Speech Research

A. Despain
G. MacDonald
O. Rothaus

May 1984

JSR-82-601

JASON
The MITRE Corporation
1820 Dolley Madison Boulevard
McLean, Virginia 22102

TABLE OF CONTENTS

SPEECH RESEARCH

1.0 INTRODUCTION

A small JASON study group--Despain, MacDonald, and Rothaus--picked up the cudgels again in the summer of 1982 and had another "go" at some speech problems. The problems they addressed, were word spotting in continuous speech, speaker recognition, speech intelligibility, and automatic discrimination between a pair of speakers.

All of these appear to be very difficult problems indeed, yet representative of typical problems in the field. Accordingly, we decided to view them as paradigms and so sought to devise, or at least have a look at, highly novel or untried methodologies which might prove useful here or elsewhere.

In the course of this report, we will describe several of the ideas we have come up with.

This report falls naturally into two parts. ~~The first of these~~ (Part I) contains some peripheral remarks on the use of "correct" statistics in speech modeling, introduces a fairly general setting for non-linear modeling of speech wave forms, together with some considerations relevant to determining the parameters of the non-linear modeling. The major thrust of this portion, however, is the introduction and

description of many of the properties of the bispectrum of time signals. While related to the ordinary spectrum, the bispectrum captures many additional features of the signal. Bispectral analysis has already proven useful in the analysis of water waves and seismic signals; there is every reason to hope that it will be a valuable tool in speech research.

Part two introduces a novel procedure for estimating the LPC coefficients of a waveform. It is based on the observation that a given filter transfer function has many "realizations" in auto-regressive modeling, each of which suggest different methods of finding the coefficients. It serves, for one thing, to explain fully why running the waveform backwards gives essentially the same coefficients as running it forwards. (KR)

There was some additional speech related activity; due principally to W. Press, with an appendix by O. Rothaus, but this material is being published separately.

During the course of our summer study, we had the opportunity to talk profitably with visitors, two from CRD: Lee Neuwirth and Allen Poritz, and one from NSA: Edward P. Neuburg. We thank them all for their useful comments.

PART I

2.0 NON-LINEAR SPEECH MODELING AND BISPECTRAL ANALYSIS

2.1 Introduction

The mathematical modeling of speech for such applications as word or speaker recognition has been intensively studied over the past twenty years. The most intense effort was undoubtedly the speech program sponsored by ARPA in the 1970s which was terminated in 1977 with the claim of success. Five years later it is clear that the methods developed fail when applied to such problems as automated word spotting on a noisy channel with uncooperative speakers. Another, perhaps even simpler problem where available techniques also fail, is in automatically discriminating between two speakers on a noisy channel with the speakers having different voice characteristics. The problem is further simplified in that at any one time one and only one speaker is speaking. What is required is a time series marking the times when the speakers change.

Current methods are successful when the speech is highly constrained; the vocabulary is limited to a few hundred words, the speakers voice characteristics are known, the semantics are well specified and the environment is characterized by low noise levels. There appears to be a number of reasons why current methods fail when applied to the kinds of problems described in the above paragraph.

1. The signal processing techniques are almost entirely linear up to the digital decision, e.g., whether a format was present or not. Yet it is well known that the human mechanisms for producing speech contain non-linear elements(1) and the literature on the psychophysics of detection of beats establishes that within the human auditory system there are some processes that imply envelope detection.

2. The current speech algorithms are almost all based on using minimum least squares as a measure of goodness of fit.(2) It is well known that the use of such a norm puts a heavy emphasis on "outliers" and increases the vulnerability of the algorithms to distortion by noise. We do not know of any exploratory effort to use other norms, for example, absolute mean deviations, nor to adaptively alter measures of goodness in response to changing signal to noise conditions.

3. Spectral methods used in speech analysis are those that have been widely used and whose deficiencies have only recently been recognized. In these methods, in order to obtain a reliable estimate of the power spectrum where noise is present, there is a smoothing of the autocovariance function by a time domain window before its Fourier transform is taken or equivalently the squared magnitude of the Fourier transform is smoothed. These methods have two grave deficiencies. The smoothing alters the physically measured quantities; the data is deliberately distorted for mathematical simplicity. Secondly, there is the implied assumption that the autocovariance vanishes outside the time period which is being analyzed. In the maximum entropy spectral

analysis (MESA), the estimator uses all lags without smoothing by assuming maximum uncertainty about the time series outside the period under analysis.(3) In the Maximum Likelihood Method of spectral analysis the window is computed at every frequency in such a way that it adapts itself to the signal plus noise under analysis.(4) We do not know of the exploitation of these techniques in the analysis of speech. This is surprising in that they are optimal estimators for autoregressive (all pole) time series while the classical techniques are more appropriate for moving average (all zeros) time series.

4. It is universally assumed that the statistics of speech can be described using parametric statistics. Indeed, it is commonly assumed that the statistics of speech are Gaussian. The highly structured nature of speech makes these assumptions highly questionable. In speech as in many other areas of statistical analysis it is assumed that a minor error in the mathematical model causes only a small error in the final results. Unfortunately this does not always hold true. During the past two decades it has become increasingly clear that common statistical procedures, particularly those optimized for an underlying Gaussian distribution, are exceedingly sensitive to seemingly minor deviations from the assumption.(5)* These comments are closely allied to comment 2 above. We know of no application of non-parametric statistics or robust methods to the analysis of speech.

---

* A geophysicist, Harold Jeffreys, of Bayesian fame, was well aware of these considerations and had a running battle with the most famous of all statisticians, Ronald Fischer, on this point in the 1920s and 1930s. Today's statisticians are slowly rediscovering Jeffrey's results.

The above points in fact are the outline of what could be an extensive speech research program. In this note, we consider only the first point by examining one method for analyzing the non-linear characteristics of speech.

3.0 REPRESENTATION OF NON-LINEAR SYSTEMS

A possible mathematical model for speech is one in which an input signal $x(t)$ is operated on by a "black box" which forms $x(t)$ into the actual speech $y(t)$. For the purposes of the model the black box is assumed to be time invariant (stationary) and the input $x(t)$ has a finite power

$$\frac{1}{2T}\int_{-T}^{T} x^2 dt < \infty \tag{3.1}$$

where $2T$ is the duration of the input. A general functional representation of the output $y(t)$ can be written as

$$y(t) = \int h_1(\tau)\, x(t-\tau)d\tau + \iint h_2(\tau_1,\tau_2)\, x(t-\tau_1)\, x(t-\tau_2)d\tau_1 d\tau_2$$

$$+ \ldots + \int \ldots \int h_n(\tau_1, \ldots \tau_n) x(t-\tau_1) \ldots x(t-\tau_n)d\tau_1 \ldots d\tau_n$$

$$+ \ldots \quad . \tag{3.2}$$

Such a representation was studied in detail by Volterra[6] in the early 20th Century and extended to statistical problems by Weiner.[7]

We need not concern ourselves with the exact domain of validity of such an expansion, though the alert reader will recognize its

relation to the Wierstrass Approximation Theorem. For the purposes of the present treatment; the functions h_i in Equation (3.2) can be considered generalized functions, and contain delta functions and derivatives thereof.[8] With this generality available, there is no point letting the black box depend on the derivatives of the input $x(t)$, since this can now be absorbed into the kernels h_i.

Even more critically, if we are working perforce with discrete inputs and outputs, derivatives are replaced by finite differences.

The leading term on the right hand side of Equation (3.2) is the basis of the linear examination of speech and many other problems in signal processing. The higher order terms have received far less attention, though Wiener's and Volterra's work has recently been treated in detail by Schetzen[9] and Rugh.[10]

The most natural case to consider in speech modeling, since it is next most difficult to the linear, is to assume the functional depends on $x(t)$ only up to the quadratic terms:

$$y(t) = \int h_1(\tau)x(t-\tau)d\tau + \int h_2(\tau_1,\tau_2)x(t-\tau_1)x(t-\tau_2)d\tau_1 d\tau_2 \tag{3.3}$$

We may assume without losing generality that h_2 is a symmetric function of its arguments. This being done, the filter functions h_1 and h_2 may be extracted in a variety of ways. One

method, for example, is to look at scaled up output for scaled down excitation, since, regarding $y(t)$ as a functional of $x(t)$, we may write

$$\int h_1(\tau)\, x(t-\tau)dt = \lim_{\varepsilon \to 0} \frac{1}{\varepsilon} y[\varepsilon x(t)] \,. \tag{3.4}$$

An alternative procedure is to note that

$$2 \int h_1(\tau)\, x(t-\tau)d\tau = y[x(t)] - y[-x(t)] \,. \tag{3.5}$$

In actual practice, the last suggests a way of pulling the linear part of the filter out of the speech signal, though it requires us to assume that the glottal pulse excitation has a very high power at the fundamental frequency relative to the power at the harmonics. For with this assumption, $-x(t)$ is approximately $x(t)$ translated by half a pitch period.

With the linear part known, we know by subtraction the quantity

$$\iint h_2(\tau_1\tau_2)\, x(t-\tau_1)\, x(t-\tau_2)d\tau_1 d\tau_2$$

and then

$$\int \int h_2(\tau_1,\tau_2)x_1(t-\tau_1)x_2(t-\tau_2)dt_1 dt_2 \,,$$

where x_1 and x_2 are independent and arbitrary excitations, can be obtained by the familiar process of polarizing a quadratic form to get the underlying symmetric bilinear form.

The mathematical procedure above is probably not the one to use in practice. The proper procedure is probably to assume the excitation, and then get $h_2(\tau_1,\tau_2)$ by Fourier analysis.

4.0 FREQUENCY DOMAIN REPRESENTATION OF NONLINEAR SYSTEMS

The linear impulse function $h_1(\tau)$ has a corresponding transfer function in the frequency domain, $L(f)$, defined by

$$h_1(\tau) = \int L(f)\, e^{-2\pi i f \tau} df \,. \tag{4.1}$$

Similarly, the quadratic impulse function $h_2(\tau_1,\tau_2)$ has an allied transfer function $Q(f_1,f_2)$ given by

$$h_2(\tau_1,\tau_2) = \iint Q(f_1,f_2)\, e^{-2\pi i (f_1\tau_1 + f_2\tau_2)}\, df_1 df_2 \tag{4.2}$$

assuming that the integrals in Equations (4.1) and (4.2) exist. Higher order terms in the functional expansion in Equation (3.2) will similarly have associated with them higher order transfer functions.

The analysis of linear systems has been greatly aided by the fact that convolution in the time domain is equivalent to multiplication in the frequency domain. A similar result holds for nonlinear systems except that multiple order transformations must be used. If $x(t_1, \ldots, t_n)$ is an integrable deterministic function then its transform pairs are

$$X(f_1, \ldots, f_n) = \int \ldots \int x(t_1, \ldots, t_n)$$

$$e^{2\pi i(f_1 t_1 + \ldots f_n t_n)} dt_1, \ldots, dt_n$$

$$x(t_1, \ldots, t_n) = \int \ldots \int X(f_1, \ldots, f_n)$$

$$e^{-2\pi i(f_1 t_1 + \ldots + f_n t_n)} df_1 \ldots, df_n$$

If $x(t)$ is a stochastic function, these integrals are replaced by Fourier-Stieltjes integrals.

The value of the higher order transform theory lies in that

$$\int \ldots \int h_n(\tau_1, \ldots, \tau_n)\, x(t_1 - \tau_1, \ldots, t_n - \tau_n) d\tau_1 \ldots d\tau_n$$

has an nth order transform

$$H_n(f_1, \ldots, f_2)\, X(f_1, \ldots, f_n) \cdot$$

For example, the quadratic term in Equation (2.2),

$$y(t_1, t_2) = \iint h_2(\tau_1, \tau_2)\, x(t_1 - \tau_1)\, x(t_2 - \tau) d\tau_1 d\tau_2 \tag{4.3}$$

will have in the frequency domain the corresponding relation

$$Y_2(f_1,f_2) = H_2(f_1,f_2)\, X(f_1)\, X(f_2) \quad .$$

We are, of course, interested only in the special case $t_1 = t_2$

$$y_2(t) = y_2(t,t) = \int\int h_2(\tau_1,\tau_2)\, x(t-\tau_1)\, x(t-\tau_2)\, d\tau_1 d\tau_2$$

which is the quadratic term in the functional expansion in Equation (3.2).

In constructing mathematical models for speech, one should inquire as to whether the higher order functionals in Equation (3.2) are of significance. We can explore this question by examining the higher order statistics of the speech signal $y(t)$. We assume that $y(t)$ is a stationary random function of time; any trends have been removed prior to analysis. The spectral representation of $y(t)$ is given by the Fourier-Stieltjes integral

$$y(t) = \int d^{2\pi i f t} dZ(f) \quad .$$

The ordinary (second order) spectral density of $y(t)$ is defined by

$$P(f)df = E[dZ(f)dZ(-f)] \tag{4.4}$$

The power spectrum $P(f)$ is related to the autocovariance function $R(\tau)$

$$P(f) = \int R(\tau)e^{-2\pi i f\tau} d\tau$$

where $R(\tau)$ is defined by the ensemble mean

$$R(\tau) = E[y(t)y(t + \tau)]$$

or

$$R(\tau) \simeq \frac{1}{2T} \int_{-T}^{T} y(t)y(t + \tau)dt$$

if time averaging is equivalent to ensemble averaging.

Higher order spectral densities can similarly be defined. In particular the third order or bispectral density $B(f_1,f_2)$ is defined by

$$B(f_1,f_2)df_1df_2 = E[dZ(f_1)\ dZ(f_2)\ dZ(f_3)] \tag{4.5}$$

with $f_1 + f_2 + f_3 = 0$.

The bispectrum $B(f_1,f_2)$ is related to the ensemble average of the third order lagged product by

$$B(f_1,f_2) = \iint S(\tau_1,\tau_2)\ e^{-2\pi i(f_1\tau_1+f_2\tau_2)} d\tau_1 d\tau_2$$

where

$$S(\tau_1,\tau_2) = E[y(t)\ y(t+\tau_1)\ y(t+\tau_2)] \ .$$

Equations (4.4) and (4.5) provide insight into the interpretation of the spectrum and bispectra. The spectrum represents the contribution to the mean square $\overline{y^2(t)}$ from the product of two Fourier components whose frequencies add to zero. The bispectrum represents the contribution to the mean cube from the product of those three Fourier components whose resultant frequency is zero.

Equations (4.4) and (4.5) lead immediately to the symmetry relations for the spectrum and bispectrum

$$P(f) = P(-f)$$

and $$B(f_1,f_2) = B(f_2,f_1) = B(f_1,-f_1-f_2)$$

$$= B(-f_1-f_2,f_1) = B(f_2,-f_1-f_2)$$

$$= B(-f_1-f_2,f_2) \ .$$

The spectrum is real and is determined by its values on the half line. The bispectrum is complex and is determined by the values in an octant; for example $0 \leq f_1 < \infty$, $0 \leq f_2 \leq f_1$.

The dimensionless ratio

$$\frac{E(y^3)}{[E(y^2)]^{3/2}}$$

is called the skewness and is usually finite for non-Gaussian processes. A related ratio for the bispectrum, called bicoherency is:

$$\frac{B(f_1,f_2)}{[P(f_1)\ P(f_2)\ P(f_3)]^{1/2}}$$

where $f_3 = -f_1 - f_2$. It may be computed in practice as follows. Pass the signal $y(t)$ through three band pass filters centered at f_1, f_2 , and $(-f_1-f_2)$ respectively, each with band width Δ , to get three new signals $y_1(t)$, $y_2(t)$, $y_3(t)$ respectively.

Then the ratio

$$\frac{E(y_1 \cdot y_2 \cdot y_3)}{[E(y_1^2)\ E(y_2^2)\ E(y_3^2)]^{1/2}}$$

is (approximately) the bicoherency times $\Delta^{1/2}$. Narrowing the filter width Δ is equivalent to time averaging over increasing time interals $1/\Delta$.

5.0 SOME TOY PROBLEMS

In order to acquire some intuition about the bispectrum, its properties and possible applications, we consider four simple problems.

First, let us assume we have a signal process $y(t)$ derived from an excitation process $x(t)$ by $y(t) = \int h(\tau)\, x(t-\tau)\, d\tau$. If we now assume we can compute the bispectra from the Fourier transform of lagged triple products by time averaging, then we obtain by a simple formal manipulation (which would probably be hard to justify in all details):

$$B_y(f_1,f_2) = B_x(f_1,f_2)H(f_1)H(f_2)H(-f_1,-f_2) \tag{5.1}$$

where B_x and B_y are the bispectra of input and output processes respectively and $\hat{H}$ is the Fourier transform of the filter, i.e.,

$$H(f) = \int h(\tau)\, e^{-2\pi i \tau f}\, d\tau$$

The important relation (5.1) already begins to show us some of the utility of the bispectrum. Unlike the power spectrum, we see that the bispectrum of the output process depends on the phase relations in the Fourier transform of the filter function.

Carrying the analysis a little further, let us note that the triple lagged correlation for Gaussian white noise is identically zero, hence the bispectrum of Gaussian white noise is zero. We conclude:

- The bispectrum of linearly filtered Gaussian white noise is zero.
- To the extent fricative sounds may be modeled on linearly filtered Gaussian noise, we expect then they have zero bispectrum.
- The bispectrum of a pure sinusoidal signal is zero. On the other hand, the bispectrum of the glottal pitch excitation will only show peaks at pairs of frequencies, each of which are multiples of the fundamental frequency. This study suggests that the bispectrum of voiced sounds will show peaks along certain lines radiating from the origin.

Next, consider a signal $x(t)$ that consists of two cosine waves with differing frequencies and phase

$$x(t) = A \cos\left(2\pi f_1 t + \phi_1\right) + B \cos\left(2\pi f_2 t + \phi_2\right)$$

The signal is passed through a quadratic device to yield a new signal $y(t)$

$$y(t) = x^2(t)$$

$$= A^2\cos^2(2\pi f_1 t+\phi_1) + B^2\cos^2(2\pi f_2 t+\phi_2)$$

$$+ 2AB\cos(2\pi f_1 t+\phi_1)\cos(2\pi f_2 t+\phi_2)$$

which can be written as

$$y(t) = \frac{1}{2}(A^2+B^2) + (A^2/2)\cos[2(2\pi f_1+\phi)]$$

$$+ (B^2/2)\cos[2(2\pi f_2+\phi_2)]$$

$$+ AB\{\cos[2\pi(f_1+f_2) + (\phi_1+\phi_2)]$$

$$+ \cos[2\pi(f_1-f_2) + (\phi_1-\phi_2)]\}$$

The squaring operation thus produces a DC component, $1/2(A^2 + B^2)$, components at twice the frequencies initially present and heterodyne frequencies at $f_1 + f_2$ and $f_1 - f_2$. Such relationships with corresponding phase relation information would readily be picked up in a bispectral analysis.

The next example consists of passing non-Gaussian white noise through a linear device. Consider a stochastic process $x(t)$ with the following properties

$$E[x(t)] = 0$$

$$E[x^2(t)] = 1$$

$$E[x^3(t)] = \beta$$

The characteristics of the linear device are given by the impulse response function $h(\tau)$ or its corresponding transfer function $H(f)$. The spectral density of the process $y(t)$

$$y(t) = \int h(\tau)x(t - \tau)d\tau$$

is then

$$P(f) = |H(f)|^2$$

and the bispectral density is

$$B(f_1,f_2) = \beta H(f_1)\ H(f_2)\ H(-f_1-f_2)$$

We next consider a Poisson process. The generated process is of the form

$$y(t) = \sum_k h(t-x_k)$$

where $\ldots x_{-N} \ldots, x_{-1}, 0, x_1, \ldots, x_k, \ldots$ are the times of events for a Poisson process with

$$E(x_{k+1}-x_k) = \mu \quad .$$

The transfer function $H(f)$ is defined by

$$h(t) = \int H(f)\, e^{2\pi i f t} df \quad .$$

The spectral and bispectral densities are then

$$P(f) = \frac{1}{\mu} |H(f)|^2$$

$$B(f_1,f_2) = \frac{1}{\mu} H(f_1)\, H(f_2)\, H(-f_1-f_2) \quad .$$

The last two examples illustrate that different processes can yield similar bispectral. A normalized bispectrum, the bicoherency is defined by

$$Bc = \frac{B(f_1,f_2)}{[P(f_1)\, P(f_2)\, P(f_1+f_2)]^{1/2}}$$

and is useful in many applications. For the case of the Poisson process

$$Bc = (\mu)^{1/2} \frac{H(f_1)}{|H(f_1)|} \cdot \frac{H(f_2)}{|H(f_2)|} \cdot \frac{H(-f_1-f_2)}{|H(-f_1-f_2)|}$$

which exhibits the phase relations between the frequencies which satisfy the requirement

$$f_1 + f_2 + f_3 = 0 \quad .$$

For the case of non-Gaussian noise passed through a linear device we have that

$$Bc = \beta \frac{H(f_1)}{|H(f_1)|} \cdot \frac{H(f_2)}{|H(f_2)|} \cdot \frac{H(-f_1-f_2)}{|H(-f_1-f_2)|} \quad .$$

As a final example we consider a Gaussian process with zero mean and unit variance

$$E\,[x(t)] = 0 \qquad E[x(t)^2] = 1$$

with a spectral density $X(f)$. We pass the Gaussian process through a nonlinear device. Let $x(t)$ be operated on by the nonlinear device to provide a new process $y(t)$

$$y(t) = N[x(t)] \quad .$$

If N is odd in the variable t

$$N(t) = -N(-t)$$

then the bispectrum vanishes so that it would not be possible to distinguish between a signal $x(t)$ operated on by N and Gaussian noise. For the analysis of odd nonlinear operators we must examine the higher order terms in Equation 3.2. Alternatively, if N contains even terms, for example

$$y(t) = x(t) + \alpha x^2(t)$$

then the bispectrum no longer vanishes. The process has a spectral density

$$P(f) = X(f) + 2\alpha^2 \int X(g)\, X(f-g)\, dg$$

and a bispectral density

$$B(f_1,f_2) = 2\alpha[X(f_1)\, X(f_2) + X(f_1)\, X(f_1+f_2)]$$

$$+ X(f_2)\, X(f_1+f_2)] + \text{terms of fourth order}.$$

For a Gaussian process

$$E[y(t)\, y(t+\tau_1)\, y(t+\tau_2)]$$

vanishes for all τ_1 and τ_2. Thus if the bispectrum of $y(t)$ is not identically zero then the process is not Gaussian.

6.0 NUMERICAL COMPUTATIONS

In actual numerical calculations of real time series a number of computational approaches to estimating the bispectrum have been employed. In the earliest applications, narrow band filtering (complex demodulation) plus averaging in the time domain was used[11,12,13]. Haubrich[14] obtained estimates by averaging over several records and Brillinger and Rosenblatt[15] averaged over the frequency domain. With the wide spread use of the Fast Fourier Transform (FFT), averaging over the frequency domain has become more popular.[16] However, the computational aspects of bispectral analysis are still in their infancy as well as the statistical character of the bispectral estimates. Because of this we will rev ·w the three principal approaches to the estimation of the bispectrum.

With digital sampling at interval Δt , energy associated with frequencies greater than the Nyquist frequency

$$f_N = (2\Delta t)^{-1}$$

appear in the power spectrum in the alias of a lower frequency. It is convenient in numerical work to refer all frequencies to the dimensionless Nyquist units f/f_N . To avoid aliasing in the power spectrum, one must sample at a rate sufficient to assure that the cumulative spectral density above the unit Nyquist frequency is very

much lower than for frequencies below the unit Nyquist frequency. Similarly, to avoid aliasing in the bispectrum, the bispectrum must be small for all frequencies f_1, f_2 outside the region $|f_1 + f_2| \leq 1$ where f_1 and f_2 are measured in Nyquist units. The restriction of this region to the octant $0 \leq f_1 \leq 1$, $0 \leq f_2 \leq f_1$ yields a triangular region in f_1, f_2 space with the interval $0 \leq f_1 \leq 1$ as a base line and the point $(1/2, 1/2)$ as the apex. For a time series sampled at equal time intervals, this triangular region completely defines the bispectrum.

High frequencies are often removed from a time series by low pass filter. If this procedure is adapted, care must be taken to avoid phase distortion since the bispectrum is a means of investigating phase relations in stochatic processes. For example, if filters of the autoregressive type are to be used, then they should be run twice in opposite directions.

In general, the bispectrum is estimated over some bandwidth Δ. The number of time samples N_o, required to estimate the bispectrum with a frequency interval Δ is at least

$$N_o = \frac{1}{\Delta t \Delta} .$$

The total length of the record N can be written

$$N = MN_o \tag{6.1}$$

where for ease of use of the FFT, N should be a power of 2 and M is preferably an odd integer. It should be noted that the common practice of adding zeros to make N a power of two should be avoided since this distorts the real data. If points need to be filled in, then maximum entropy methods should be employed, or alternatively, one should allow overlap between consecutive records. If one wishes to cover the entire triangular region $0 \leq f_1 \leq 1$, $|f_1 + f_2| \leq 1$ with estimates spaced at intervals Δ, then the bispectrum must be evaluated at at least $N_o^2/16$ points.

For a given time series, the first step in numerically evaluating the bispectrum is to subtract the mean from all values and remove any linear trend. In order to use the computational advantages of the FFT the time series should be windowed or tapered by multiplying by a function that smoothly decays to zero at both ends of the time series. A number of techniques for tapering have been suggested in the literature[17] of the form

$$y'(t) = C(t/N)\, y(t) = C(\tau)\, y(t) \,.$$

A widely used family of tapers is the cosine taper defined by

$$C(\tau) = 1/2 \left(1 - \cos \frac{\pi}{\alpha} \tau\right) \text{ for } 0 \leq \tau < \alpha$$

$$= 1 \quad \alpha \leq \tau \leq 1 - \alpha$$

$$= 1/2\left[1 - \cos \frac{\pi}{\alpha} (1-\tau)\right] , \quad 1 - \alpha < \tau \leq 1 \,.$$

For $\alpha = 1/2$, there is full tapering and the tapering is termed the Hanning window.

Tapering before taking the Fourier transform has the advantage of reducing the leakage of spectral peaks to frequencies far away but it coarsens the spectral resolution by a factor of $(1 - \alpha)$ for the cosine tapers. Tapering has the further disadvantage of changing the expectation and variances of the estimates of the spectral and bispectral densities. Huber et al.[18] show that tapering increases the variance of the spectral and bispectral densities by approximate factors of $D_4 D_2^{-2}$ and $D_6 D_3^{-2}$ respectively where

$$D_k = \int_0^1 C^k(\tau) d\tau$$

for large values of M (see Equation 5.1). For $M = 1$ Huber et al. find

$$D_4 D_2^{-2} = \frac{35}{18} \qquad \alpha = 1/2$$

$$= 1 + \frac{67}{64} \alpha + \cdots \qquad \text{for small } \alpha$$

$$D_6 D_3^{-2} = \frac{231}{100} \qquad \alpha = 1/2$$

$$= 1 + \frac{615}{512} \alpha + \cdots \qquad \text{for small } \alpha \; .$$

The Hanning taper increases the variance of bispectral estimate by a factor of 2.31. It is thus desirable to increase N, the length of the record, reducing α keeping $N\alpha$ constant, since this reduces the leakage of power from the spectral peaks.[18]

Leakage remains a serious problem even after windowing if the spectral peaks are several decibels above the mean spectral density. This problem is a particular problem for spectra that vary as the inverse frequency, a behavior that is found in many natural phenomenon.

After tapering in the time domain use of the FFT yields

$$Y(f) = Y_k = \frac{1}{N} \sum_{t=o}^{N-1} y_t \, e^{-2\pi i(kt/N)} \quad 0 < k < N/2$$

where y_t denote the N discretely sampled values

$$y_o, y_1, \cdots, y_t, \cdots y_{N-1} \ .$$

The frequency f is determined by k since

$$f = k/N\Delta t = k\Delta/M \ .$$

The estimated power spectral density in the band Δ_f is then

$$\hat{P}(f)\Delta = \sum_{j=-L}^{L} |Y_{k+j}|^2 \qquad M = 2L + 1 \ .$$

This relation indicates the importance of selecting a relatively high value for M so that a number of values of Y_k enter into the estimation of $P(f)$.

The estimation of the bispectral density can proceed in a number of ways. If the averaging is over quadratic window then

$$\hat{B}\,(f_1,f_2)\,\Delta^2 = \hat{B}\,[k_1'(N\Delta t),\, k_2'(N\Delta t)]\,\Delta^2$$

$$= \sum_{j_1=-L}^{L} \sum_{j_2=-L}^{L} Y_{j_1+k_1}\, Y_{j_2+k_2}\, Y_{-j_1-j_2-k_1-k_2} \tag{6.2}$$

Huber et al.[18] suggest averaging over a symetrical hexagonal window such that

$$\hat{B}(f_1,f_2)\,\Delta^2 = \frac{4M^2}{3M^2+1} \sum Y_{j_1+k_1}\, Y_{j_2+k_2}\, Y_{-j_1-j_2-k_1-k_2} \tag{6.3}$$

where the summation is taken over all values of j_1 and j_2 that satisfy the conditions

$$|j_1| < L\,, \quad |j_2| < L\,, \quad |j_1+j_2| < L\,.$$

It would appear that (6.3) is probably the most efficient way of computing the bispectrum.

Most of the sampling properties that have been analyzed are asymptotic.[15] A detailed discussion of the asymptotic behavior of bispectral estimates is given by Lii et al.[19] The covariance of two complex variables X and Y can be written as

$$\text{cov}\ (X,Y) = E(XY^*) - E(X)\ E(Y^*)$$

where Y^* is the complex conjugate of Y. The variance of Y is then

$$\text{var}(Y) = \text{cov}\ (Y,Y)$$

and the asymptotic variance of the bispectral estimate $\hat{B}\ (f_1,f_2)$ is

$$\text{var}\ [\hat{B}\ (f_1,f_2)] = \frac{N\Delta}{M^2}\ P(f_1)\ P(f_2)\ P(f_1+f_2)$$

where N is the total length of the record (Eq. 5.1) so that

$$\text{var}[\hat{B}\ (f_1,f_2)] = \frac{1}{M\Delta}\ P(f_1)\ P(f_2)\ P(f_1+f_2)$$

and the asymptotic variance of the estimate of the bicoherency $\hat{B}c$ is

$$\text{var}(\hat{B}c) = \frac{1}{M\Delta}$$

if the statistical variability of the denominator of $\hat{B}c$ is neglected, a quadratic window is used and there is no tapering. The effects of tapering on the variance of the estimates have already been

considered. If hexagonal windows[18] are used then the above variances must be multiplied by

$$\frac{4M^2}{3M^2+1} \approx \frac{4}{3} .$$

7.0 APPLICATIONS OF BISPECTRAL ANALYSIS

The notion that frequency representation for a cumulant might be useful is probably due to Kolmogorov as acknowledged by Shiryaev.[20] Tukey[21,22] independently recognized the value of the information that could be derived from bispectral analysis. The first actual calculation of the bispectrum was on data of the height of ocean waves[11]. Since then there have been a number of exploratory efforts in a number of fields, not including speech. The analysis of the ocean waves prior to breaking seemed to show a non-linear behavior that was predicted by the Navier-Stokes equations. The authors concluded that the bispectrum analysis in this field confirmed the theoretical predictions of the shape of weakly non-linear waves.[11]

In an exploratory investigation of seasonal weather patterns, MacDonald indicated that bispectral analysis showed a strong non-linear interaction between the band at a cycle per year and the higher harmonics at low latitudes but no such interaction at high latitudes.[12] There has been no further work along these lines. Preliminary investigations have been made of non-linear interaction in the generation of microseisems[14] and of "pearl" oscillations in geomagnetic data[23] but again these studies have not led to a greater understanding of the phenomenon of interest. As a result of a bispectral analysis, there has been the suggestion that certain peaks in the spectra of the great Chilean and Alaskan earthquakes arise from the

interaction of different modes of free oscillations of the earth.[24] Cartwright[25] used bispectral analysis to examine the interaction of tides with the continuum ocean spectrum. Roden and Bendiker[26] have examined profiles of oceanic variable using bispectral techniques.

Outside of geophysics, bispectral analysis has been applied in a number of fields. Lii et al.[19] examined turbulence at high velocities and Reynolds number to show that the contribution of wavenumber triplets to the rate of vorticity production and spectral transfer are non-local in character. Lumley and Takeuchi[27] and Helland et al.[28] have also applied bispectral methods to the analysis of turbulence. Hasselman[29] showed that the non-linear transfer function for a ship moving in irregular waves can be obtained from bispectral estimates. In a bispectral analysis of eletroencephalogram signals, Barnett et al.[30] find that significant quadratic coupling between components making up the electroencephalogram occured only for awake subjects having a high level of alpha activity. Godfrey[13] had indicated the usefulness of bispectral analysis in detecting nonlinear behavior of economic time series.

An important application bispectrum analysis has been described by Sato et al.[31] The noise of gears in a machine are recorded acoustically. .1 the normal state of operation the noise exhibits a nonvanishing bispectrum because of the inherent nonlinearities of the machine. As scores develop on the gear surfaces, the modulus of the bispectrum decrease as the noise generated by the randor imperfections

in the gears increases. A conventional power spectrum analysis would fail to detect the enhancement of the noise since the principle harmonics would be present but only slightly broadened by the noise. This provides a method of determining the wear on the gears without actual direct inspection of the gears.

8.0 CONJECTURES REGARDING THE APPLICATION OF BISPECTRAL ANALYSIS TO SPEECH.

Despite the numerous applications of bispectral analysis to a wide variety of problems, bispectral methods have not been employed in the analysis of speech. The speech signal clearly contains nonlinear components; for example, the fricatives almost surely involve the interaction of noise with the signal which might be primitively modeled by

$$x_t = \alpha\, x_{t-1} + \beta\, x_{t-1}\, \varepsilon_{t-1} + \varepsilon_t$$

where x_t is the signal at time t and ε_t is the "noise." The first conjecture about speech is that the bispectrum shows significant structural properties.

One important application of bispectrum analysis, if indeed the bispectrum of speech is nonvanishing, is to use bispectral analysis to determine whether or not a magnetic tape which contains speech and noise is worth keeping if high noise levels were introduced when the speech was recorded or noise is due to the aging of the tape or processing of the tape. This application is suggested by the work of Sato et al.[31] A means of testing this conjective is to record noise free speech, add white noise to the recording until the speech becomes unintelligible and measure the bispectrum at the various stages of intelligibility degradation. Using this means it may be possible to

distinguish between those tapes that should be kept for further analysis and those that can be destroyed without loss of recoverable information.

A second conjecture is that two speakers will have voice characteristics that have different bispectra because of differences in vocal chord structure and geometry of the mouth, nose and larynx. If this conjecture is true, then it should be possible to distinguish which of two, or more, speakers is talking at only one time estimating the bispectra using digital or analog real time estimates.

A third problem relates to word recognition. To date word recognition methods have depended very heavily on the use of linear methods, particularly linear predictive coding (LPC). It may turn out that the nonlinear aspects of word recognition are of significance. For word spotting, the bispectrum may be useful.

REFERENCES FOR PART I

1. Flanagan, J. 1972. *Speech Analysis, Synthesis and Preception.* Springer-Verlag, New York.

2. Markel, J. and A. Gray. 1976. *Linear Prediction of Speech.* Springer-Verlag, New York.

3. Burg, J. 1967. *Maximum Entropy Spectral Analysis.* Paper presented at the 37th International Society of Exploration Geophysics, Oklahoma City, October 31, 1967.

4. Capon, J. 1969. High-resolution frequency-wavenumber spectral analysis. *Proc. IEEE* 57:1408-18.

5. Huber, P. 1981. *Robust Statistics.* J. Wiley, New York.

6. Volterra, V. 1959. *Theory of Functionals and of Integral and Integro-Differential Equations.* Dover, New York.

7. Wiener, N. 1958. *Nonlinear Problems in Random Theory.* Technology Press and J. Wiley, New York.

8. Lighthill, M. 1958. *Introduction to Fourier Analysis and Generalized Functions.* Cambridge University Press.

9. Schetzen, M. 1980. *The Volterra and Wiener Theories of Nonlinear Systems.* J. Wiley, New York.

10. Rugh, W. 1981. *Nonlinear Systems Theory: The Volterra/Wiener Approach.* Johns Hopkins Press, Baltimore.

11. Hasselmann, K., W. Munk and G. MacDonald. 1963. Bispectra of ocean waves in *Advanced Seminar on Spectral Analysis of Time Series.* M. Rosenblatt (ed.). J. Wiley, 125-139.

12. MacDonald, G. 1963. The bispectrum of atmospheric pressure records. *Proc. IBM Sci. Comput. Symp. on Statistics.* IBM, White Plains, NY, 247-264.

13. Godfrey, M. 1965. An exploratory study of the bispectrum of economic time series. *Appl. Stat..* 14:48-69.

14. Haubrich, R. 1965. Earth noise, 5-500 millicycles per second. *J. Geophys. Res..* 70:1415-1427.

15. Brillinger, D. and M. Rosenblatt. 1967. Computation and interpretation of k-th order spectra. *Advance Seminar on Spectral Analysis of Time Series*, B. Harris (ed.) 189-232.

16. Barnet, T., L. Johnson, P. Naitoh, N. Hicks and C. Nute. 1971. Bispectrum analysis of electroencephalogram signals during waking and sleeping. *Science*. 172:401-402.

17. Bingham, C., M. Godfrey and J. Tukey. 1967. Modern techniques of power spectrum estimation. *IEEE Trans. Audio Electroacoust.* AU-15:56-66.

18. Huber, P., R. Kleiner, T. Gasser and G. Dumermuth. 1971. Statistical methods for investigating phase relations in stationary stochastic processes. *IEEE Trans. Audio Electroacoust.* AU-19:78-86.

19. Lii, K., M. Rosenblatt and C. Van Atta, 1976 Bispectral measurements in turbulence, *J. Fluid Mech.* 77:45-62.

20. Shiryaer, A. 1960 Some problem in the spectral theory of higher-order moments. *J. Theor. Prob. Appl.* 5: 265-284.

21. Tukey, J. 1959 An introduction to the measurement of spectra in *Probability and Statistics*, U. Grenander. ed. Wiley, New York, 300-330.

22. Tukey, J. 1958 Personal communication to G. MacDonald and W. Munk, July, La Jolla, California.

23. Madden, T. 1964 Spectral, cross spectral and bispectral analysis of low frequency electromagnetic phenomena in *Natural Electromagnetic Phenomena*. Below 30 k c/s, D. Blied, ed. Plenum Press, New York, 429-450.

24. Zadro, M. and N. Capoto 1967 Spectral and bispectral analysis and Q of the free oscillations of the earth. *Supple. Nuovo Cimentor* 6: 67-68.

25. Cartwright, D. 1968 A unified analysis of tides and surges round North and East Britain, *Phil. Trans. A.* A, 26351-55.

26. Roden, G. and D. Bendiner 1974 Bispectra and cross-spectra of temperature, sound velocity and density function with depth off Northeastern Japan. *J. Phip. Ocean.* 3:308-317.

27. Lumleg, J. and K. Takevchi 1976 Application of central-limit theorems to turbulence and higher order spectra. *J. Fluid Mech.* 74, 433-468.

28. Helland, K, K. Lii and M. Rosenblatt 1977 Bispectra of atmospheric and wind tunnel turbulence in *Application of Statistics*, P. Krishhaiahed. North Holland, 223-248.

29. Hasselmann, K., 1966 On nonlinear ship motions in irregular waves, *J. Ship Research*, 10:64-68.

30. Barnett, T., L. Johnson, P. Naitoh, N. Hicks, and C. Nute Bispectrum analysis of electroencephalogram signals during waking and sleeping, *Science*; 122-401-402.

31. Sato, T., K. Sasaki and Y.N. Kamond. 1977, Real-time bispectral analysis of gear noise and its application to contactless diagnosis, *J. Acoust. Soc. Am.*, 62:382-387.

PART II

9.0 A "NOVEL" VIEW OF LINEAR PREDICTIVE CODING.

9.1 Introduction

In the typical set-up for linear predictive coding (LPC), we start with a (not necessarily causal) relation between excitation u and output s of the form

$$s_n = \sum_{k=-\infty}^{\infty} \alpha_{n-k} u_k \ .$$

One forms, at a purely formal level, the Laurent series: $S(x) = \sum_{-\infty}^{\infty} s_n x^n$, $H(x) = \sum_{-\infty}^{\infty} \alpha_n x^n$, $U(x) = \sum_{-\infty}^{\infty} u_n x^n$, to get $S(x) = H(x)\, U(x)$, a relation which is true at the level of formal Laurent series, but not necessarily at the level of function, as we will show in a moment.

For technical convenience, we will assume that the excitation is finitely supported, so that the series for $U(x)$ makes sense as a function. (In somewhat greater generality, we could permit the excitation to be such that $U(x)$ is analytic except at zero and infinity).

In practical applications, it is often assumed that the filter transfer $H(x)$ is actually a rational function, or has a very "good"

approximation by a rational function, which replaces it in further discussion. We will make the same assumption.

The filter is called minimum phase if all the poles of $H(x)$, except possibly for a pole at zero, are outside the unit circle.

In auto-regressive (AR) modeling, or all pole modeling, it is assumed that $H(x)$ is the reciprocal of a polynomial, except for a stray power of x, which is absorbed into the excitation, because we are free to choose time $n = 0$ independently in excitation and output.

We are free to assume, then, that $H(x) = \frac{c}{P(x)}$, where $P(x) = a_k x^k + a_{k-1} x^{k-1} + \ldots + a_1 x + 1$. By changing polarity and gain in the excitation if necessary, we may even assume $c = 1$.

From this point on we assume that $H(x) = \frac{1}{P(x)}$, with $P(x)$ as above.

9.2 How does the typical analysis proceed from this point? One simply notes that $P(x)\, S(x) = U(x)$, so that

$$s_n = -(a_1 s_{n-1} + a_2 s_{n-2} + \ldots a_k s_{n-k}) + u_n ,$$

and given a (patch of) signal, one estimates the a's by a familiar least squares residual error procedure, which comes down to solving k linear equations for unknown coefficients $a_1, a_2, \ldots, a_k$.

Let's make this a little more precise. For convenience in exposition, we assume a bilaterally infinite signal output is given. (This is the so-called auto-correlation method). We will assume additionally that the total signal energy $\sum_{-\infty}^{\infty} s_n^2$ is finite, so the auto-covariance $r(p) = \sum_{-\infty}^{\infty} s_n s_{n-p}$ also exists for all p. This is the same as saying that at the level of formal Laurent series,

$$R(x) = \sum_{-\infty}^{\infty} r(p)\, x^p = S(x)\, S(1/x) .$$

We then seek numbers $y_1, y_2, \ldots, y_k$ so as to minimize

$$\sum_{n=-\infty}^{\infty} \left[s_n + (y_1 s_{n-1} + \ldots + y_k s_{n-k})\right]^2$$

Putting in a dummy variable y_o, held equal to +1, this comes to minimizing the quadratic form

$$\sum_{0<i,j\leq k} y_i y_j\, r(i-j)$$

which boils down to solving the linear system:

$$\sum_{i=1}^{k} y_i r(j-i) = -r(j)$$

$j = 1, 2, \ldots, k$.

If we define the polynomial (with variable coefficients) by $Y(x) = \sum_{v=0}^{k} y_v x^v$, $y_o = 1$, then our linear system of equations says simply that the formal Laurent series $Y(x)\, R(x)$ has vanishing coefficients for the first through k^{th} power of x.

Note that $Y(x)R(x) = Y(x)S(x)S(1/x)$. For a given Laurent series $L(x)$ let us agree that $L(x) = o(x^p)$, $p > 1$, means that the coefficients of $x, x^2, \ldots, x^p$ are all zero. In this notation, our linear system of equations is simply the requirement:

$Y(x)S(x)S(1/x) = o(x^k)$, with the boundary condition $Y(o) = y_o = 1$.

There are a few facts worth noting here.

(1) The residual square error is the constant coefficient in $Y(x)\ Y(1/x)\ S(x)\ S(1/x)$.

(2) Running the signal backwards is equivalent to replacing $S(x)$ by $S(1/x)$, and so does not alter the equation for $Y(x)$.

(3) As is known, $Y(x)$ always has all of its zero outside the unit circle.

(4) Computing $S(x) = \frac{1}{P(x)} U(x)$ using power series for $\frac{1}{P(x)}$ about the origin will give a causal relation between output and excitation (i.e. the signal depends on earlier values of the excitation.

(5) The signal computed by (4) has finite energy if and only if $P(x)$ has all of its zeros outside the unit circle. (We mean the statement to apply to all finitely supported excitations).

(6) If the signal is computed as in (4) and $P(x)$ has its zeros outside the unit circle, then the solution $Y(x)$ to our linear system $Y(x)\ S(x)\ S(1/x) = o(x^k)$ will be $Y(x) = P(x)$ if the excitation consists of a single pulse, approximately so if the excitation consists of widely separated approximately equal pulses, but otherwise will be generally quite different from $P(x)$.

9.3 All of this is quite familiar, except possibly for our notation. What we have to note that is, we believe, new is that there are many relations between excitation and output for which $P(x)\ S(x) = U(x)$, over and above the one described in (4) of the previous section.

This is quite easy to see. $\frac{1}{P(x)}$ has a convergent Laurent series expansion in the annular region lying between any two consecutive zeros of $P(x)$. Using such an expansion gives a linear relation between excitation and output which is definitely non-causal, i.e., the signal depends on the remote future of the excitation.

If we take the Laurent expansion beyond the last zero of $P(x)$, we obtain a "pure" non-causal filter, i.e., the signal depends only on the future of the excitation. The other Laurent expansions give dependence of signal both on future and past.

For a signal $S(x)$ computed by any one of the Laurent series above, we always have $P(x)\,S(x) = U(x)$. This observation raises a number of interesting questions. Foremost among these is the following: When we are modeling by LPC using a least square residue criterion, how do we know which of the various realizations of the filter we are really getting? Of equal importance is: Can one conceive of filters which realistically operate in the non-causal modes described above?

Let's address the second questions first. Here we will simply be argumentative, leaving it to the reader to decide how convincing our argument is. In the formation of the speech signal, it is clear that the brain and central nervous system is well ahead in time of the actual muscular articulation and glottal excitation. So the nervous system is in a position to modify the signal on the basis of excitation yet in the future. Now some will argue that the time scale of neural pulses is large compared to glottal pulses. But the brain and central nervous system are large scale parallel processors, with lots of feedback, so it is not altogether inconceivable that some anticipatory signal reforming is carried on. At the same time, from the point of view of ear-brain received signal processing, it is quite plausible that the signal for brain interpretation is reformed out of the received signal by looking both forwards and backwards in time, at a fine temporal resolution.

Whether or not the reader is convinced by these arguments, he will shortly see that our willingness to look at the non-causal versions

of the filter give- · · some extra processing power that we did not have before, without any serious increase in processing cost. Moreover, we feel that we have better ways of recovering non-minimum-phase filters.

Now let's have a look at the first question. We think that the answer is simply that you do not, and you have to keep an open mind. If we really had a bilaterally infinite signal, then it is possible to tell from the annulus of convergence of $S(x)$ something of what is going on. But, of course, we are never in that situation.

Before proceeding further, it is useful for us to introduce a little additional notation. We will call a finite Laurent series which goes forward no further than the k^{th} power of x and backward no further than the ℓ^{th} power of $\frac{1}{x}$ a $[k,\ell]$ polynomial. If the coefficients of $\left(\frac{1}{x}\right)$, $\left(\frac{1}{x}\right)^2$, ..., $\left(\frac{1}{x}\right)^{\ell}$ and $x, x^2, \ldots, x^k$ are all zero in a Laurent series $L(x)$, we will write $L(x) = o\left(x^{[k,\ell]}\right)$.

Now, let's look at the realization of our filter in which $\frac{1}{P(x)}$ is expanded in a series beyond the last zero of $P(x)$. Here, we will find that s_n is given linearly in terms of $u_n, u_{n+1}, u_{n+2}, \ldots$, etc. It is clear that the natural way to model this situation is to assume that

$$s_n = -\left(b_1 s_{n+1} + b_2 s_{n+2} + \ldots + b_k s_{n+k}\right) + u_n$$ and minimize

$$\sum_{n=-\infty}^{\infty} \left[s_n + \left(b_1 s_{n+1} + b_2 s_{n+2} + \ldots + b_k s_{n+k}\right)\right]^2$$

In terms of the notation just introduced, this boils down to solving the linear system:

$$Z(x)\ S(x)\ S(1/x) = o(x^{[o,k]})$$

where $Z(x)$ is a $[o,k]$ polynomial with constant coefficient equal to 1.

The reader will recall from sec. 9.2 the polynomial $Y(x)$ which solved $Y(x)\ S(x)\ S(1/x) = 0(x^{k})$. It is then clear that $Z(x) = Y(1/x)$, and that our squared residual error is the same in both cases. It is important to note, however, that our inferred relation between signal and excitation is not the same in both cases, since one is "pure" causal filtering, the other "pure" non-casual filtering.

We still have to determine how to model the remaining realizations of the filter. In these cases, it seems natural and plausible to suppose that

$$s_n = -(c_1 s_{n-1} + c_2\ s_{n-2} + \ldots + c_p s_{n-p}$$

$$+ d_1\ s_{n+1} + d_2\ s_{n+2} + \ldots + d_q\ s_{n+q}) + u_n\ ,$$

$p + q = k$, and to choose the c's and d's so as to minimize

$$\sum_n (s_n + c_1\ s_{n-1} + \ldots + c_p\ s_{n-p} + d_1\ s_{n+1} + \ldots + d_q\ s_{n+q})^2$$

This boils down to solving the linear system

$$W(x)\ S(x)\ S(1/x) = o\left(x^{[p,q]}\right) ,$$

where $W(x)$ is a $[p,q]$ polynomial with constant coefficient equal to 1. There is an obvious relation between the solutions for the pair $[p,q]$ and the pair $[q,p]$, which has the following important consequence. If k is even then the solution $W(x)$ for the pair $[p,p]$, $p = k/2$ has the property $W(x) = W(1/x)$. Hence, we have only to solve $k/2$ linear equations, rather than k linear equations, which can be a substantial savings.

It is clear that if we go forward p places and backwards p places, we will get a smaller residual error than if we simply went forward p places, because we have more free variables over which to minimize. Yet in both cases we have only p linear equations to solve. We may lose the advantage conferred by having a Toeplitz system to which the method of Levinson applies, but in the covariance method with finite signal patch, this method does not strictly apply anyhow.

We also want to note for future reference that in the case of a general pair $[p,q]$, with solution $W(x)$, the residual square error is the constant term in the Laurent series for $W(x)\ W(1/x)\ S(x)\ S(1/x)$.

9.4 We want to draw some conclusions, based on our above discussion, for LPC modeling.

Suppose we have a signal and excitations related by $S(x) = \frac{1}{P(x)} U(x)$, and we do not know which of the various realizations of the filter we should be dealing with. If it were known in advance that all the zeros of $P(x)$ are outside the unit circle, then probably conventional LPC modeling is adequate. But suppose the filter is not minimum phase. For the sake of argument, let us suppose that no zero of $P(x)$ lies on the unit circle. Then for all the modelings of the filter there is one and only one for which the signal has finite energy--namely the one in which the annular region of convergence of $\frac{1}{P(x)}$ includes the unit circle. This then, is the only realization in which it makes sense to talk about the auto-correlation and the power spectrum of the signal. It is quite possible that the assorted methods for estimating the auto-correlation from a finite patch of signal, such as windowing, maximum entropy, or maximum likelihood, are implicitly assuming that realization of the filter for which the power spectrum exists. (We have not investigated this interesting and possibly quite important question).

We do not generally know in advance, of course, whether our filter is minimum phase. This strongly suggests to us that the best general method of LPC modeling is to pick a suitable value of k, and go forward k steps and backward k steps. This is to say, we assume that

$$s_n = c_1 s_{n-1} + c_2 s_{n-2} + \ldots + c_k s_{n-k}$$

$$+ d_1 s_{n+1} + d_2 s_{n+2} + \ldots + d_k s_{n+k} + u_n .$$

and minimize the residual square error as before. Whether or not we have an infinite signal, we may assume $c_i = d_i$, so we still have only k linear equations to solve, and we will clearly get a <u>smaller</u> residual square error than if we had carried out conventional k step LPC analysis.

The fact that we have a smaller residual error suggests that this more general modeling can be incorporated into a considerably improved data compression system. One might, for example, find the c's (hence the d's), find the derived excitation, transmitting both, and reform the signal at the other end by solving for s_{n+k} in terms of the excitation u_n and earlier values of the signal.

For the purpose of using our generalized LPC modeling in word recognition, some method of scoring template against signal is needed. We believe a simple modification of the Itakura procedure should work here.

9.5 We will finally make a few remarks about the behavior of residual square error for general LPC modeling as opposed to conventional. To this end, suppose we have a polynomial $P(x)$, $P(x) = 1 + a_1 x + \ldots + a_k x^k$, with all its roots outside the unit

circle and the signal we are analyzing is given by $S(x) = \frac{1}{P(x)} U(x)$, where $\frac{1}{P(x)}$ is expanded as a power series about the origin. Suppose for this signal we tried a general $[k,k]$ LPC modeling. What will we find, and what will the residual error be?

For the sake of our argument, yet still to give us a general idea of what's going on, we suppose the excitation consists of a single unit pulse at $n = 0$, so $U(x) = 1$.

For our more general modeling, we seek a $[k,k]$ polynomial $W(x)$, constant coefficient one, such that $W(x)\ S(x)\ S(1/x) = o\left(x^{[k,k]}\right)$. Now $P(x)\ P(1/x)\ S(x)\ S(1/x) = 1$, so if $c = 1 + a_1^2 + a_1^2 + \ldots + a_k^2$, then $W(x) = \frac{1}{c} P(x)\ P(1/x)$. The residual square error being the constant term in $W(x)W(1/x)S(x)S(1/x)$, we find it to be $\frac{1}{c}$, a substantial improvment over the error of 1 in conventional LPC modeling.

REFERENCES FOR PART II

1. Makhoul, J. Linear Prediction: A Tutorial Review, Proc JEEE, 63, No. 4, (1975).

2. Marvel, J., and A. Gray. Linear Prediction of Speech, Springer-Verlag (1976).

DISTRIBUTION LIST

Dr. Marv Atkins
Deputy Director, Science & Tech.
Defense Nuclear Agency
Washington, D.C. 20305

Dr. Robert Cooper [2]
Director, DARPA
1400 Wilson Boulevard
Arlington, VA 22209

The Honorable Richard DeLauer
Under Secretary of Defense (R&E)
Office of the Secretary of Defense
The Pentagon, Room 3E1006
Washington, D.C. 20301

Dr. Alvin M. Despain
1192 Grizzly Peak Boulevard
Berkeley, CA 9470[illegible]

Director [2]
National Security Agency
Fort Meade, MD 20755
ATTN: (b)(3):50 USC §402 Note

CAPT Craig E. Dorman
Department of the Navy, OP-095T
The Pentagon, Room 5D576
Washington, D.C. 20350

CDR Timothy Dugan
NFOIO Detachment, Suitland
4301 Suitland Road
Washington, D.C. 20390

Dr. Larry Gershwin
NIO for Strategic Programs
P.O. Box 1925
Washington, D.C. 20505

Dr. S. William Gouse, W300
Vice President and General Manager
The MITRE Corporation
1820 Dolley Madison Blvd.
McLean, VA 22102

Dr. Edward Harper
SSBN, Security Director
OP-021T
The Pentagon, Room 4D534
Washington, D.C. 20350

Mr. R. Evan Hineman
Deputy Director for Science & Technology
P.O. Box 1925
Washington, D.C 20505

(b)(3):50 USC §403(g) Section 6
CIA/DDS&T
P.O. Box 1925
Washington, D.C. 20505

The MITRE Corporation [25]
1820 Dolley Madison Blvd.
McLean, VA 22102
ATTN: JASON Library, W002

Mr. Jack Kalish
Deputy Program Manager
The Pentagon
Washington, D.C. 20301

Mr. John F. Kaufmann
Dep. Dir. for Program Analysis
Office of Energy Research, ER-31
Room F326
U.S. Department of Energy
Washington, D.C. 20545

Dr. George A. Keyworth
Director
Office of Science & Tech. Policy
Old Executive Office Building
17th & Pennsylvania, N.W.
Washington, D.C. 20500

MAJ GEN Donald L. Lamberson
Assistant Deputy Chief of Staff
(RD&A) HQ USAF/RD
Washington, D.C. 20330

Dr. Donald M. LeVine, W385 [3]
The MITRE Corporation
1820 Dolley Madison Blvd.
McLean, VA 22102

Mr. V. Larry Lynn
Deputy Director, DARPA
1400 Wilson Boulevard
Arlington, VA 22209

Dr. Gordon MacDonald, W300
The MITRE Corporation
1820 Dolley Madison Boulevard
McLean, VA 22102

Dr. Joseph Mangano [2]
DARPA/DEO
9th floor, Directed Energy Office
1400 Wilson Boulevard
Arlington, VA 22209

Mr. John McMahon
Dep. Dir. Cen. Intelligence
P.O. Box 1925
Washington, D.C. 20505

Director
National Security Agency
Fort Meade, MD 20755
ATTN: (b)(3):50 USC §402 Note

Dr. Marvin Moss
Technical Director
Office of Naval Research
800 N. Quincy Street
Arlington, VA 22217

(b)(3):50 USC §403(g) Section 6
P.O. Box 1925
Washington, D.C. 20505

Director
National Security Agency
Fort Meade, MD 20755
ATTN: (b)(3):50 USC §402 Note [4]
DDR-FANX 3

Prof. William A. Nierenberg
Scripps Institution of
Oceanography
University of California, S.D.
La Jolla, CA 92093

Mr. Alan J. Roberts
Vice President & General Manager
Washington C^3 Operations
The MITRE Corporation
1820 Dolley Madison Boulevard
Box 208
McLean, VA 22102

Mr. Richard Ross [2]
P.O. Box 1925
Washington, D.C. 20505

Dr. Oscar S. Rothaus
106 Devon Road
Ithaca, NY 14850

Dr. Phil Selwyn
Technical Director
Office of Naval Technology
800 N. Quincy Street
Arlington, VA 22217

Dr. Eugene Sevin [2]
Defense Nuclear Agency
Washington, D.C. 20305

Dr. Joel A. Snow [2]
Senior Technical Advisor
Office of Energy Research
U.S. DOE, M.S. E084
Washington, D.C. 20585

Mr. Alexander J. Tachmindji
Senior Vice President & General Manager
The MITRE Corporation
P.O. Box 208
Bedford, MA 01730

Dr. Vigdor Teplitz
ACDA
320 21st Street, N.W.
Room 4484
Washington, D.C. 20451

Dr. Al Trivelpiece
Director, Office of Energy Research, U.S. DOE
M.S. 6E084
Washington, D.C. 20585

Mr. James P. Wade, Jr.
Prin. Dep. Under Secretary of Defense for R&E
The Pentagon, Room 3E1014
Washington, D.C. 20301

Mr. Leo Young
OUSDRE (R&AT)
The Pentagon, Room 3D1067
Washington, D.C. 20301