of data digitized before the first motion and also because experience has shown that great increases in operator length do not improve predictability proportionately. Our data has 1/20 second digitization intervals, however, we have discovered that our seismograms have little energy in the spectrum above 5 cps. Therefore, only alternate digitized points were used. The resulting prediction operator length is 7 seconds.

The finiteness of this operator caused our actual output to deviate from the theoretical output in the following way: The operator cannot successfully use noise with wavelength of the order of 7 seconds and longer in prediction, since it is only 7 seconds in length. Reference to graphs in our previous reports indicates that 5% to 20% of the power in the spectrum may fall within this range. Although this low frequency is apparent in some of the

* for large improvement see section II

prediction error traces, the visual quality of the records is not impaired, however, due to the very lowness of this frequency.

IV. Results

Results are presented in the form of the following figures. The results are good in every case and sometimes quite remarkable.

Figure IV-1-2 Prediction Error Filtering Examples Notice particularly on the UP component. The filtered trace becomes markedly non-white at 209.8 seconds. The first break on the unfiltered trace is not readily apparent until 210.8. Considerable signal to noise energy improvement is noticed on all traces. It is very difficult to pick out the first break on the left component, but the filtered left component is markedly non-white by 211.5 sec.

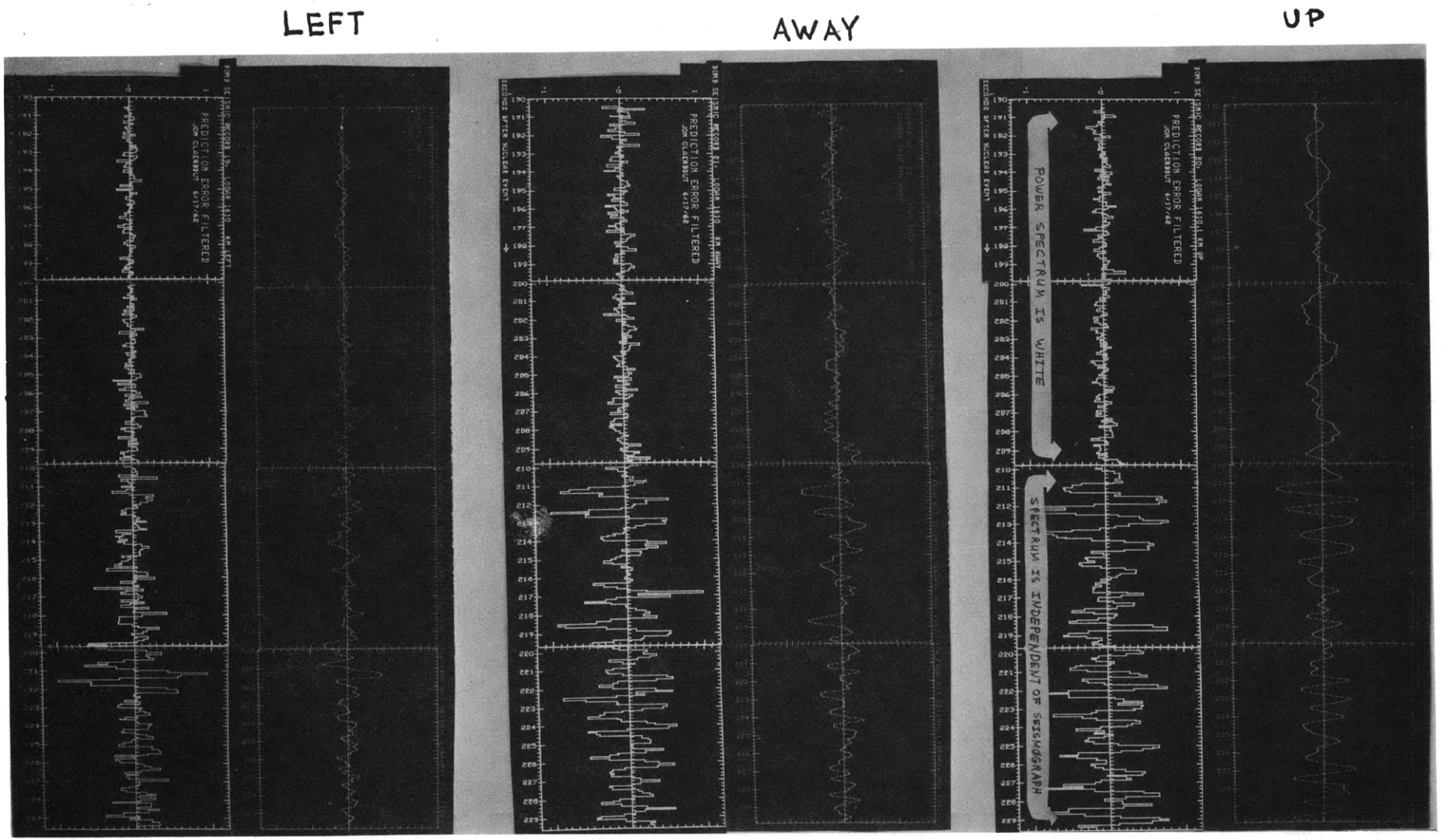


Figure VII-3 Prediction Error Filtering Examples The signal-to-noise ratio is approximately the power in the seismogram after the first break divided by the power before the first break. It is noted to be increased for all 3 components and markedly so for the UP and TOWARD components. The first break is difficult to pick on the LEFT trace. The filtered LEFT, however, becomes markedly non-white at 231.2 seconds, which we know from the vertical component to be the correct first break time.

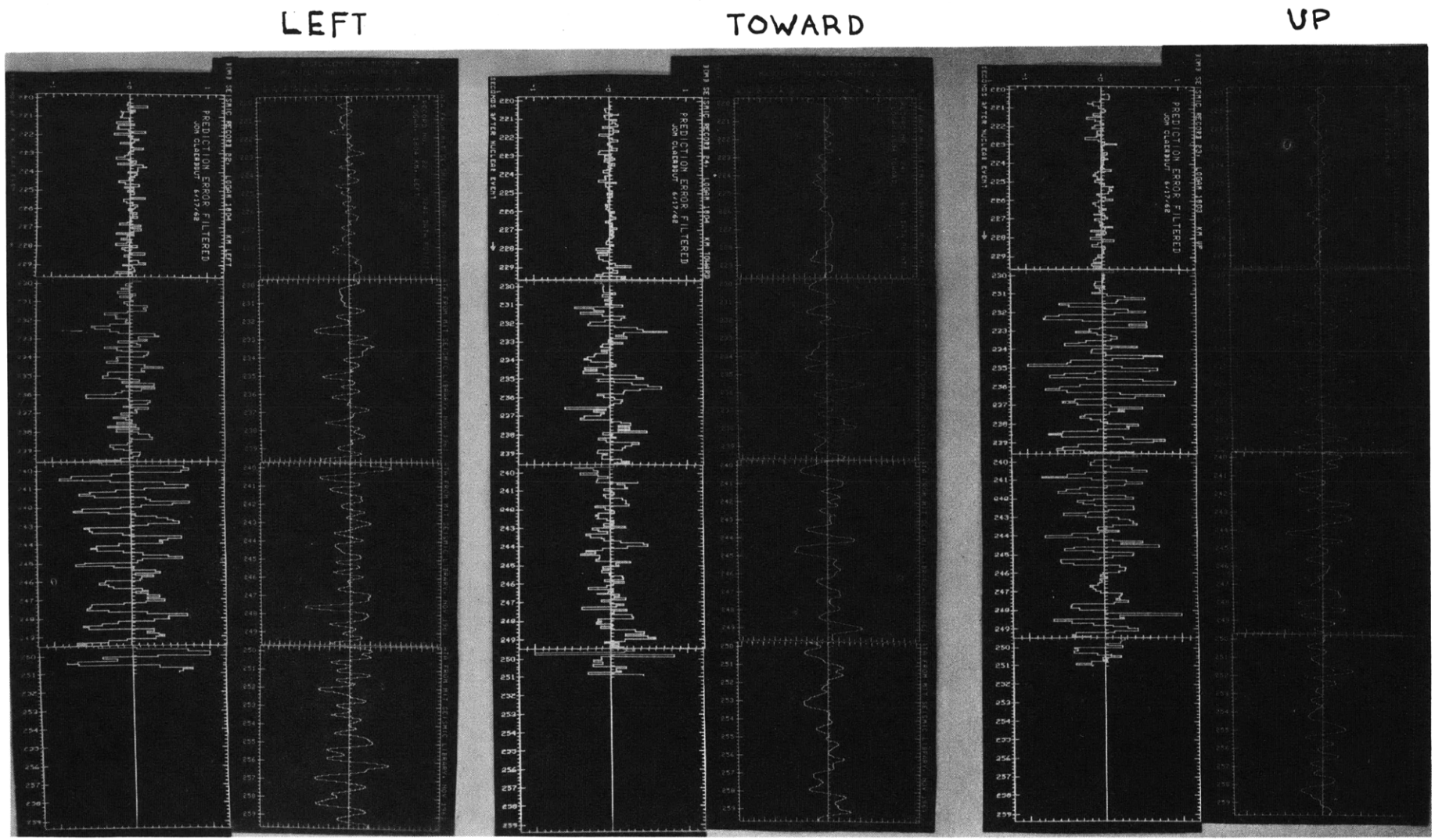


Figure IV-1-4 Prediction Error Filtering Example Again a marked improvement in signal-to-noise ratio is noted on all components. The magnitude of the first motion is often increased with respect to the noise. In some cases it is not increased. If it is not increased this reflects the fact that the original seismogram is a bit misleading. The noise trace was just about to move abruptly up or down when the signal came along and reinforced this motion. Thus the first motion is not as big as it might seem on the original data.

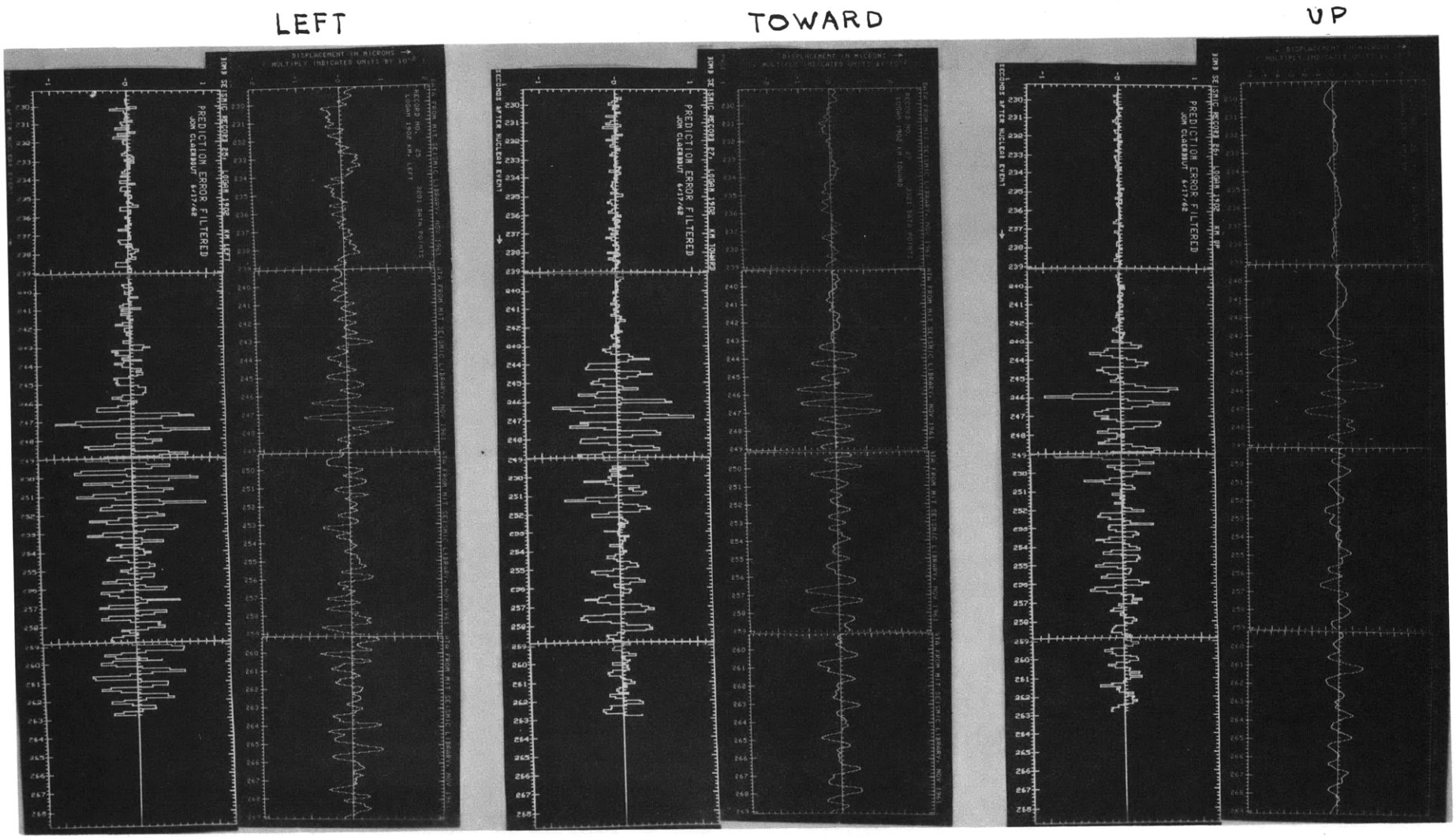
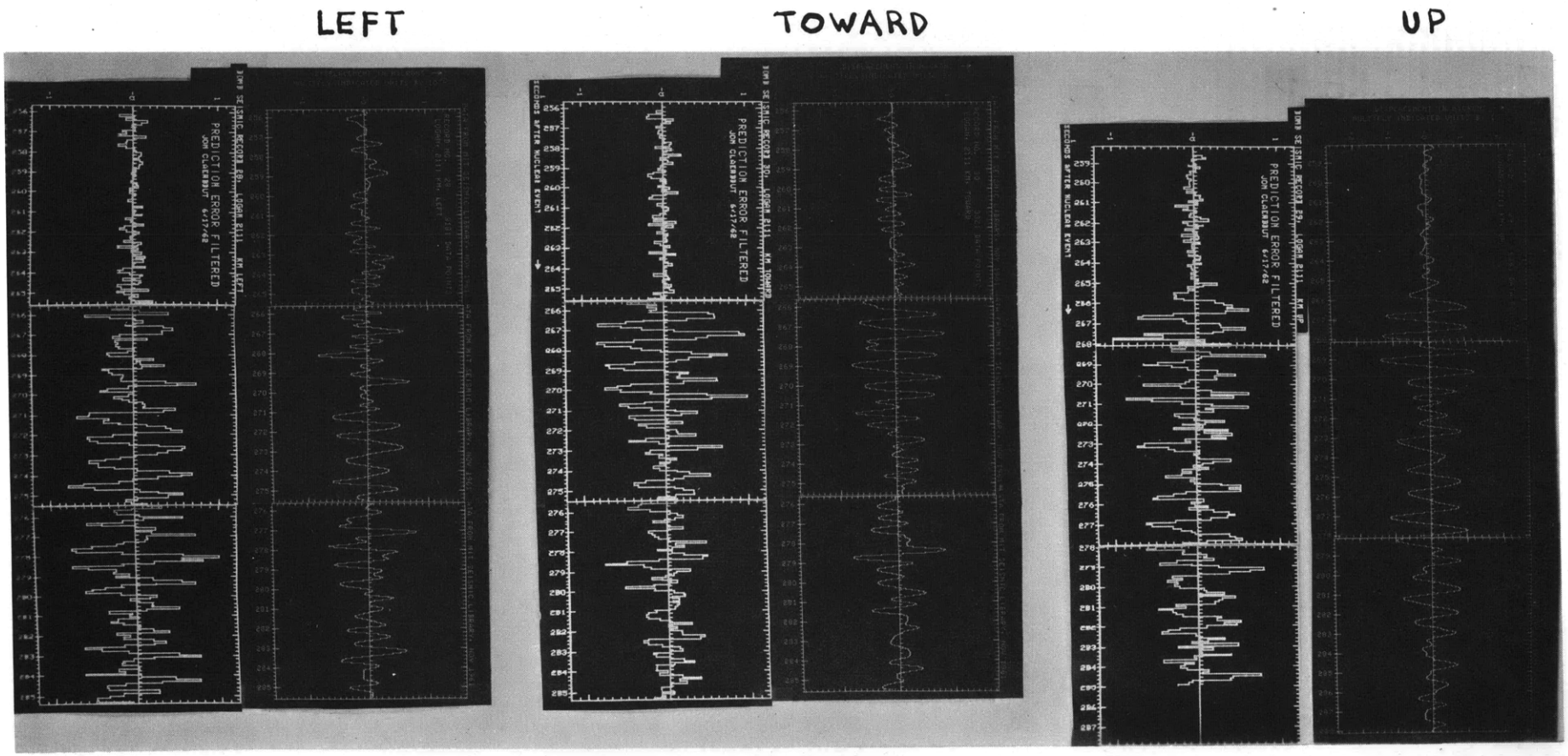


Figure VII-5 Prediction Error Examples Again a marked improvement is noted in signal-to-noise-energy ratio, especially on the left trace. The first break is difficult to determine on the UP component in either the filtered or the unfiltered trace.



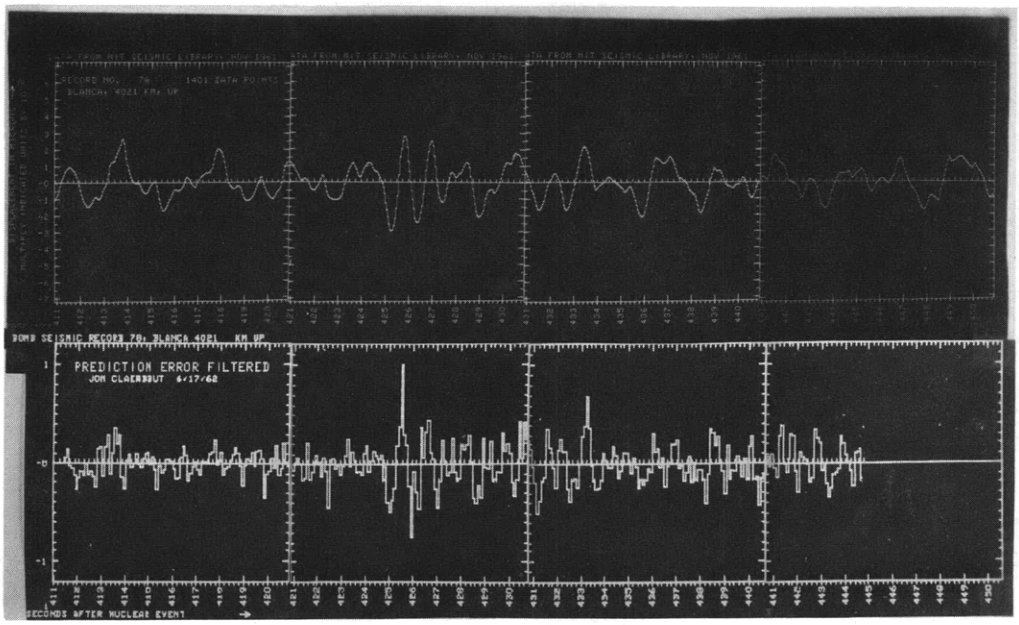


Figure VII-6 The p-wave is clearly located at 425 seconds. This is an example where all but perhaps a skilled seismologist would not be able to pick p from the original record, but where it is quite clear from the prediction error record.

SECTION VIII: Travelling Auto-Spectra of Nuclear Shot Seismograms

A travelling spectrum is a succession of spectral estimates of a time function taken at successive time intervals. Thus it is a function of both frequency and time. This concept, although is is a mathematical amalgam, may be useful in the analysis of non-stationary time series where the successive spectral estimates change in some physically meaningful way.

It was not certain what could be learned by taking the travelling spectra of seismograms of p- and s-waves from nuclear shots since simple theory predicts no dispersion for these phases in a homogeneous isotropic medium. But considerable change of waveform (i.e. dispersion) is known to occur in the real earth. Therefore, although one has no detailed ideas of what information it might be able to extract in regard to nuclear detection, it was thought there might be value in computing the travelling spectra, especially since by utilizing a special technique (Simpson et al., 1961a, Appendix J) it was possible to compute a travelling 24-point spectrum of a typical seismogram on the IBM 709 in the amount of time it takes to read this sentence. In fact, it is too easy to use the computer to generate many more numbers and curves than are readily interpretable. For the first investigation travelling spectra was computed for all the digitized data which was available.

Since the travelling spectrum is a function of two parameters, frequency and time, and since our program can compute values almost as fast as they can be printed, there was a significant problem in data presentation. I took two approaches. The first was to print twelve numbers per line of printed page, these being the spectral amplitude estimates scaled to a maximum of 5, rounded to an integer, and then taken to the 10th power. The result

is intended to resemble 13 bar graphs running down the page, representing spectral estimates at 13 frequencies as a function of time. Time of p-wave and s-wave arrival is indicated. The second approach to the data presentation problem is to make these bar graphs on the scope. This allows finer presentation of amplitude.

A selected few of the results are presented in the figures. Some things are notable. On Figure II-3-1 is presented the travelling spectra from two nuclear shots over almost identical paths. The spectra are similar, but far from being identical. On Figure II-3-5 the s-wave arrival is apparent on the travelling spectra as an increase in high-frequency energy. On Figure II-3-6 a phase arrival is noted in which there appears to be some dispersion. This phase has not yet been identified.

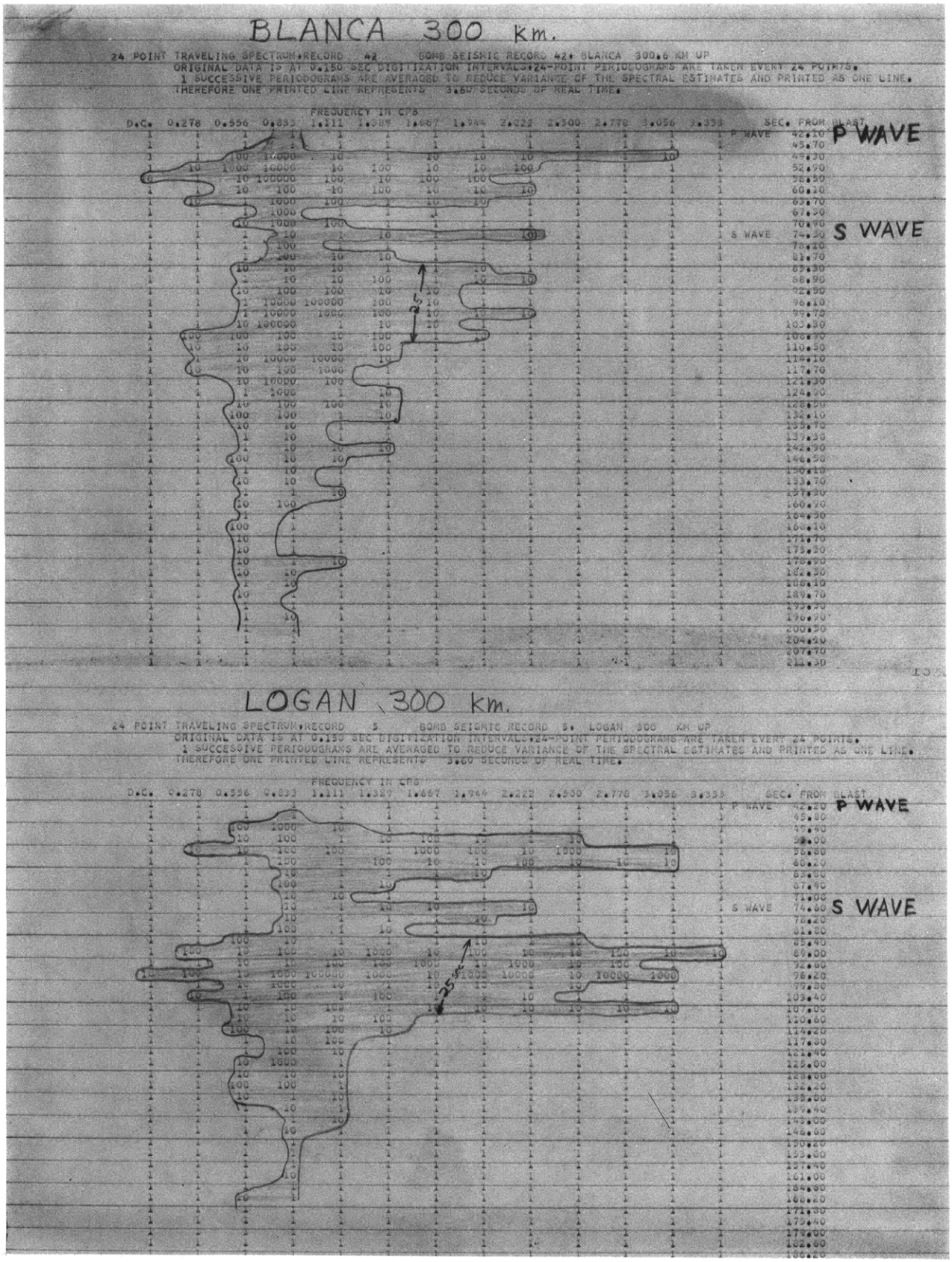


Figure VIII - 1 Similarity, but not identity of traveling spectra from two nuclear shots over almost identical paths.

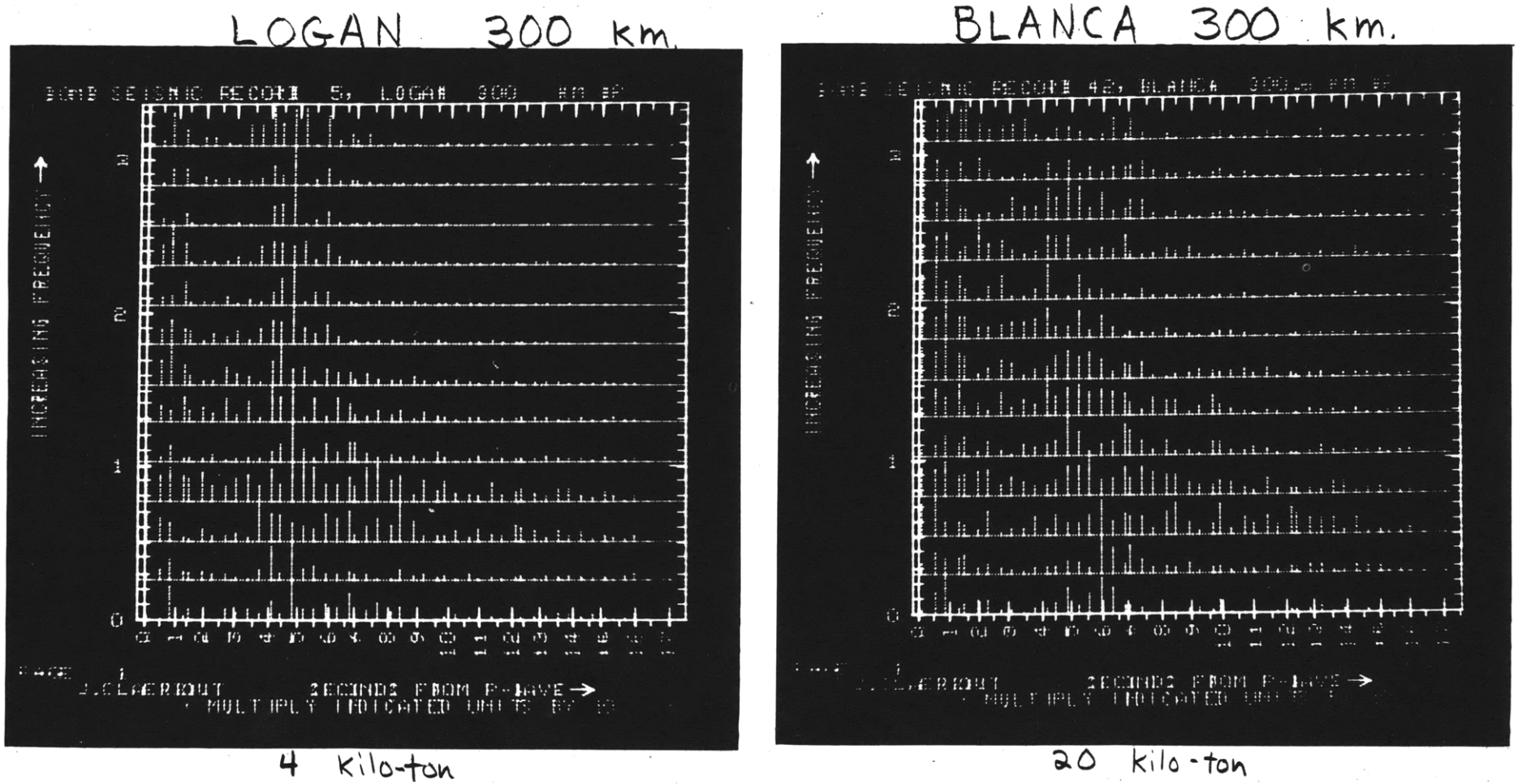


Figure VIII-2 Traveling Spectra from Logan and Blanca This is another type of presentation of the information in Figure II-3-1.

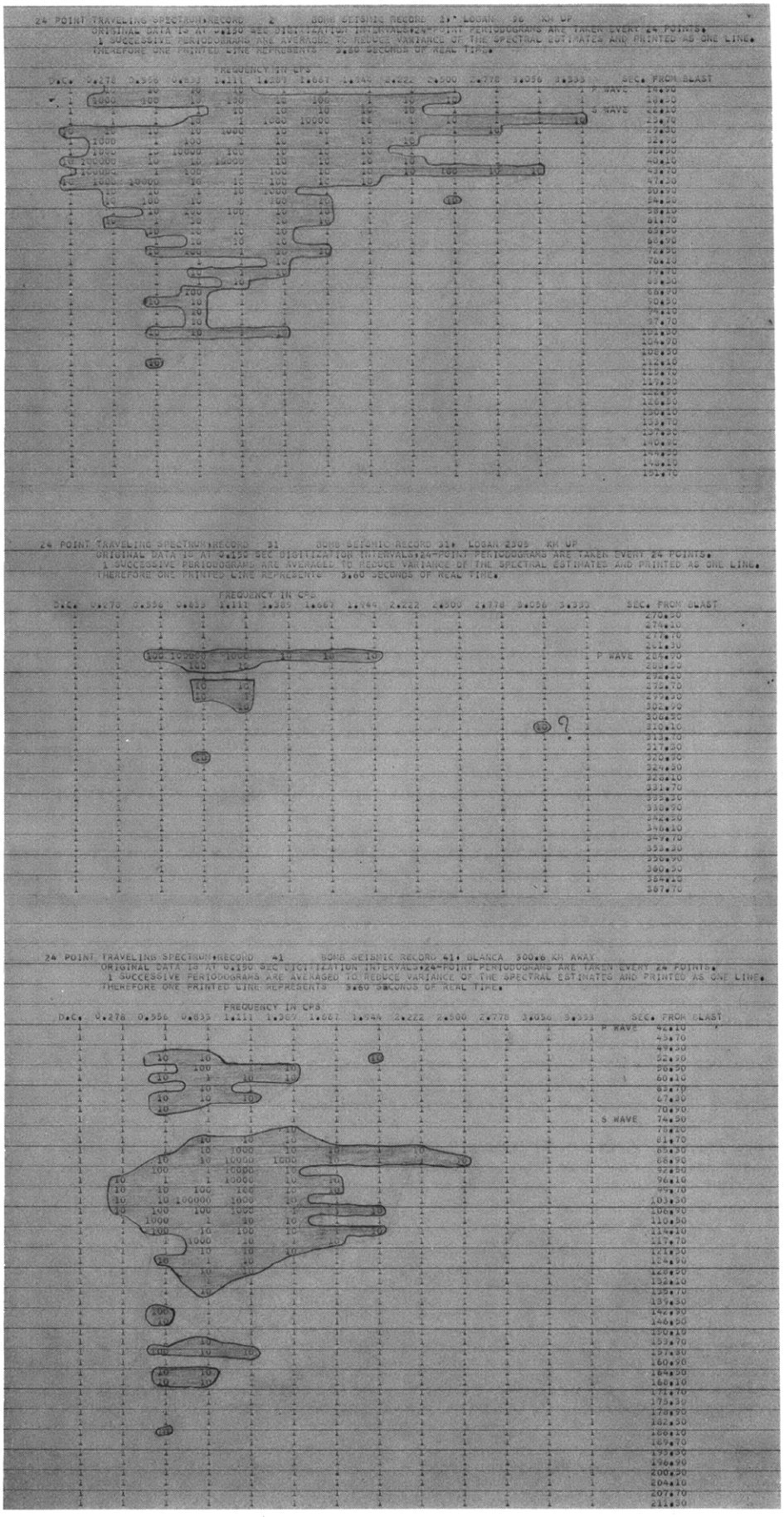
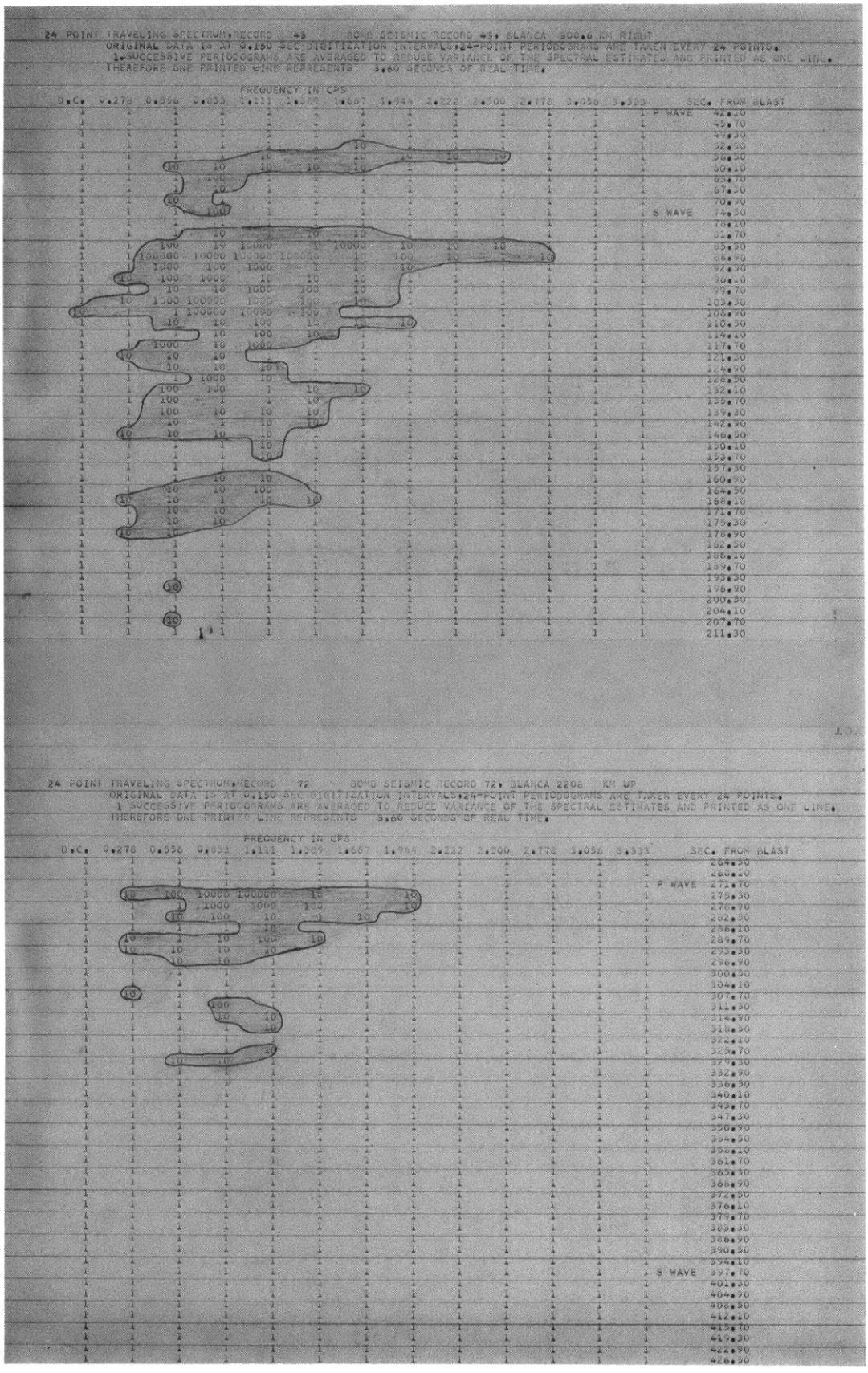


Figure VIII - 3 Some Miscellaneous Traveling Spectra



24 POINT TRAVELING SPECTRUM, RECORD 43 BOMB SEISMIC RECORD 43, BLANCA 300.6 KM RIGHT
ORIGINAL DATA IS AT 0.150 SEC DIGITIZATION INTERVALS, 24-POINT PERIODOGRAMS ARE TAKEN EVERY 24 POINTS.
1 SUCCESSIVE PERIODOGRAMS ARE AVERAGED TO REDUCE VARIANCE OF THE SPECTRAL ESTIMATES AND PRINTED AS ONE LINE.
THEREFORE ONE PRINTED LINE REPRESENTS 3.60 SECONDS OF REAL TIME.

FREQUENCY IN CPS

D.C.	0.278	0.556	0.833	1.111	1.389	1.667	1.944	2.222	2.500	2.778	3.056	3.333	SEC. FROM BLAST
1	1	1	1	1	1	1	1	1	1	1	1	1	P WAVE 42.10
1	1	1	1	1	1	1	1	1	1	1	1	1	45.70
1	1	1	1	1	1	1	1	1	1	1	1	1	47.30
1	1	1	1	1	1	10	1	1	1	1	1	1	52.50
1	1	1	1	10	1	10	10	10	10	1	1	1	56.50
1	1	10	10	10	10	10	1	1	1	1	1	1	60.10
1	1	1	100	1	1	1	1	1	1	1	1	1	63.70
1	1	1	10	1	1	1	1	1	1	1	1	1	67.30
1	1	10	1	1	1	1	1	1	1	1	1	1	70.90
1	1	1	100	1	1	1	1	1	1	1	1	1	S WAVE 74.50
1	1	1	1	10	1	1	1	1	1	1	1	1	78.10
1	1	1	10	10	10	10	1	1	1	1	1	1	81.70
1	1	100	10	10000	1	10000	10	10	10	1	1	1	85.30
1	1	100000	10000	100000	100000	10	100	10	1	10	1	1	88.90
1	10	1000	100	1000	1	10	10	1	1	1	1	1	92.50
1	1	100	1000	10	10	10	1	1	1	1	1	1	96.10
1	10	10	10	1000	100	10	1	1	1	1	1	1	99.70
1	10	1000	100000	1000	100	10	1	1	1	1	1	1	203.30
10	1	1	100000	10000	100	10	1	1	1	1	1	1	106.90
1	1	10	10	100	10	10	10	1	1	1	1	1	110.50
1	1	1	10	100	10	1	1	1	1	1	1	1	114.10
1	1	1000	10	1000	1	1	1	1	1	1	1	1	117.70
1	10	10	10	1	1	1	1	1	1	1	1	1	121.30
1	1	10	10	10	1	1	1	1	1	1	1	1	124.90
1	1	1	1000	10	1	1	1	1	1	1	1	1	128.50
1	1	100	100	1	10	10	1	1	1	1	1	1	132.10
1	1	100	1	1	10	1	1	1	1	1	1	1	135.70
1	1	100	10	10	10	1	1	1	1	1	1	1	139.30
1	10	1	10	10	1	1	1	1	1	1	1	1	142.90
1	10	10	10	1	1	1	1	1	1	1	1	1	146.50
1	1	1	1	10	1	1	1	1	1	1	1	1	150.10
1	1	1	1	10	1	1	1	1	1	1	1	1	153.70
1	1	1	1	1	1	1	1	1	1	1	1	1	157.30
1	1	1	10	10	1	1	1	1	1	1	1	1	160.90
1	10	10	10	100	1	1	1	1	1	1	1	1	164.50
1	10	10	1	10	10	1	1	1	1	1	1	1	168.10
1	1	10	10	1	1	1	1	1	1	1	1	1	171.70
1	1	10	10	1	1	1	1	1	1	1	1	1	175.30
1	10	10	1	1	1	1	1	1	1	1	1	1	178.90
1	1	1	1	1	1	1	1	1	1	1	1	1	182.50
1	1	1	1	1	1	1	1	1	1	1	1	1	186.10
1	1	1	1	1	1	1	1	1	1	1	1	1	189.70
1	1	1	1	1	1	1	1	1	1	1	1	1	193.30
1	1	10	1	1	1	1	1	1	1	1	1	1	196.90
1	1	1	1	1	1	1	1	1	1	1	1	1	200.50
1	1	10	1	1	1	1	1	1	1	1	1	1	204.10
1	1	1	1	1	1	1	1	1	1	1	1	1	207.70
1	1	1	1	1	1	1	1	1	1	1	1	1	211.30

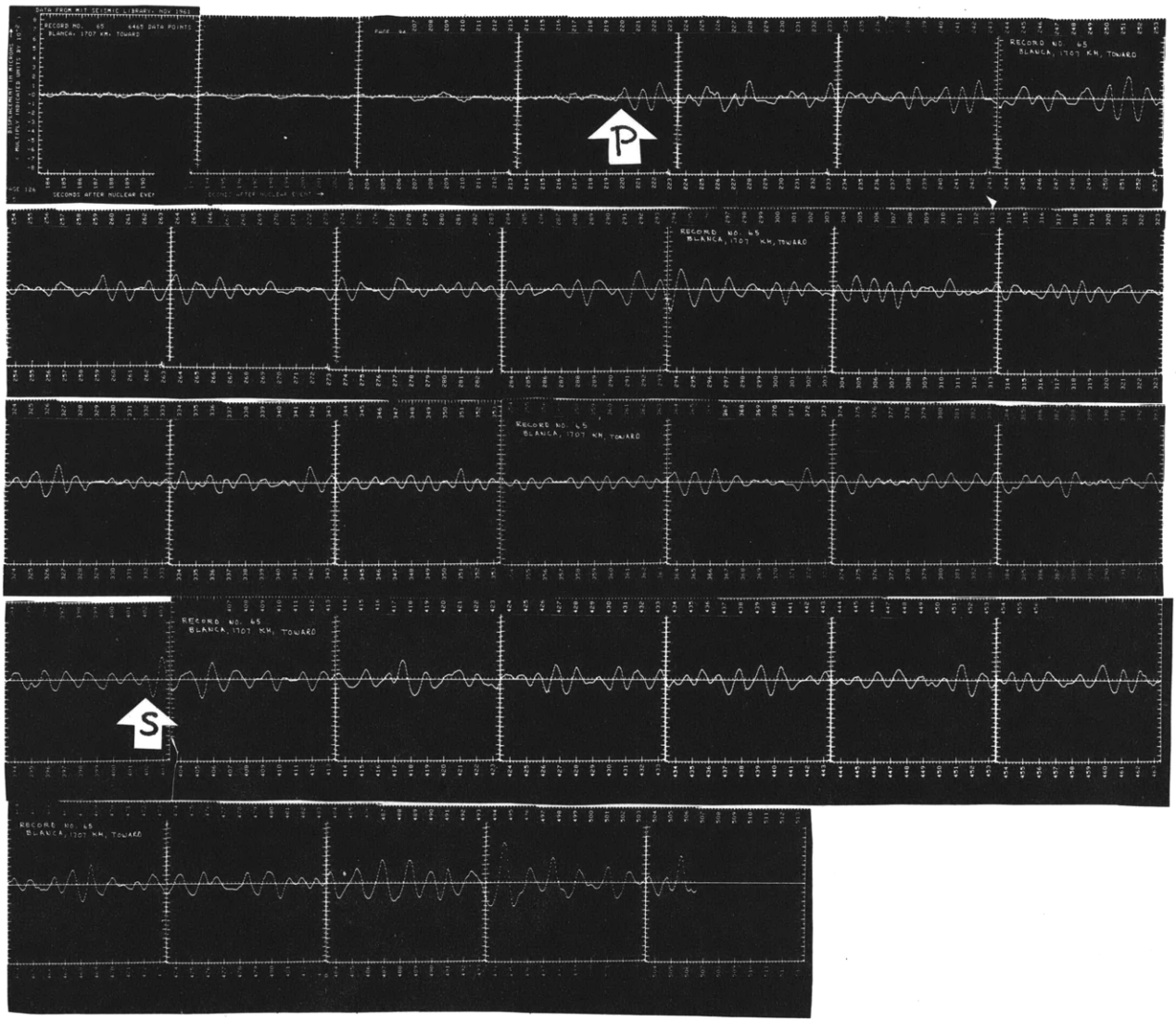
24 POINT TRAVELING SPECTRUM, RECORD 72 BOMB SEISMIC RECORD 72, BLANCA 220.8 KM UP
ORIGINAL DATA IS AT 0.150 SEC DIGITIZATION INTERVALS, 24-POINT PERIODOGRAMS ARE TAKEN EVERY 24 POINTS.
1 SUCCESSIVE PERIODOGRAMS ARE AVERAGED TO REDUCE VARIANCE OF THE SPECTRAL ESTIMATES AND PRINTED AS ONE LINE.
THEREFORE ONE PRINTED LINE REPRESENTS 3.60 SECONDS OF REAL TIME.

FREQUENCY IN CPS

D.C.	0.278	0.556	0.833	1.111	1.389	1.667	1.944	2.222	2.500	2.778	3.056	3.333	SEC. FROM BLAST
1	1	1	1	1	1	1	1	1	1	1	1	1	264.50
1	1	1	1	1	1	1	1	1	1	1	1	1	268.10
1	1	1	1	1	1	1	1	1	1	1	1	1	P WAVE 271.70
1	10	100	10000	100000	10	1	10	1	1	1	1	1	275.30
1	1	1	1000	1000	100	1	10	1	1	1	1	1	278.90
1	1	10	100	10	1	10	1	1	1	1	1	1	282.50
1	1	1	1	10	1	1	1	1	1	1	1	1	286.10
1	10	1	10	100	10	1	1	1	1	1	1	1	289.70
1	10	10	10	10	1	1	1	1	1	1	1	1	293.30
1	1	10	10	1	1	1	1	1	1	1	1	1	296.90
1	1	1	1	1	1	1	1	1	1	1	1	1	300.50
1	10	1	1	1	1	1	1	1	1	1	1	1	304.10
1	1	1	100	1	1	1	1	1	1	1	1	1	307.70
1	1	1	10	10	1	1	1	1	1	1	1	1	311.30
1	1	1	1	10	1	1	1	1	1	1	1	1	314.90
1	1	1	1	1	1	1	1	1	1	1	1	1	318.50
1	1	1	1	10	1	1	1	1	1	1	1	1	322.10
1	1	10	10	1	1	1	1	1	1	1	1	1	325.70
1	1	1	1	1	1	1	1	1	1	1	1	1	329.30
1	1	1	1	1	1	1	1	1	1	1	1	1	332.90
1	1	1	1	1	1	1	1	1	1	1	1	1	336.50
1	1	1	1	1	1	1	1	1	1	1	1	1	340.10
1	1	1	1	1	1	1	1	1	1	1	1	1	343.70
1	1	1	1	1	1	1	1	1	1	1	1	1	347.30
1	1	1	1	1	1	1	1	1	1	1	1	1	350.90
1	1	1	1	1	1	1	1	1	1	1	1	1	354.50
1	1	1	1	1	1	1	1	1	1	1	1	1	358.10
1	1	1	1	1	1	1	1	1	1	1	1	1	361.70
1	1	1	1	1	1	1	1	1	1	1	1	1	365.30
1	1	1	1	1	1	1	1	1	1	1	1	1	368.90
1	1	1	1	1	1	1	1	1	1	1	1	1	372.50
1	1	1	1	1	1	1	1	1	1	1	1	1	376.10
1	1	1	1	1	1	1	1	1	1	1	1	1	379.70
1	1	1	1	1	1	1	1	1	1	1	1	1	383.30
1	1	1	1	1	1	1	1	1	1	1	1	1	386.90
1	1	1	1	1	1	1	1	1	1	1	1	1	390.50
1	1	1	1	1	1	1	1	1	1	1	1	1	394.10
1	1	1	1	1	1	1	1	1	1	1	1	1	S WAVE 397.70
1	1	1	1	1	1	1	1	1	1	1	1	1	401.30
1	1	1	1	1	1	1	1	1	1	1	1	1	404.90
1	1	1	1	1	1	1	1	1	1	1	1	1	408.50
1	1	1	1	1	1	1	1	1	1	1	1	1	412.10
1	1	1	1	1	1	1	1	1	1	1	1	1	415.70
1	1	1	1	1	1	1	1	1	1	1	1	1	419.30
1	1	1	1	1	1	1	1	1	1	1	1	1	422.90
1	1	1	1	1	1	1	1	1	1	1	1	1	426.90

Figure VIII - 4 Some Miscellaneous Traveling Spectra

S phase determination from original record



S phase determination from traveling Spectra

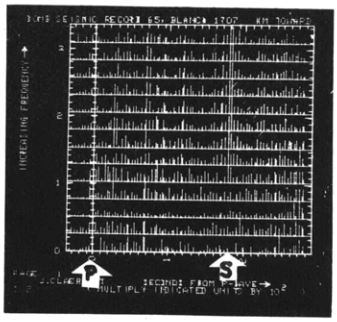


Figure VIII - 5 S wave time determination by means of traveling spectra

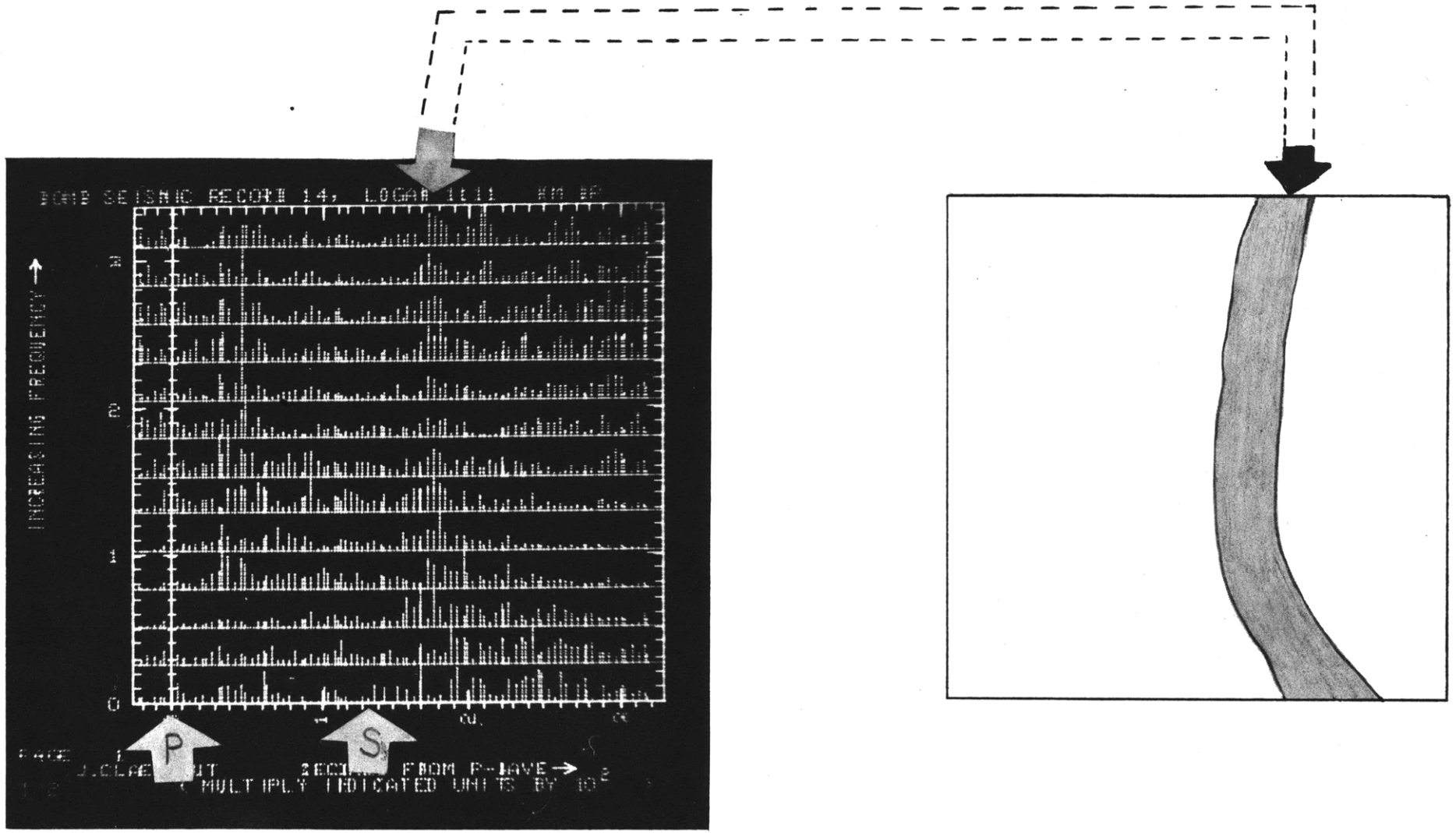


Figure VII-6 Does the shaded area represent some kind of dispersed phase ?

SECTION IX Filtering for Signal-to-Noise Improvement

Abstract

A filter is derived which can remarkably increase the signal-to-noise energy ratio on seismic records. In the examples considered the ratio was increased by factors of up to about 20. The construction of this filter is based on assumptions about signal spectra and noise spectra. The filter distortion, however, is severe and the method is not expected to be useful when applied to first motion studies. Thus the method should be useful for determining the existence of very weak arrivals. A possible application of this is in the detection of Leet's (Leet 1962) "lonesome P" phase. This application was tried but results were inconclusive due to inadequate relevant digitized data. Other experiments, perhaps less directly relevant to nuclear detection, gave excellent results.

I. Introduction

In the previous subsection we have seen filtered seismograms in which the signal-to-noise ratio was substantially enhanced. The filter in that subsection was based only upon a knowledge of the noise power spectrum. In many geophysical problems, some knowledge of the signal may reasonably be assumed. One might make the relatively weak assumption that the energy-density spectrum of the signal is known, or one could make the stronger assumption that both amplitude and phase spectrum (and thus the waveform) were known. It is advantageous to make the strongest realistic assumption possible because then the solution filters are "tailor-made" to the problem. It is dangerous, however, to make strong assumptions which are not justified, since we may not know how sensitive our solutions will be to small deviation from the assumptions. On the other hand, any sensible assumption is probably good if

the solution is not particularly sensitive to deviations from the assumption.

II. Feasibility Experiment

In the examples considered in this subsection we assume knowledge only of the noise spectrum and the signal spectrum although the method which will be applied is generally applicable to the stronger assumption of noise spectrum and signal waveform. The mathematical method is to take our assumed noise and signal spectra and construct a filter which is optimum in the Wiener sense. The details of the method are explicitly developed in Appendices B and C. The general idea is that the square error will be minimized, error being both 1) filter output when the only input is noise and 2) filter output other than signal when only signal is input. It was further assumed that at a given seismic receiver noise is present most of the time and signal is by comparison rarely present.

One might wonder how sensitive this filter is to small perturbations in the assumed signal and noise spectra. The answer is that it depends upon the spectra. This can be seen by examining Figure (IX-1) in which is displayed the spectra from one of the test cases. The filter, as might be expected, has greatest spectral components in the regions of high signal-to-noise ratio. It can be noted that high ratios at frequencies where both signal and noise have low energies do not strongly affect the filter. Thus the filter seems to have a sensible spectrum and although it is peaked rather sharply, it does not appear that any minor alteration in assumed signal and noise spectra would cause major alterations in the filter spectra.

The particular filter used in the examples tries to reproduce the signal after a 3 second delay. To facilitate comparison, however, the time scale was relabled in such

a way as to remove the delay. Distortion of the signal (caused by trying to suppress noise) now may cause precursers to the signal as soon as 3 seconds early. In fact, the filter will have considerable distortion since we have set up the problem so that the filter should suppress noise and then we have also said that noise will be the most frequent input. Thus the filter will try very hard to suppress noise, and much signal distortion will almost always result. For this reason the filter is not a good one for first motion studies.

This part of the experiment is based on the following assumptions:

1) The spectrum of a p-wave signal from a nuclear blast arriving on the LEFT-RIGHT component will be similar to that on the more clearly observable UP component.

2) The microseisms noise spectrum does not change significantly from the minute before to the minute after a p-arrival.

The results in the particular examples studied which are displayed in Figures (IX-2 to IX-6) indicate that these assumptions cannot be too bad. In them the signal spectrum was determined from the first 25 seconds of p-wave on the vertical component. The noise spectrum is computed from the horizontal component before the p-arrival time. (This time is known from the vertical component.)

III Detection Experiment

In the prediction error experiment (Section VII of this report) one of the prediction error filters increased the signal-to-noise energy ratio to such an extent that the p-wave was easily recognizable where it had previously required a good deal of imagination to recognize (see Figure VII-6). Since this phase is what Leet calls "lonesome-p" (more than 2500 kilometers distant and no observable s-waves or surface waves) and its presence may

be quite significant for nuclear detection we considered the general problem of trying to increase our ability to detect just the existence of a signal in a very high relative noise level. This led to an elaborate mathematical scheme written up in SECTION V. The final equations would be difficult to program satisfactorily using standard methods and it was felt that further theoretical study would lead to simplifications both theoretically and computationally; therefore, its use is not included.

The symmetrical Wiener-Levinson filter is quite similar in concept and in simple numerical examples gave similar numerical answers. Furthermore, one feels that the Wiener-Levinson symmetrical filter should be able to do a better job of increasing the signal-to-noise ratio than the prediction error filter because the former is derived from both signal and noise information whereas the latter is derived only from noise information.

The essential assumption in this experiment is that we have some means of getting knowledge of the lonesome-p spectrum. The various possible means of getting this knowledge represents a big study in itself. In order to proceed, we make the following assumption: the spectrum will not change radically from Logan to Blanca for similar distances and similar paths. Since Blanca was a stronger blast than Logan it was hoped that we would be able to find a distance at which the p-phase could be observed on Blanca, but not on Logan. Then we would compute spectrum of the p-phase on Blanca and the spectrum of the noise before Logan and construct a filter. This filter would then be applied to Logan in the hopes of observing p on Logan. Unfortunately, our available digitized data did not allow even this experiment. The closest approximation was Blanca 2208 km UP and Logan 2111 km UP. Unfortunately, 1) this is nearer than the distances Leet specified for

lonesome-p (2500 km on out), 2) the distances may be different enough to cause a change in the spectrum, 3) the phase is clearly evident on the Logan record even without any filtering. The best we can hope for is that we can show improvement in the signal-to-noise ratio. Unfortunately, the amount of noise digitized before the signal arrival was so small as the make unrealistic an estimate of the improvement ratio. Nevertheless, the experiment was performed and is depicted in Figure (IX-47). Better data was clearly needed.

IV Conclusion

Given noise spectra and signal spectra which are as different from each other as is typical with microseisms and p-waves, we can construct a filter which substantially improves signal-to-noise energy ratio. Because of distortion, this filter is not useful if a detailed study of the waveform is to be made.

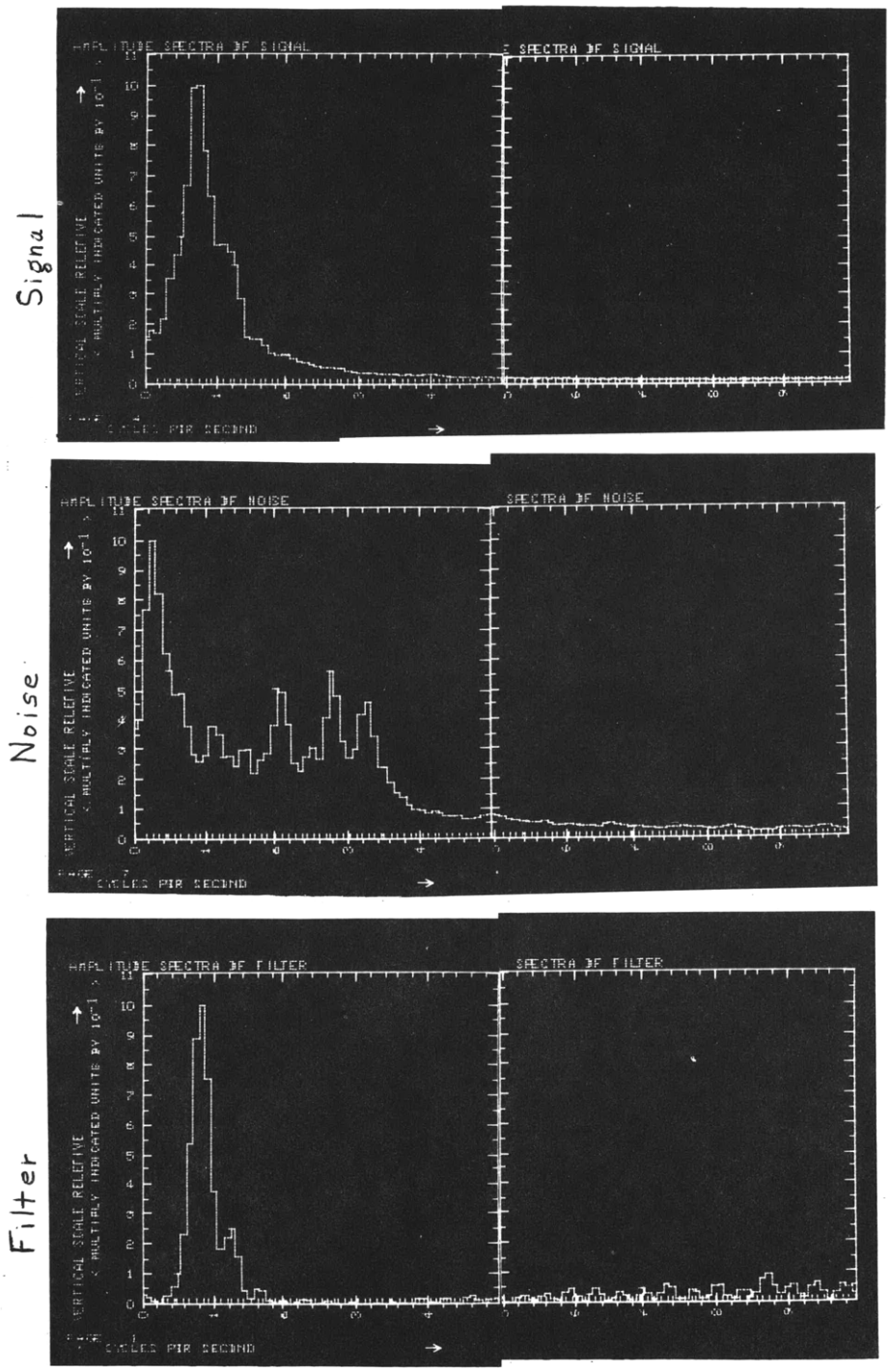
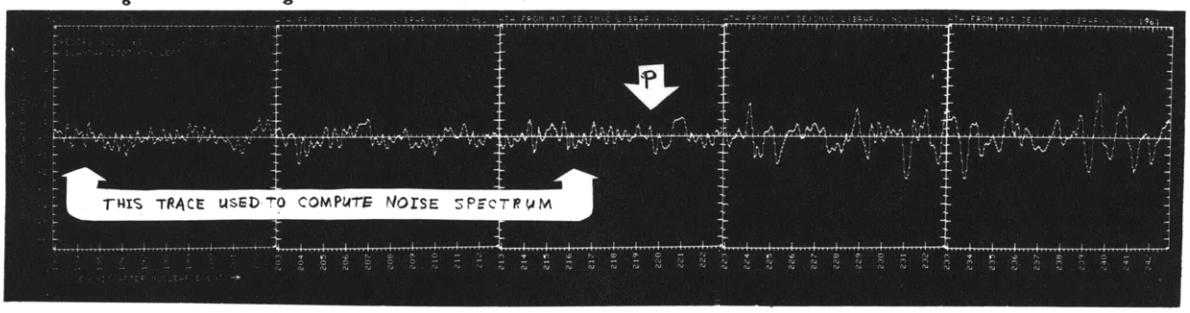
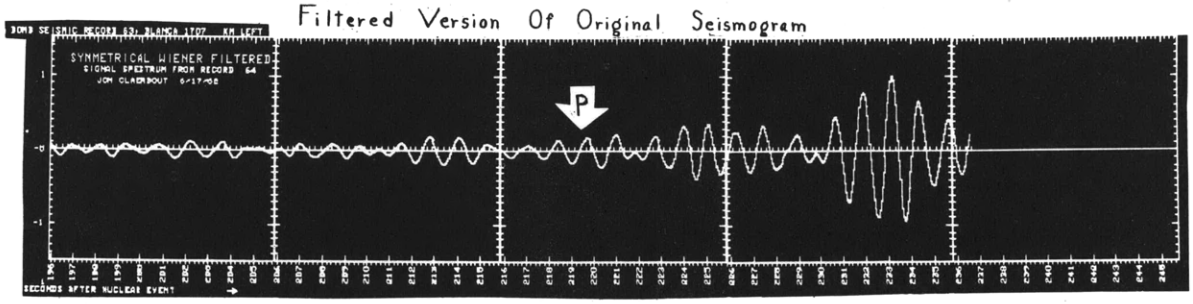


Figure IX-1 Spectra Of Signal Taken From Seismogram #64, Noise From Seismogram #63, And Spectrum Of Resulting Wiener Filter. Resulting Experiment Depicted In Figure IX-2.

Original Seismogram



Filtered Version Of Original Seismogram



Signal Or Vertical Component

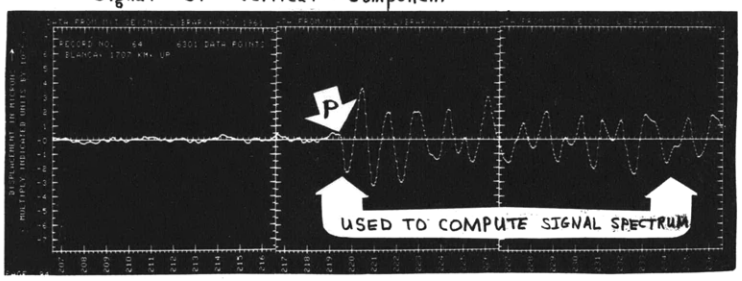
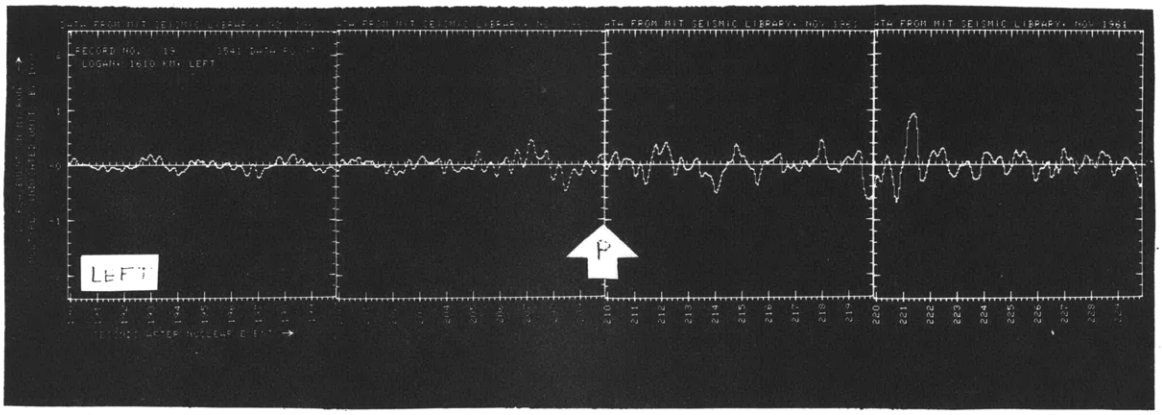
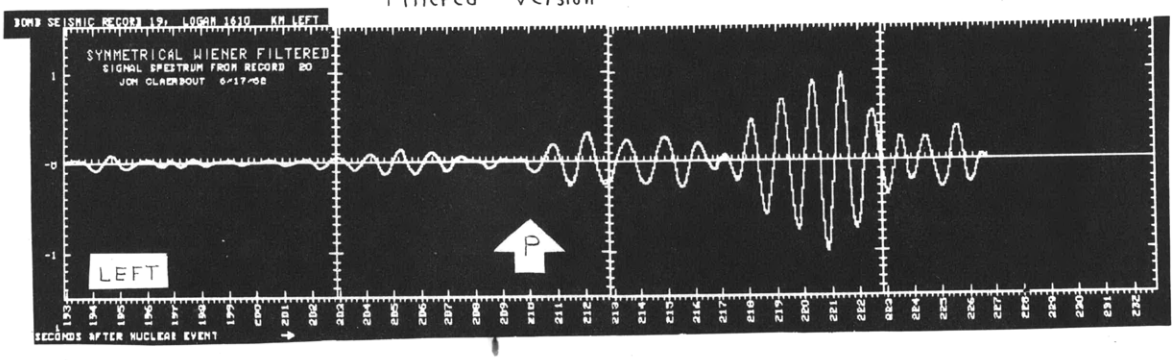


Figure IX - 2 Example Of Wiener Filtering Noise spectra is taken from before the p-wave on the left component (top) and signal spectra is taken from the higher quality vertical component (bottom). These spectra are used to generate a filter which is used on the original seismogram (top) to produce a filtered version (middle). Considerable signal-to-noise ratio improvement is noted, but the first motion is not any more apparent. The exact time of first motion is known from the vertical.

Original Seismogram



Filtered Version



UP Component

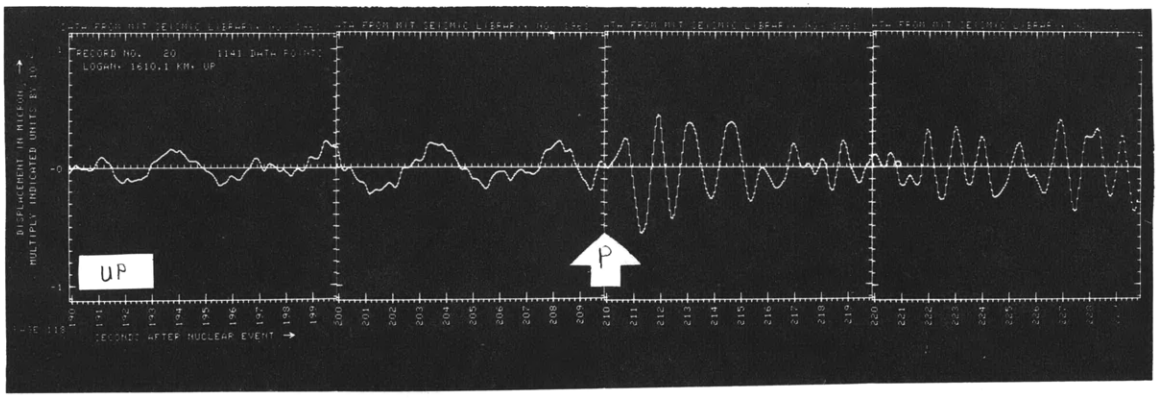
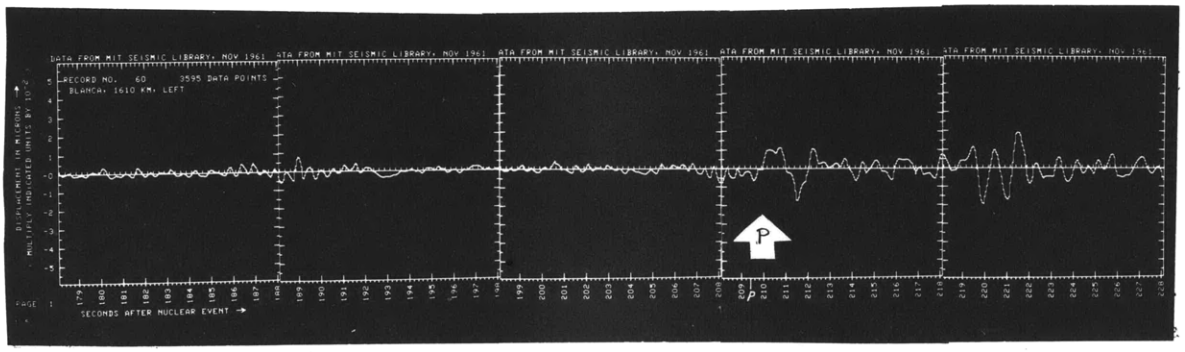
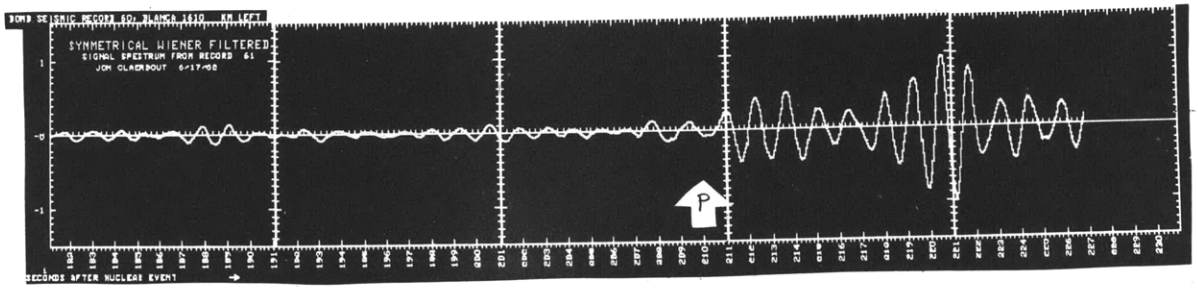


Figure IX - 3 Another Example Of Wiener Filtering

Original Seismogram



Filtered Version



UP Component

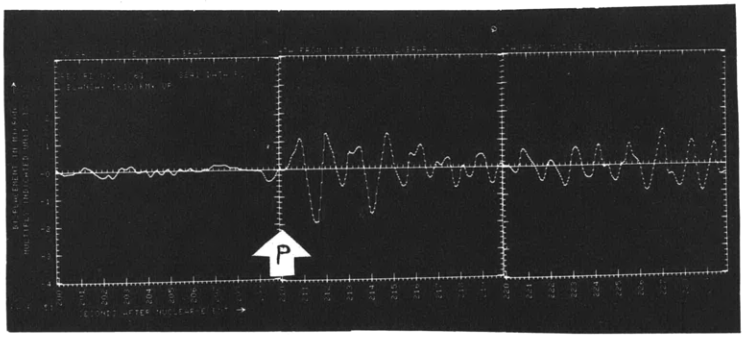
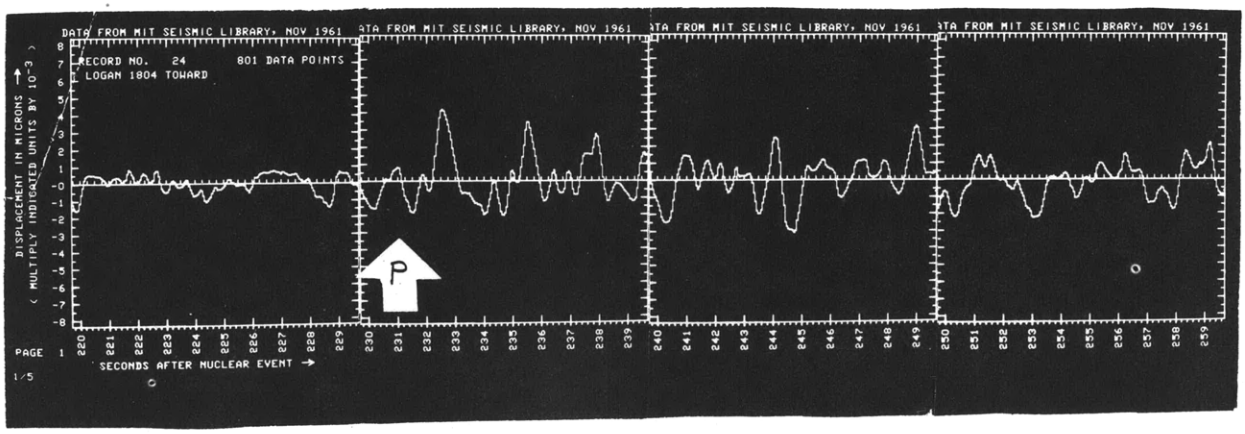
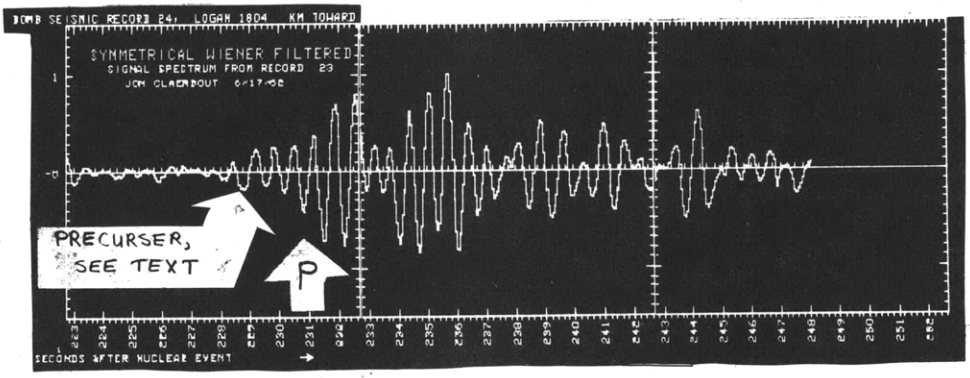


Figure IX-4 Another Example Of Wiener Filtering

Original Seismogram



Filtered Version



UP Component

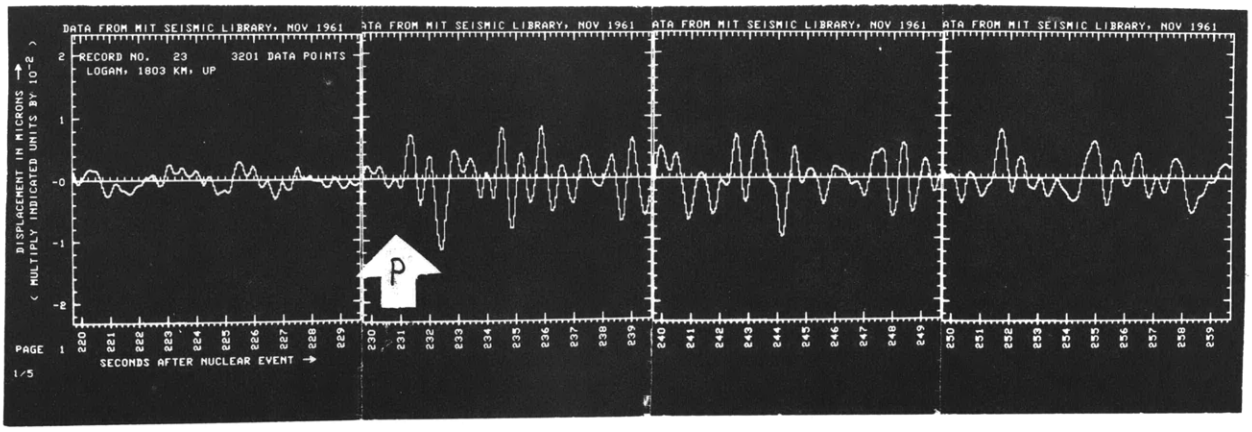
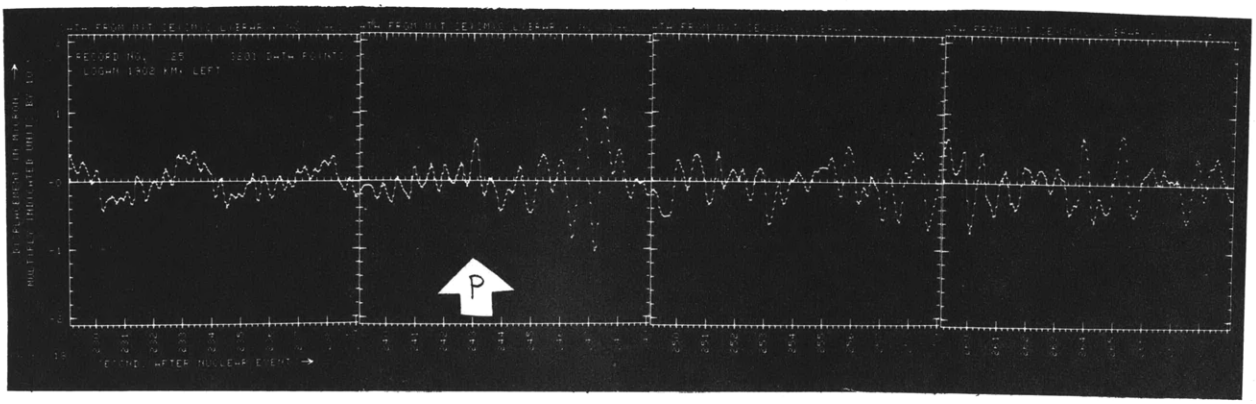
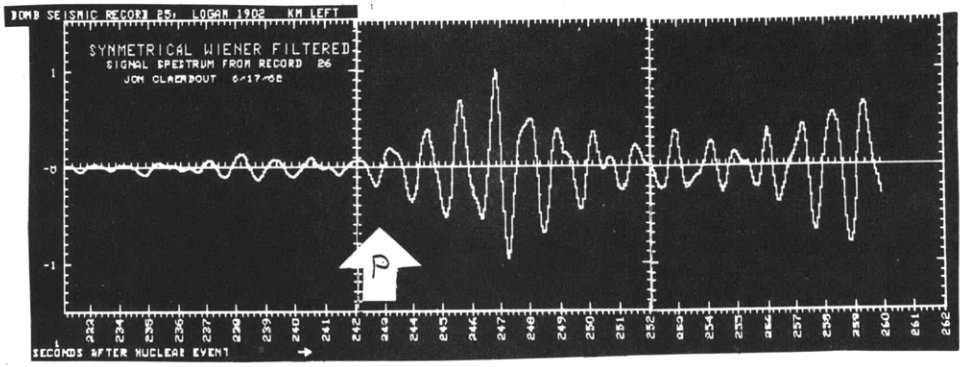


Figure IX - 5 Another Example Of Wiener Filtering

Original Seismogram



Filtered Version



UP Component

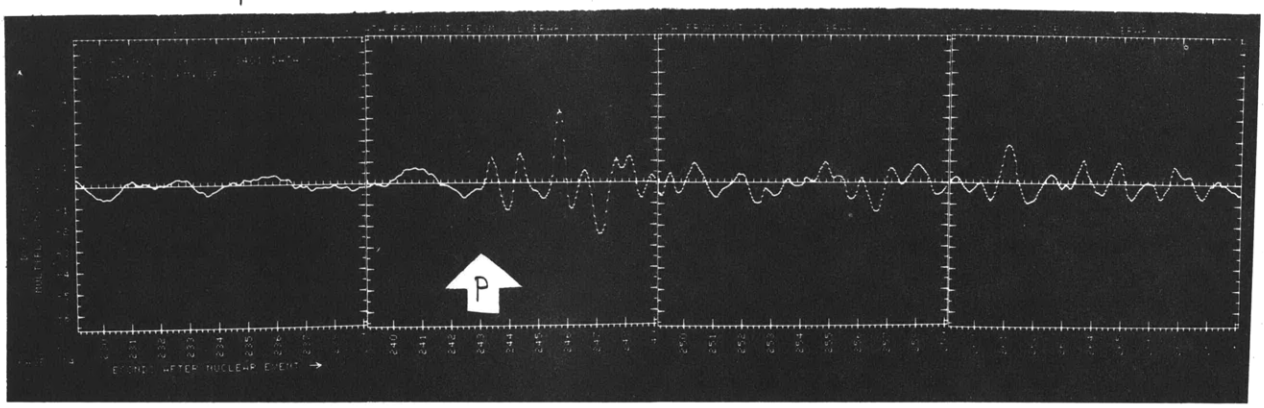
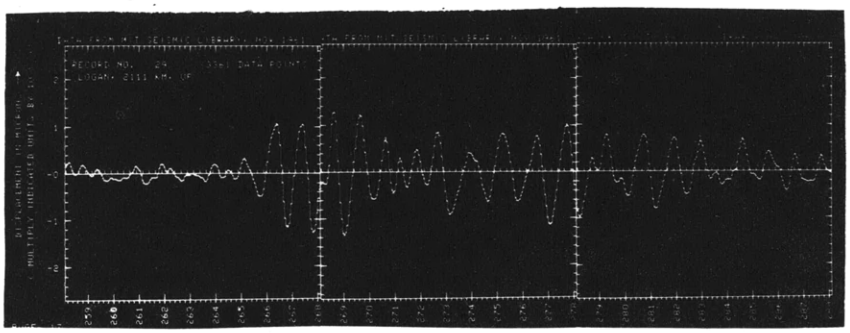
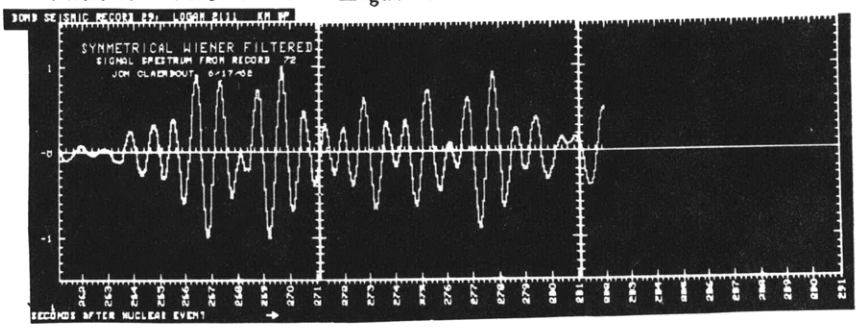


Figure IX - 6 Another Example Of Wiener Filtering

Logan 2111 UP



Filtered Version Of Logan 2111 UP



Assumed Spectrum From Blanca 2208 UP

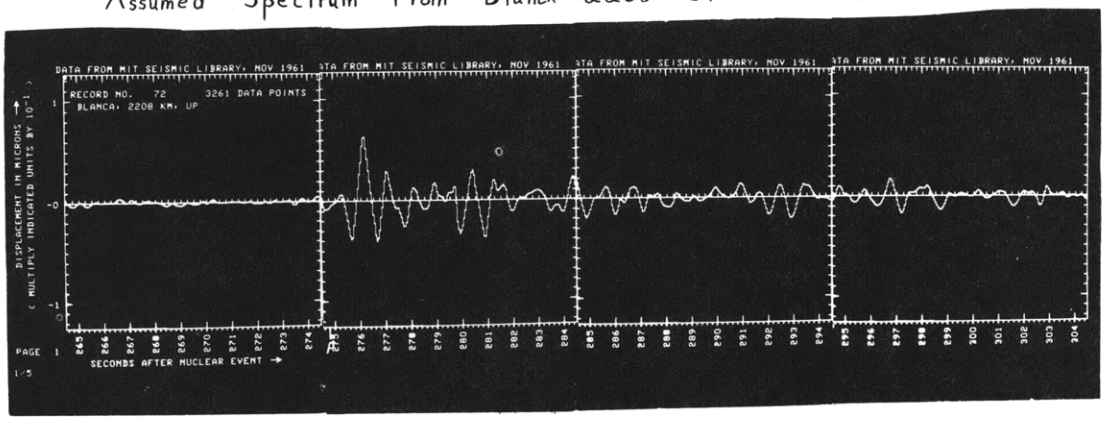


Figure IX-7 Results Of Detection Experiment.
See Text.

ACKNOWLEDGEMENT

I am indebted to Professor Stephen M. Simpson who first arroused my interest in digital filters and who made it possible for me to work on their application to both the detection of seismic events and the discrimination of natural phenomena from clandestine underground nuclear tests. I regard it as a great privilege to be associated with an enterprise which may contribute to world peace and stability.

I am indebted to Professor Theodore R. Madden for many discussions which helped me to keep a strong bond of physical intuition with mathematics.

And finally I am most indebted to Professor Enders A. Robinson for much help and encouragement. His Ph.D. thesis was my most important reference. It was he who first introduced me to least squares inverse wavelets and later drew my attention to the connection with the theory of orthogonal polynomials.

Jon F. Claerbout

BIBLIOGRAPHY

Geronimus, Ya., (1960), Polynomials Orthogonal on a Circle and Interval, Pergamon Press.

Lanczos, (1956), Applied Analysis, Prentice Hall.

Levinson, N., (1947), The Wiener RMS Error Criterion in Filter Design and Prediction, Appendix B of Wiener's book below.

Robinson, E., (1954), Predictive Decomposition of Time Series with Applications to Seismic Exploration, Ph.D. Thesis Mass. Inst. of Tech.

Simpson, et. al., (1961-62), Reports to Advanced Research Projects Agency: Project Vela Uniform.

Szego, (1939), Orthogonal Polynomials, New York.

Whittle, P., (1963), Prediction and Regulation, (in press).

Wiener, N., (1950), Extrapolation, Interpretation and Smoothing of Stationary Time Series, Technology Press.

Wold, (1938), A Study in the Analysis of Stationary Time Series, Almqvist & Wiksell: Stockholm.