FINDING NASAL ZEROS

by

Michael H. O'Malley

Memorandum No. ERL-M531

July 1975

ELECTRONICS RESEARCH LABORATORY

College of Engineering
University of California, Berkeley
94720

FINDING NASAL ZEROS[†]

by

Michael H. O'Malley
Computer Science Division
Department of Electrical Engineering and Computer Sciences
and the Electronics Research Laboratory
University of California
Berkeley, California 94720

SUR Note 185

July 1975

Linear prediction is becoming the principal technique for speech analysis. Unfortunately, for speech which contains significant spectral zeros, linear predictors mis-estimate the poles and miss the zeros. Since 25% of speech is nasal, nasalized or lateral, this defect is quite significant.

In a recent conversation, Alan Oppenheim described to me a technique for finding both the poles and zeros in a speech wave. His technique is based on a combination of homomorphic deconvolution and linear predictive analysis. The following note is the result of my comparing different versions of his technique. This note is intended to be the first part of a study of pole and zero locations in nasalized sonorants in natural speech. However, prompted by several requests for the algorithm, I am putting out this preliminary version for comments and corrections.

Oppenheim's technique is based on the observation that, if a spectrum is a rational function with both poles and zeros, differentiating the logarithm of the spectrum transforms both the poles and zeros into poles. This transformation, which is an extension of the usual homomorphic deconvolution, can then be followed by linear predictive analysis. Since

[†]Research sponsored by U.S. Army Research Office Contract DAHCO7-75-G0088.

linear prediction fits an optimal all pole model to a given spectrum, it is an appropriate tool for finding the poles of the transformed spectrum. The poles which are found with this technique should represent both the poles and zeros of the original signal. The zeros are the poles of the transformed signal which were not present in the original signal.

There are three problems with this technique. First, the original signal may not have been minimum phase and thus may have had some poles or zeros outside of the unit circle. The resultant transformed spectrum will then have poles outside the unit circle which cannot be modeled by linear predictive analysis. Secondly, the periodicity of the glottal source induces spectral zeros which are then also transformed into poles. Thirdly, the transformed spectrum has a large number of additional zeros which again cannot be modeled by linear predictive analysis. The first problem can be solved by basing the linear predictive model only on the positive time portion of the cepstrum, since that portion arises from the minimal phase part of the original signal. The second problem can be solved by the use of proper windows and by cepstral smoothing. The third problem, however, remains. This paper presents some empirical data on the effect of these extraneous zeros on the accuracy of the analysis.

Several versions of Oppenheim's technique are possible. Either the complex logarithm or the computationally simpler log magnitude may be used to compute the cepstrum. Either the full cepstrum or only the positive time portion may be used when doing the linear predictive analysis. Different amounts of cepstral smoothing may be used to remove the effects of voice periodicity and the associated spectral zeros on the spectrum. The following two sections present an analysis of one version of the technique and compare the resulting spectrum to an idealized all pole model. The effects of different versions of the technique are then presented along with an algorithm for performing the analysis.

Complex Logarithmic Analysis

Let $x(t)$ represent a steady state voiced speech signal of period T_0 and duration T such as might arise from a nasalized vowel. If $g(t)$ is the glottal waveform, $h(t)$ is the impulse response of the vocal tract, $r(t)$ is the impulse response of the radiation from the mouth, $p(t)$ is a train of pulses T_0 seconds apart, and $w(t)$ is an analysis window of width T, then

$$x(t) = \left(g(t)*h(t)*r(t)*p(t)\right)w(t) .$$

If $x(t)$ is now sampled every T_s seconds, its z-transform is

$$X(z) = \left(G(z)H(z)R(z)P(z)\right)*W(z) .$$

If the window $w(t)$ is smooth enough and wide enough, its effect on G, H, and R can be ignored. Thus

$$X(z) \cong G(z)H(z)R(z)\left(P(z)*W(z)\right) .$$

The derivative of the logarithm of the spectrum is

$$\frac{d}{dz}\log(X(z)) = \frac{d}{dz}\left(\log(Y(z)) + \log(P(z)*W(z))\right) ,$$

where

$$Y(z) = G(z)H(z)R(z) .$$

As is well known, the effect of the last term may be removed by low pass "filtering" in the cepstral[1] domain -- so-called cepstral smoothing. The differentiation may also be done in the cepstral domain by multiplying the cepstrum by the "quefrency" n.

In general, G, H, and R are rational functions although G is not usually minimum phase. Thus $Y(z)$ is the ratio of polynomials $P(z)$ and $Q(z)$. If $\hat{X}_c(z)$ is the cepstrally smoothed and differentiated logarithmic spectrum, then from the above analysis

$$\hat{X}_c(z) = \left(\log Y(z)\right)' = Y'(z)/Y(z) = \frac{\left(P(z)/Q(z)\right)'}{\left(P(z)/Q(z)\right)}$$
$$= \frac{P'(z)}{P(z)} - \frac{Q'(z)}{Q(z)} = \frac{Q(z)P'(z) - P(z)Q'(z)}{P(z)Q(z)}$$

where prime represents differentiation with respect to z. $\hat{X}_c(z)$ is thus a rational function which has poles at both the poles p_i and at the zeros z_i of $Y(z)$; the zeros of the speech wave have been transformed into poles.

Unfortunately, new zeros have been added at the points where

$$Q(z)P'(z) - P(z)Q'(z) = 0 .$$

If there are n poles and m zeros in the original speech wave then there will be up to $n(m-1)$ or $m(n-1)$ added zeros.

Comparison With an All Pole Spectrum

Since linear prediction finds an all pole, minimum phase filter which minimizes the energy ratio, it is instructive to compare $\hat{X}_c(z)$

[1]The sampled data cepstrum is defined as the inverse digital Fourier transform (DFT) of the logarithm of the DFT of a signal. The DFT is the z-transform evaluated at equal intervals on the unit circle. $\log(u) = \ln|u| + i(\theta + 2\pi k)$ for all k.

to an idealized all pole spectrum $\hat{X}_p(z)$

$$\hat{X}_p(z) = \frac{1}{P(z)Q(z)} .$$

From the preceding analysis, it is easy to see that

$$\frac{\hat{X}_c(z)}{\hat{X}_p(z)} = P(z)Q'(z) - Q(z)P'(z) .$$

This function is rather complex but insight may be gained by considering a simple pair of complex conjugate poles at p_1 and p_1^*. In this case, it can be shown that $\hat{X}_c(z)$ has an added zero at $e^{\sigma_1 T_s}$ where $|p_1| = e^{\sigma_1 T_S}$.

For the case of an all pole, narrow bandwidth spectrum (a vowel-like sound), there will be an added zero approximately centered between each pair of poles. On the other hand, poles due to zeros in the original spectrum, when between the other poles, reduce the number of additional zeros. Further experimentation is necessary to see how these additional zeros affect linear predictive analysis in the case of natural speech.

Minimum Phase Signals

In Oppenheim and Schafer (1975) it is shown that, if $x(t)$ is the convolution of a minimum phase and a maximum phase signal,

$$x(t) = x_{min}(t) * x_{max}(t) ,$$

then the complex cepstrum $\hat{x}(n)$ is the sum of two components. The positive time component is due to the minimum phase signal and the negative time component is due to the maximum phase signal. By ignoring the negative time cepstrum (or, in the case of sampled data signals, the second half of the cepstral array), an analysis can be based on the minimum phase signal alone. Differentiating the logarithm of the spectrum of a minimum phase signal does not generate negative time components in the cepstrum although it will increase aliasing in sampled data signals. Thus $\hat{x}_c(n)$ for $n > 0$ (the positive time component of the smoothed, differentiated logarithmic spectrum) comes from the minimum phase component of the original signal. Furthermore, the poles of $\hat{x}_c(n)$ for $n > 0$ are inside the unit circle as are the zeros. Thus $\hat{x}_c(n)$ for $n > 0$ is, itself, minimum phase.

Log Magnitude Analysis

Because of the computational efficiency, it is tempting to use the log magnitude rather than the complex logarithm. Unfortunately, the derivative of the log magnitude spectrum is zero at the inflection points of the original spectrum. This has the effect of putting a zero in front of each pole and zero of the original spectrum. However, the complex logarithm of the spectrum of the autocorrelation function $\phi_{xx}(t)$ is real and so, except for a factor of two, is equal to the log magnitude spectrum of $x(t)$. Since the autocorrelation is the convolution of a minimum and maximum phase component, each with the same spectral magnitude as $x(t)$, the positive time log magnitude cepstrum is equivalent to the complex cepstrum of a minimum phase signal which has the same spectrum as the original signal $x(t)$. Thus if $x(t)$ is minimal phase, the positive time log magnitude cepstrum is equal to the complex cepstrum. In any case, if an analysis which matches the spectral magnitude of $x(t)$ is adequate, then the log magnitude may be used. Since all the transforms are then real and since the phase angle need not be found, the computation should be from two to four times faster.

Cepstral Domain Processing

The first step in the analysis procedure is to get the cepstrum $\hat{x}(n)$ from the original signal $x(t)$. If there are N values in the cepstral array, then removing the "negative time" component may be done by letting $\hat{x}(n) = 0$ for $n > \frac{N}{2}$ (remember that $\hat{x}(n)$ is periodic). Cepstral smoothing then consists of letting

$$\hat{x}(n) = 0 \quad \text{for} \quad n > \frac{4 \text{ msec}}{T_s} .$$

The z-transform of $n\hat{x}(n)$ is $-z\hat{X}'(z)$ or

$$-(n-1)\hat{x}(n-1) \Leftrightarrow \hat{X}'(z) \quad \text{for} \quad n > 0 .$$

Thus the differentiation may be done by multiplying the cepstrum by $-n$ and shifting. Combining these three operations gives $\hat{x}_c(n)$ from the normal cepstrum $\hat{x}(n)$.

Separating the Original Poles and Zeros

It is necessary to distinguish those poles which are due to formants in the speech signal from those which are due to spectral zeros. One way to perform this separation is to compute the residue of each pole of $\hat{X}_c(z)$. The partial fraction expansion of $\hat{X}_c(z)$ is easily seen to be:

$$\hat{X}_c(z) = \frac{P'(z)}{P(z)} - \frac{Q'(z)}{Q(z)} = \sum_{i=1}^{m} \frac{(z-z_i)'}{z-z_i} - \sum_{i=1}^{n} \frac{(z-p_i)'}{z-p_i} = \sum_{i=1}^{m} \frac{1}{z-z_i} + \sum_{i=1}^{n} \frac{-1}{z-p_i}$$

where z_i is one of the m zeros and p_i is one of the n poles of

$Y(z)$. Thus the residue of $\hat{X}_c(z)$ is 1 at the original zeros and -1 at the original poles.

The Algorithm

The following algorithm may be used to find nasal zeros:

I. Find the cepstrum.

- A. $x(t)$ is two to three periods of a speech waveform.
- B. Window $x(t)$ and add zero samples to prevent aliasing. The window should be narrow enough to remove some of the harmonic structure.
- C. Compute the log magnitude spectrum $\log|X(z)| = \hat{X}(z)$. The complex logarithm is better although more difficult.
- D. Compute the inverse DFT $\hat{x}(n)$. Since $\hat{X}(z)$ and $\hat{x}(n)$ are real and since only the first few values of $\hat{x}(n)$ are needed, considerable savings can be made.

II. Do cepstral domain processing to get $\hat{x}_c(n)$ for $n > 0$.

- A. $\hat{x}_c(n) = -(n-1)\hat{x}(n-1)$ for $0 < n \leq 40$

  $= 0$ otherwise.

III. Do linear predictive analysis.

- A. Solve the linear predictive equations. Since there are additional poles due to the nasal zeros, more coefficients will be needed.

IV. Find the poles.

- A. Use a polynomial root solver or find the peaks in the inverse filter spectrum.

V. Find the nasal zeros.

- A. Compute the residue or ignore those poles which were also found by the usual formant tracking techniques.

Computational Results

The algorithm was tested using synthetic speech sampled at 10 kHz. It was found that, for a 120 Hz fundamental, a 20 msec window in the time domain and a 4 msec window in the cepstral domain gave sufficient smoothing to remove the effects of the excitation function. 31.2 msec of silence was added to prevent aliasing.

Figure 1 is the cepstrally smoothed, log magnitude spectrum for a signal with formants at 300, 1000, 1700, 2700, 3500, and 4500 Hz and with a zero at 600 Hz. Figure 2 compares this spectrum to the squared magnitude of the derivative of the complex log spectrum $|\hat{X}_c(z)|^2$ for the signal. Except for the first formant which it missed, most of the peaks are quite close to the correct location. However, in other speech samples, some of the peaks were missing. The errors seem to be due to the added zeros in the transformed signal and to poles and zeros of $G(z)$ and $R(z)$. Perhaps an analysis of the minimum phase component of the signal rather than of a minimum phase signal with the equivalent spectrum would have produced better results. The next section presents a different technique which was discovered in the process of testing Oppenheim's analysis.

All-Pass Equivalent Analysis

For a minimum phase signal, poles and zeros near the unit circle may be identified by their effect on the phase of the spectrum. Figure 3 compares the smoothed derivative of the phase of the minimum phase equivalent of $X(t)$ with the log magnitude spectrum $\log|X(z)|$. As can be seen, this phase curve is better for finding poles and zeros than is $\hat{X}_c(z)$.

To demonstrate this, let $x(t)$ be a signal with phase $\phi(\omega)$ and magnitude $\gamma(\omega)$. In the vicinity of a pole at $s_1 = \sigma_1 + i\omega_1$

$$r(\omega) \cong \frac{1}{(1+u^2)^{1/2}}$$

$$\phi(\omega) \cong \tan^{-1} u$$

$$\phi'(\omega) \cong \frac{1}{1+u^2}$$

where $u = \frac{\omega-\omega_1}{\sigma_1}$, $\sigma_1 < 0$. Thus the peaks in $\phi'(\omega)$ are narrower than the peaks in $r(\omega)$ and so the poles and zeros are easier to find.

It turns out that a very simple modification of the zero finding algorithm enables us to make use of the properties of $\phi'(\omega)$. As mentioned earlier the imaginary component of the complex logarithm is the phase, $\phi(\omega)$, of the signal, $x(t)$.

$$\phi(\omega) = \mathrm{Im}(\ln(X(z)))\Big|_{z=e^{-i\omega T_s}} .$$

Now

$$\frac{d}{d\omega}\ln(X(z)) = \frac{dz}{d\omega}\ln'(X(z))$$

$$= -iT_s z \ln'(X(z))\Big|_{z=e^{-i\omega T_s}} .$$

Thus

$$\phi'(\omega) = \frac{d}{d\omega}\mathrm{Im}(\ln(X(z))) = \mathrm{Im}\frac{d}{d\omega}\ln(X(z)) = -T_s \mathrm{Re}(z \ln'(X(z)))\Big|_{z=e^{-i\omega T_s}} .$$

Since $n\hat{x}(n) \Leftrightarrow -z \ln'(X(z))$,

$\frac{\phi'(\omega)}{T_s}$ is the real part of the spectrum of $n\hat{x}(n)$.

To be useful for finding poles and zeros, $\phi'(\omega)$ must be derived from an appropriate minimum phase signal. Fortunately, the positive time cepstrum, $\hat{x}(n)$, $0 < n < \frac{.004}{T_s}$, comes from the smoothed minimum phase component of $x(t)$. Thus the real part of the transform of $T_s n\hat{x}(n)$ for $0 < n < \frac{.004}{T_s}$ is equal to the desired $\phi'(\omega)$.

It is interesting to consider the case in which $r(\omega)$ is constant. Under this condition, $\text{Im}(z \ln'(X(z))) = 0$ and $\phi'(\omega) \Leftrightarrow T_s n\hat{x}(n)$. Thus $n\hat{x}(n)$ is symmetric and so $\hat{x}(n)$ is antisymmetric. A signal which meets this condition will be called all-pass since it is the impulse response of an all-pass filter.

Given the cepstrum, $\hat{x}(n)$ of any signal $x(t)$, the all-pass equivalent cepstrum (APEC), $\hat{x}_a(n)$ may be defined as twice the antisymmetric component of $\hat{x}(n)$.

$$\begin{aligned} \hat{x}_a(n) &= \hat{x}(n) \quad , \quad n > 0 \\ &= -\hat{x}(-n) \quad , \quad n < 0 \,. \end{aligned}$$

Thus $\hat{x}_a(n)$ is the cepstrum of an all-pass signal whose phase is $2\phi(\omega)$. The spectrum of $n\hat{x}_a(n)$ is $2\phi'(\omega)/T_s$.

Consider now the case where $\hat{x}(n)$ is derived from a minimum phase signal $x_{min}(t)$, where $\phi(\omega)$ is the phase of that signal. The pole and zero finding algorithm may be modified so that

$$\hat{X}_c(z) = \left. \frac{2\phi'(\omega)}{T_s} \right|_{z=e^{-i\omega T_s}}$$

by replacing step IIA with

$$\text{IIA}' \quad \begin{aligned} \hat{x}_c(n) &= n\hat{x}(n) && \text{for} \quad 0 < n \le 40 \\ &= -n\hat{x}(-n) && \text{for} \quad -40 \le n < 0 \\ &= 0 && \text{otherwise} \end{aligned}$$

The effect of this change is to perform the analysis on the all-pass equivalent $x_a(t)$, a signal with additional RHP zeros and poles corresponding to its LHP poles and zeros. As usual, $\hat{X}_c(z)$ has poles at both the poles and zeros of $x_a(t)$. Since $\hat{x}_c(n)$ now has poles in the LHP, problems may develop with the LPC. However, the original poles and zeros can usually be identified quite successfully by peak and valley picking.

In order to make the preceeding analysis somewhat clearer, let us consider an example of the s-plane pole and zero locations for various functions. Assuming that $x(t)$ is mixed phase with 2 pole and 2 zero pairs, Figure 4 shows the poles and zeros for $\phi_{xx}(t)$, $x_{min}(t)$, $x_a(t)$ and $\hat{x}_c(n)$ as found using log magnitude analysis. Figure 5 shows the effect of using the complex logarithm to analyze the minimum phase component directly. The zero locations of $\hat{x}_c(n)$ have not been verified.

Summary

Two techniques have been presented in an attempt to solve some problems which arise in the detailed analysis of speech signals. The problems are that, first, the presence of zeros distorts the measured locations of spectral poles and, secondly, that the spectral zeros, which are of interest in their own right, cannot be located. By using homomorphic deconvolution to obtain an appropriate signal and then using linear prediction to analyze it, these problems may be partially overcome. A technique based on an all-pass equivalent signal was also demonstrated. This technique was shown to be equivalent to an analysis of the slope of the phase of the cepstrally smoothed, minimum phase component of the signal.

Reference

Oppenheim, Alan and Schafer, Ronald (1975). Digital Signal Processing. Prentice Hall.

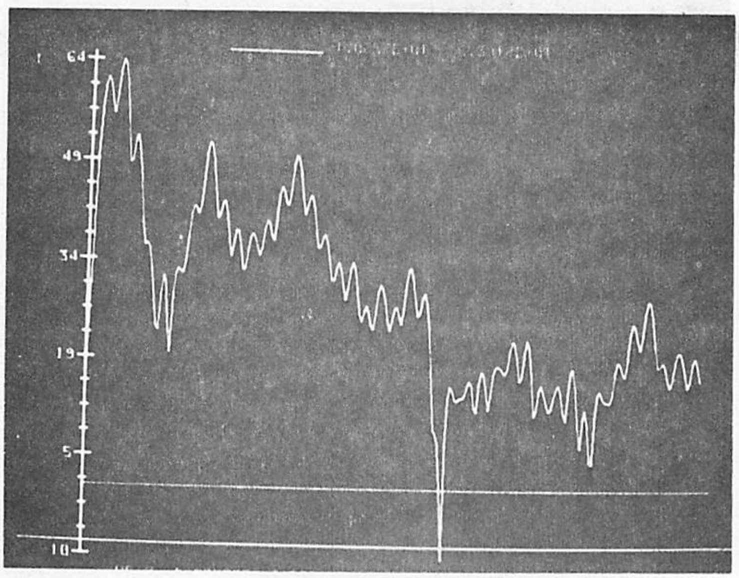

FIG 1a LOG MAGNITUDE SPECTRUM X(ω)

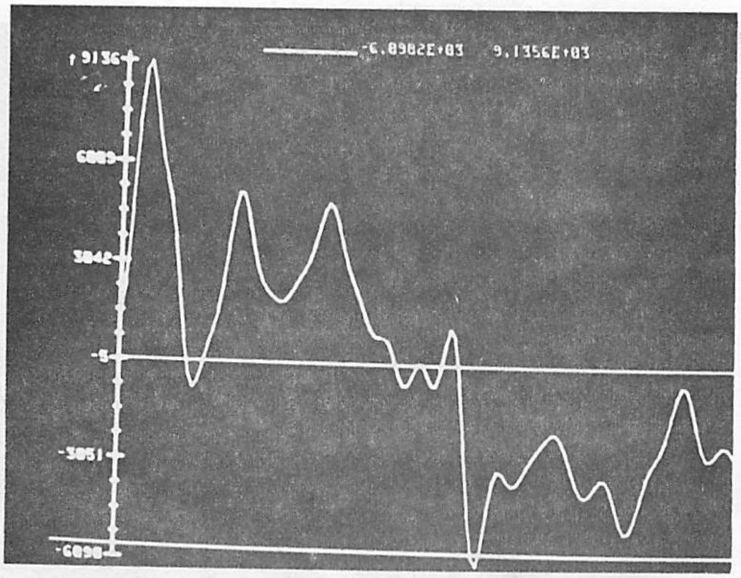

FIG 1b CEPSTRALLY SMOOTHED LOG SPECTRUM

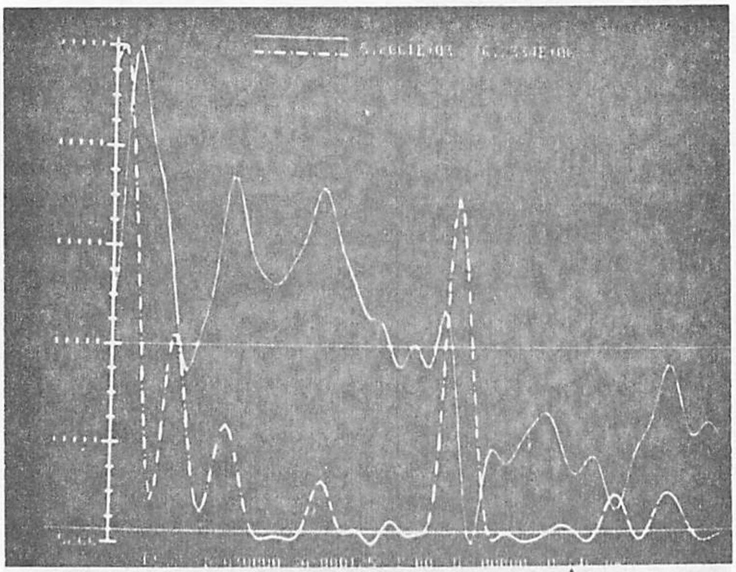

FIG 2a SMOOTH LOG SPECTRUM vs. $|\hat{X}_c(\omega)|^2$

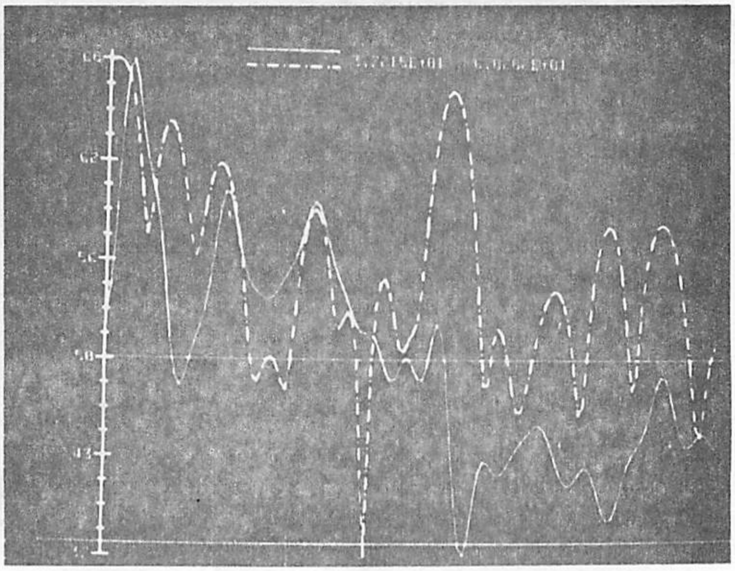

FIG 2b SMOOTH LOG SPECTRUM vs. $\log|\hat{X}_c(\omega)|$

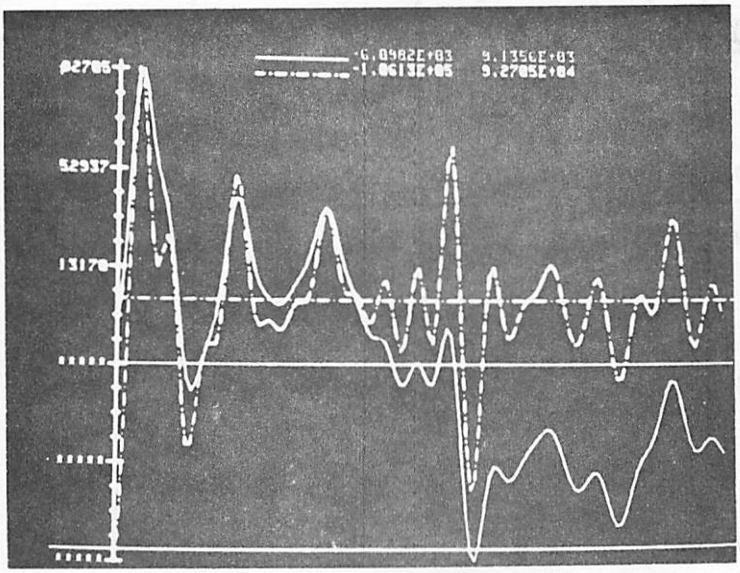

FIG 3 SMOOTH LOG SPECTRUM <u>vs</u>. $\hat{X}_c$(co) - APEC ANALYSIS

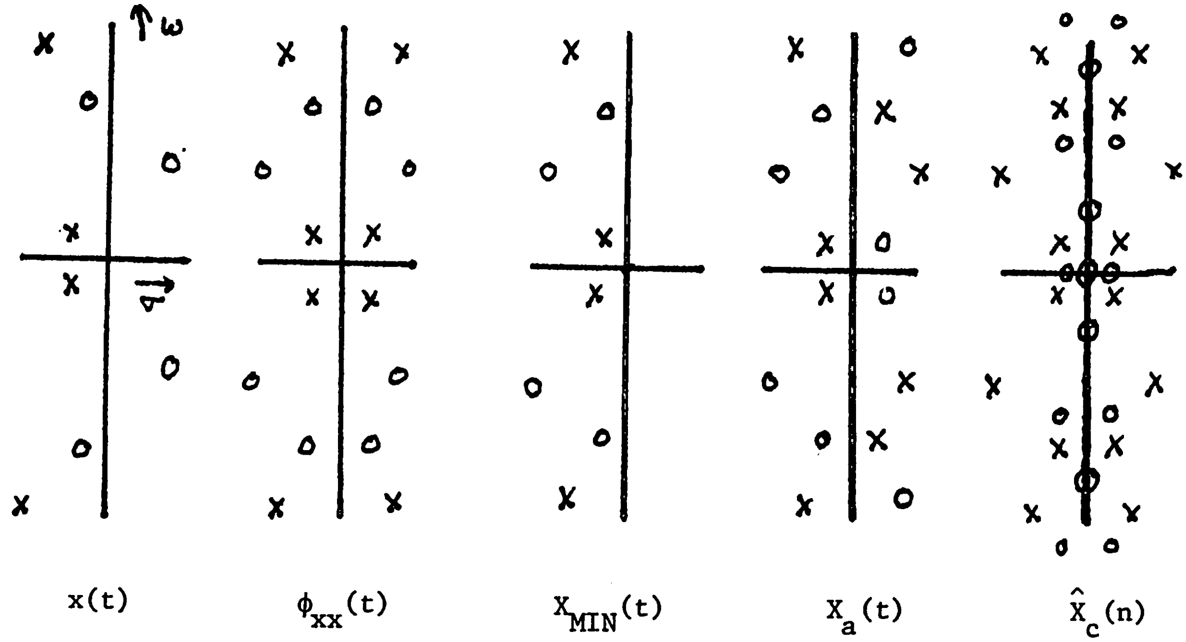

FIG 4 APEC ANALYSIS - TYPICAL POLE AND ZERO LOCATIONS

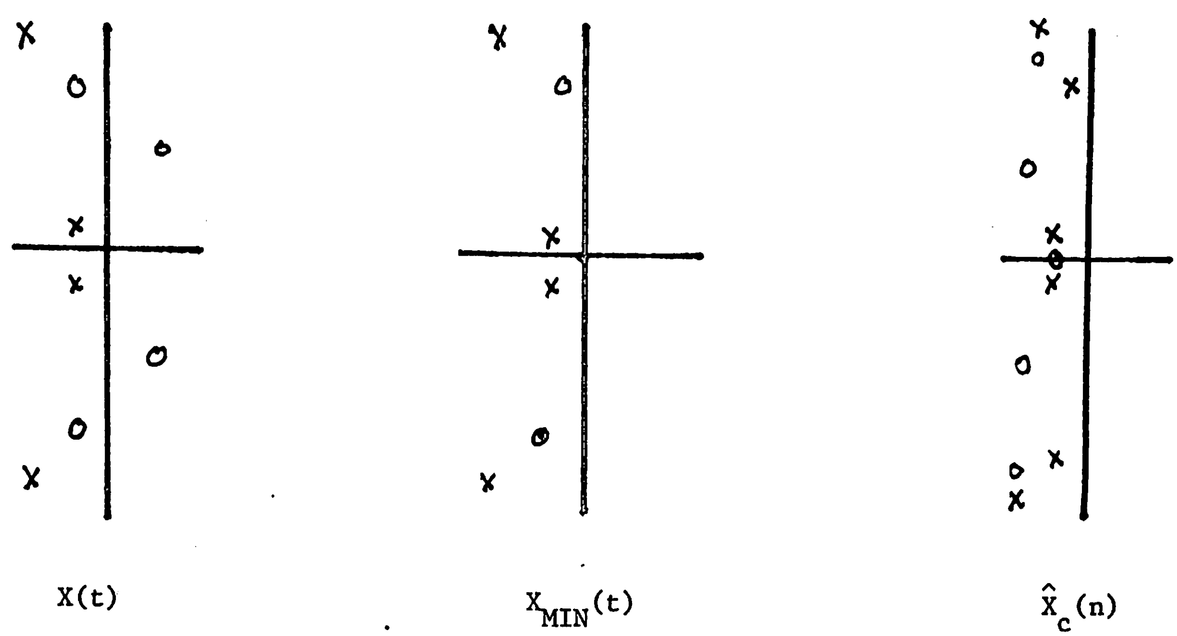

FIG 5 MINIMUM PHASE ANALYSIS - TYPICAL POLE AND ZERO LOCATIONS