

SENSORY INFORMATION PROCESSING

1 January 1976 - 30 June 1976

Semi-Annual Technical Report

Contractor: University of Utah
Contract Number: DAHC15-73-C-0363
Effective Date: 1 July 1973
Expiration Date: 30 September 1976
Amount of Contract: \$2,825,000.00
Project Code: 3D30

Principal Investigator: Dr. Thomas G. Stockham, Jr.
Telephone: (801) 581-8224

Contracting Officer: Mr. Edgar S. Allen
DSSW

Approved for public release;
distribution unlimited.

Sponsored by
Defense Advanced Research Projects Agency
ARPA Order Number 2477

UTEC-CSc-77-117

TABLE OF CONTENTS

| | Page |
|---|------|
| I. REPORT SUMMARY | 1 |
| II. RESEARCH ACTIVITIES--SENSORY INFORMATION PROCESSING | |
| Section 1. Removing Atmospheric Turbulence Blur | 3 |
| Section 2. Pole-Zero Modeling Using Autocorrelation Prediction | 25 |
| Section 3. Short Time Spectrum Acoustic Processing | 37 |
| Section 4. Log Spectral Estimation for Stationary and Nonstationary Processes | 39 |
| Section 5. Word Recognition in Continuous Speech Using Linear Prediction Analysis | 41 |
| Section 6. Dynamic Filtering of Degraded Speech Using Autocorrelation Prediction | 43 |
| Section 7. Linear Predictive Coding with a Glottal Waveform Model | 48 |
| Section 8. Speech Synthesis Using the Complex Cepstrum | 51 |
| Section 9. Image Understanding | 54 |
| Publications and Presentations | 55 |
| III. FORM DD1473 | 56 |

REPORT SUMMARY

In Section 1 Baxter and Rushforth review the atmospheric turbulence deblurring problem. Based on this review, a research plan is being developed for research in the area.

In Section 2 Atashroo reports on his mathematical techniques for modeling the "zeros" in synthetic speech. This work is continuing so these preliminary technical results are not yet conclusive.

In Section 3 Callahan abstracts the finished work on Short Time Spectrum Acoustic Processing.

In Section 4 Ingebretsen abstracts his work on two estimation parameters for audio signals.

In Section 5 Christiansen abstracts his method for recognizing words in continuous speech.

In Section 6 Peterson and Atashroo discuss a noise filtering method for speech signals.

In Section 7 Done briefly describes an attempt at linear predictive coding using a glottal waveform model. This approach did not improve speech quality sufficiently for reasonable bit rates.

Section 8 reports an attempt to use more waveform phase information in a speech synthesis method.

As indicated in Section 9 Newell will report ongoing efforts in Image Understanding in detail in the next report of this series.

SECTION 1

REMOVING ATMOSPHERIC TURBULENCE BLUR

Brent Baxter

Craig Rushforth

1. Introduction

The removal of the effects of atmospheric turbulence from optical images is a significant problem of long standing. Recent investigations by Knox and Thompson have led to the development of a restoration procedure which shows considerable promise. This procedure has not been successfully applied to real data as yet, however, nor has it been sufficiently well analyzed and simulated to provide a thorough quantitative understanding of their properties. Furthermore, these procedures will very likely require modification before they can be practically applied to large quantities of real data. We have begun an investigation of Knox's method aimed at finding suitable ways to apply it to real data.

1.1. Problem Statement

One of the most difficult and serious problems associated with high-resolution optical imaging systems is the distortion introduced by atmospheric turbulence. If images produced in the presence of atmospheric turbulence are simply recorded and processed by conventional methods, the result is often a loss of essential

high-spatial-frequency detail. This problem is serious even under favorable conditions at a good site, and it is made worse by such effects as wind, solar heating, and boundary-layer turbulence near a moving object. Much of the potential resolving power of a large telescope will be lost unless the received optical field is processed and recorded properly.

There now exist techniques whose potentials for reducing the effects of atmospheric turbulence have been demonstrated theoretically and by simulation. However, much work remains to be done before the promise of these or related techniques can be realized with real data. Foremost among the problems which remain are noise sensitivity and prohibitive computational complexity. A third important issue is related to an assumption which is explicit or implicit in almost all work published to date, namely that the imaging is isoplanatic; i.e., that the instantaneous point-spread function of the atmosphere-telescope system is the same for all points in the object. This will usually be true only for objects of very limited extent. Thus, appropriate extensions will have to be considered if these techniques are to be applied to objects larger than a single isoplanatic patch.

There is a final issue which often fails to receive the attention that it deserves, namely the skilled and careful handling and recording of images using proper techniques and

equipment. Using the wrong type of film or developing process, improper gray scale compensation, or a careless scanning and digitizing procedure may significantly detract from the effectiveness of a procedure for removing turbulence blur. Furthermore, it is difficult to assess the real merit of a technique whose results are presented using carelessly-prepared photographs. Unfortunately, these problems occur more often than one might expect.

2. Summary of Previous Work

In this section, we provide a brief summary of previous work on the problem of restoring images which have been blurred by atmospheric turbulence. We make no attempt to be comprehensive, but rather restrict our attention to work which is immediately relevant to our problem. For a more complete discussion of other work, and for a comprehensive bibliography, see [1].

2.1. Long-Time Averaging

The problem in which we are interested is characterized by an object intensity distribution which is constant for a long period of time, during which the conditions of the atmosphere change appreciably. One obvious approach to removing the effects of random atmospheric turbulence is to perform a long-time average of the image intensity distribution. This type of averaging can be effected either

by exposing film or another recording medium for a long time, or by averaging many short-time exposures. In either case, the averaging has the effect of removing the random fluctuations from the resulting image. Figure 1 shows such an image.

Unfortunately, this long-time averaging also has the effect of removing or greatly attenuating the high spatial frequencies from the image [11]. Thus, although the average image is stable, its spatial-frequency content is limited by atmospheric "seeing" rather than by the imaging system. Furthermore, only limited improvement can be achieved by applying standard restoration techniques to the long-time average image. As an example of the seriousness of this effect, consider the 200-inch Hale telescope. The diameter of the Airy disk for this telescope is about 0.05 arc seconds, while a typical seeing limit on a quiet night is about 2 arc seconds. Thus, if long-time averaging is employed, atmospheric turbulence degrades the resolving power of this system by a factor of about 40. If high-resolution imaging systems are to reach their potential in the presence of atmospheric turbulence, then, it is clear that alternative methods which yield statistical stability while preserving high spatial frequencies are required.

A partial solution to this problem was obtained by Labeyrie [12, 13], whose technique has come to be known as speckle interferometry because of the speckled character of the images used. An example of such an image is shown in Figure 5. In this technique, a series of short-time exposures of the atmospherically-degraded object are taken through a narrow-band spectral filter. The exposure time is short enough that the atmosphere is effectively "frozen" during the exposure, and successive photographs are sufficiently separated in time that they may be taken to be statistically independent for the purpose of averaging.

Instead of simply averaging these short-time exposures directly, a process which leads to a loss of high spatial frequencies as we have seen, speckle interferometry requires some additional operations before averaging. First, the Fourier transform of each image is taken, either optically or digitally. The averaging is then performed on the squared magnitude of these individual Fourier transforms.

The process of taking the square of the magnitude destroys the information regarding the phase of the image. In this fact lies both the strength and the weakness of speckle interferometry. Since it is the random phases of the high-spatial-frequency components which cause their mutual cancellation in the process of long-time averaging, the suppression of phase prevents this cancellation and thus preserves some high-frequency information which would be

lost under direct long-time averaging. Unfortunately, information about the phase of the object is also suppressed. This loss of phase information limits the class of objects to which speckle interferometry can be usefully applied. For example, this technique can yield useful information about symmetric objects or binary star systems, but not about an object with an arbitrary intensity distribution or arbitrary shape.

Labeyrie and his colleagues have verified experimentally that speckle interferometry yields essentially diffraction-limited information about such objects as binary star systems. In addition, theoretical discussions of this technique have been provided by Miller and Korff [14] and by Korff [15]. These authors have used models of propagation in the turbulent atmosphere to explain why speckle interferometry works and to generalize the results to partially-coherent objects.

2.3. The Knox-Thompson Technique

The success of speckle interferometry in reducing the effects of atmospheric turbulence, coupled with its limited applicability, has spurred a search for generalizations and improvements of this technique. A significant extension of speckle interferometry was developed by Knox and Thompson in 1974 [1]. This new procedure uses speckle interferometry as before to obtain amplitude information. To this amplitude information is added phase information obtained by averaging the Fourier transforms of the short-time exposures in a different way.

We denote the Fourier transform of the short-exposure image intensity distribution by $\tilde{I}(u)$. The quantity calculated in speckle interferometry, from which the amplitude information can be obtained, is $\langle |\tilde{I}(u)|^2 \rangle$, where the pointed brackets denote an average over many images. Knox and Thompson use this same procedure to obtain amplitude information.

Knox and Thompson have shown that differential phase may be obtained from the quantity $\langle \tilde{I}(u_1) \tilde{I}^*(u_2) \rangle$, where u_1 and u_2 are spatial frequencies differing from each other by only a small amount, and the asterisk denotes the complex conjugate. Differential phases can then be added to produce an object phase distribution which, together with the object amplitude (from speckle interferometry), characterize the

object intensity distribution out to the diffraction limit of the telescope. This technique thus offers the potential of greatly improved resolution for objects with arbitrary intensity distributions.

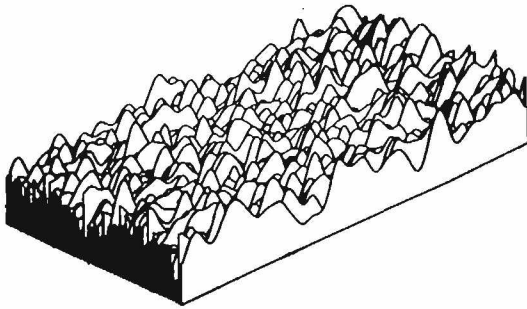
Unfortunately, the effectiveness of the Knox-Thompson technique has not as yet been demonstrated on real data. Simulations with data free of sensor noise performed by Knox [1], by ourselves, and by others [16] have led to results in general agreement with predictions, however, and there is reason to believe that the technique will work reasonably well on real images, provided that they are of high quality (noise free) and that there are enough of them to allow adequate averaging. There are indications that the technique may be quite sensitive to noise, however [2]. This concern has led to a search for alternatives which are more robust. One such approach is discussed in the next section.

Before we turn to that discussion, we briefly summarize the results of a one-dimensional simulation of the Knox-Thompson technique which we performed here at the University of Utah. Figure 2a shows a collection of sample functions of random wavefront phase induced by atmospheric turbulence according to the physical model described in [1]. A typical point spread function of the atmosphere-telescope combination is shown in Figure 2b. Figure 2c shows the intensity distribution of the object to be imaged, and

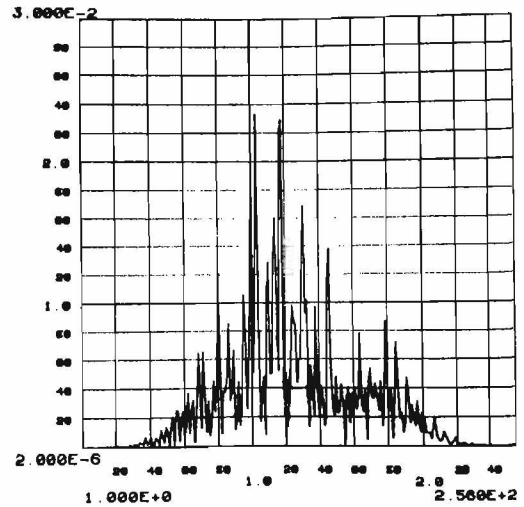
Figure 2d shows the blurred image obtained by convolving Figure 2c with Figure 2b.

Figure 3a shows the result of averaging 64 point spread functions similar to that shown in Figure 2b. Figures 3b and 3c show the RMS image Fourier transform and the squared magnitude of the object Fourier transform. Figures 4a and 4b illustrate the phase reconstruction process, and finally Figure 4c shows the estimate of the object intensity distribution obtained using the Knox-Thompson technique. Note the good agreement with Figure 2c.

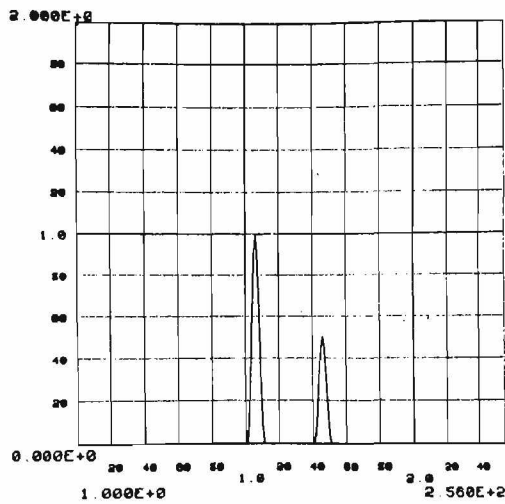
The simulation described above did not include the effects of sensor noise. Although we have not as yet attempted to quantify the effects of sensor noise on the performance of the Knox-Thompson method, we have some preliminary indications that tend to support Sherman's [2] contention that these effects can be quite severe. These indications were obtained during a preliminary attempt to apply the Knox-Thompson procedure to some photographs at the Kitt Peak Observatory. We believe that our failure to achieve satisfactory results in this preliminary experiment was due in large measure to the poor quality of the data, particularly the large amount of sensor noise present.



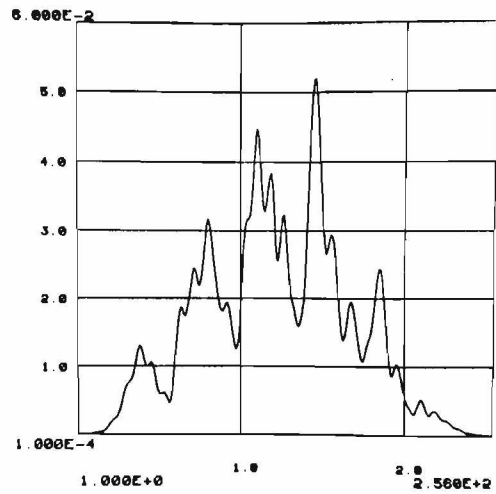
(a)



(b)



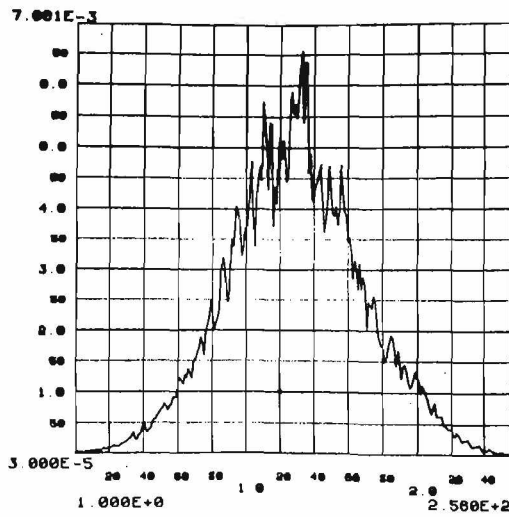
(c)



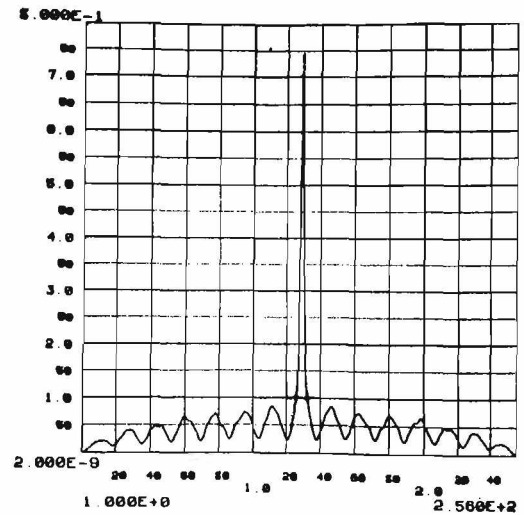
(d)

FIGURE 2

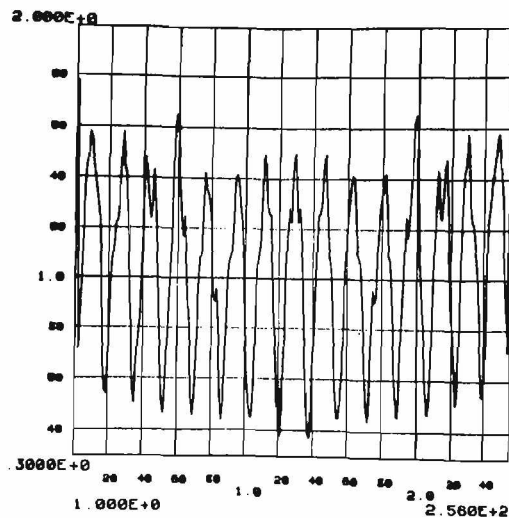
- One dimensional simulated atmospheric turbulence blur data.
- (a) Random phase at receiving aperture due to fluctuations atmospheric density. Distance across aperture upward and toward right. Successive images upward and toward left.
 - (b) A single point spread function. Note the speckle like character.
 - (c) Object intensity distribution.
 - (d) Turbulence blurred image.



(a)



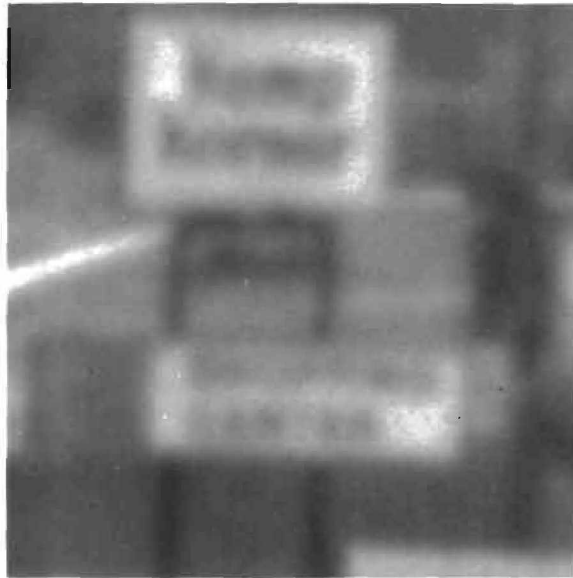
(b)



(c)

FIGURE 3

- Various types of averaging on multiple simulated images.
- (a) The average of 64 blurred images is similar to a single long exposure image. Note how the detail is lost.
 - (b) RMS average of the Fourier transform. Note the presence of energy at spatial frequencies from DC (center) to the diffraction limit. Cosine fringes suggest a double object.
 - (c) Object squared magnitude Fourier transform.



(a)



(b)

FIGURE 1

Long exposure turbulence.

- (a) Two second exposure made through a questar telescope.
- (b) Attempts to recover the high spatial frequency information result some improvement accompanied by amplification of film grain noise.

2.4. Real-Time Phase Compensation

An entirely different approach to the reduction of atmospheric turbulence effects is to try to deal with these effects before an image is recorded rather than after. In many circumstances, the effects of atmospheric turbulence can be modeled as a phase perturbation of the received wavefront at the imaging aperture. If this phase perturbation were known, or if an accurate estimate could be obtained, the effects of turbulence could be removed by phase compensation. Several investigators have considered the possibility of implementing real-time phase-compensation systems [4, 6]. In these systems, measurements of optical phase perturbation are made over the objective pupil, either explicitly or implicitly, with the result being used to control a deformable mirror or some other phase-correcting device to compensate for atmospheric turbulence in real time before the image is recorded. Such systems avoid the heavy burden of digital computation which we have described above in connection with the Knox-Thompson and Sherman techniques, but on the other hand they require more complicated optical equipment and control systems. Furthermore, the accuracy of phase compensation systems is limited by phase-estimation errors due to the finite wavefront-sensor signal-to-noise ratio and by fitting errors due to the finite number of spatial modes restored by the wavefront corrector [6]. Thus, we believe that further extensions and improvements of

both phase compensation systems and post-processing techniques of the type we have discussed above should be undertaken.

2.5. The Work of Shapiro

In an interesting series of paper [6, 9], Shapiro has established what appears to be a very useful framework for the analysis and design of optical imaging systems which attempt to remove the effects of atmospheric turbulence. In [9], he develops a normal-mode decomposition for the turbulent atmosphere which is similar to that developed previously for free-space imaging by Rushforth and Harris [17] and others. He then applies these results to a study of the ultimate performance limits on imaging through a turbulent atmosphere [6, 8]. In agreement with others, he shows that it is possible in principle to achieve diffraction-limited imaging.

Finally, Shapiro in [7] considers the conditions under which diffraction-limited imaging may be achieved even when the extent of the object is such that the imaging is no longer isoplanatic; i.e., when the atmospheric point spread function depends upon the position of the point source in the object plane. Since all the work discussed in Sections 2.1 through 2.5 is based on the assumption of isoplanatic imaging, and since in many practical situations the imaging will not be isoplanatic, Shapiro's results are of

considerable interest and importance.

In effect, what Shapiro shows is that in many cases of interest, the object can be broken up into distinct isoplanatic elements such that the contribution of each element to the image can be separated from the others. In principle, this enables us to deal with each isoplanatic element separately and then put the individual images back together to form the total image. In practice, this procedure may be formidable in its complexity, but it is nevertheless of interest to know that conditions frequently exist under which it is possible in principle.

In the process of obtaining his results on nonisoplanatic imaging, Shapiro [7] develops a model for the atmosphere which we intend to build upon in our work. He points out that the problem of optical imaging through a turbulent atmosphere shares many features with the problem of transmitting and receiving information at radio frequencies through random media such as tropospheric scatter channels. In fact, he shows that the turbulent atmosphere can be modeled as a wide-sense stationary, uncorrelated scatter (WSSUS) channel, which in many respects is the simplest and most useful scatter-channel model [10]. Shapiro modifies and applies the results of [10] to show that when the WSSUS channel is underspread (i.e., the delay-Doppler product is less than one), the effects of the various isoplanatic elements can be separated as described

above. This condition frequently occurs in practice.

Thus, we have an example in which a model from random-channel communication theory has proven very useful in the analysis of optical imaging in atmospheric turbulence. In addition, Shapiro has shown that this model can be applied to optical communication systems and to speckle interferometry. We believe that this model, with appropriate modifications, will be very useful in the work which we are proposing. We shall say more about this in the next section, where we make specific proposals to utilize this model.

To summarize, we have described in this section several methods which have the potential to achieve high-resolution imaging in the presence of atmospheric turbulence. Considerable work remains before this potential can be realized in a practical system, however. In particular, the issues of noise sensitivity, computational complexity, and nonisoplanatic imaging must be investigated more carefully. Since sensor noise will be present in any practical system, it is essential that we understand how this noise affects various restoration procedures, and that we know how to minimize these effects. The issue of computational complexity may determine whether a technique is hopelessly impractical or can be made operational. Finally, a system which is to be of use in many practical situations must be able to deal with large objects for which the imaging will

very likely be nonisoplanatic. We expect the communication theory model described in this section to be useful in these investigations as well as in suggesting alternative or modified procedures with improved performance and efficiency.

REFERENCES

- [1] Keith T. Knox, "Diffraction-Limited Imaging with Astronomical Telescopes", Ph.D. dissertation, Institute of Optics, University of Rochester, New York, 1975.

- [2] James W. Sherman, "A Posteriori Restoration of Atmospherically Degraded Images Using Multiframe Imagery", Topical Meeting on Image Processing, Optical Society of America, Pacific Grove, California, February 24-26, 1976.

- [3] C.R. Lynds, S.P. Worden, and J.W. Harvey, "Digital Image Reconstruction Applied to Alpha Orionis", Astrophysical Journal, July 1, 1976.

- [4] R.A. Muller and A. Buffington, "Real-Time Correction of Atmospherically Degraded Telescope Images through Image Sharpening", Journal of the Optical Society of America, September 1974, pp. 1200-1210.

- [5] A.M. Schneiderman, P.F. Keller, and M.G. Miller, "Laboratory Simulated Speckle Interferometry", Journal of the Optical Society of America, November 1975, pp. 1287-1291.

- [6] J.H. Shapiro, "Propagation-Medium Limitations on Phase-Compensated Atmospheric Imaging", Journal of the Optical Society of America, May 1976.

- [7] J.H. Shapiro, "Diffraction-Limited Atmospheric Imaging of Extended Objects", Journal of the Optical Society of America, May 1976, pp. 469-477.

- [8] J.H. Shapiro, "Optimum Adaptive Imaging through Atmospheric Turbulence", Applied Optics, November 1974, pp. 2609-2613.

- [9] J.H. Shapiro, "Normal-Mode Approach to Wave Propagation in the Turbulent Atmosphere", Applied Optics, November 1974, pp. 2614-2619.
- [10] R.S. Kennedy, "Fading Dispersive Communication Channels", John Wiley and Sons, Inc., 1969.
- [11] R.E. Hufnagel and N.R. Stanley, "Modulation Transfer Function Associated with Image Transmission through Turbulent Media", Journal of the Optical Society of America, January 1964, pp. 52-61.
- [12] A. Labeyrie, "Attainment of Diffraction Limited Resolution in Large Telescopes by Fourier Analyzing Speckle Patterns in Star Images", Astronomy and Astrophysics, Vol. 6, 1970, pp. 85-87.
- [13] D.Y. Gezari, A. Labeyrie, and R.V. Stachnik, "Speckle Interferometry: Diffraction-Limited Measurements of Nine Stars with the 200-Inch Telescope", Astrophysical Journal, April 1, 1972, pp. L1-L5.
- [14] M.G. Miller and D. Korff, "Resolution of Partially Coherent Objects by Use of Speckle Interferometry", Journal of the Optical Society of America, February 1974, pp. 155-161.
- [15] D. Korff, "Analysis of a Method for Obtaining Near-Diffraction-Limited Information in the Presence of Atmospheric Turbulence", Journal of the Optical Society of America, August 1973, pp. 971-980.
- [16] "A Recent Result in the Restoration of Turbulence-Degraded Imagery", ESL, Inc., Sunnyvale, California, May 12, 1975.

- [17] C.K. Rushforth and R.W. Harris, "Restoration, Resolution, and Noise", Journal of the Optical Society of America, April 1968, pp. 539-545.

- [18] J.W. Goodman and J.F. Belsher, "Fundamental Limitations in Linear Invariant Restoration of Atmospherically Degraded Images", Technical Note, Department of Electrical Engineering, Stanford University, Stanford, California, 1975.

- [19] J.C. Dainty and R. Shaw, "Image Science", Academic Press, 1974.

- [20] R.S. Kennedy, "On the Atmosphere as an Optical Communication Channel", IEEE Transactions on Information Theory, September 1968, pp. 716-725.

SECTION 2

POLE-ZERO MODELING USING AUTOCORRELATION PREDICTION

M. Ali Atashroo

Steven F. Boll

Abstract

This paper describes a method for estimating the poles and zeros of a rational filter whose spectrum matches in the linear prediction sense, a given prototype spectrum. The filter parameters are calculated from systems of linear equations that result from linear prediction on autocorrelation functions. It is shown that this method is equivalent to spectral matching by inverse filtering. The matching process is carried out in three steps. In step one, the standard linear prediction analysis procedure is modified to account for zeros present in the matching filter. The denominator coefficients (pole-predictors) are then calculated from the resulting system of autocorrelation equations. In step two, the spectral contribution to these poles is removed from the given prototype spectrum by convolution yielding a residual spectrum. In step three, the numerator coefficients, (zero-predictors) are calculated by first matching the residual spectrum with a high order all pole spectrum, then matching the inverse of the latter with the desired all-zero filter. This work has been

expanded and published in technical report UTEC-CSc-76-271,
Atashroo, M. "Pole-Zero Modeling and its Applications to
Speech Processing", U. of U. Computer Science Dept.

Spectral Matching By Inverse Filtering

The spectrum of a rational filter, $H(z)$ can be matched to a given spectrum $|S(\omega)|^2$ by requiring that when $|S(\omega)|^2$ is filtered by the inverse filter, $H(z)^{-1}$, the output spectrum is flat. That is, the filter coefficients of $H(z)$ are computed in order to satisfy the requirement that

$$|S(\omega)|^2 \cdot \frac{1}{|H(\omega)|^2} = 1 \quad (1)$$

This approximation technique is called spectral matching by inverse filtering, and defines the frequency domain interpretation of pole-zero modeling by autocorrelation prediction.

All-Pole Spectral Matching

Applying this interpretation to the approximation of a given power spectrum $|S(\omega)|^2$ by an all-pole spectrum, $G_A^2/|A(\omega)|^2$ gives

$$|S(\omega)|^2 |A(\omega)|^2 = G_A^2 \quad (2)$$

The filter coefficients $\{a_k\}$ and gain G_A are computed in order to satisfy equation 2. Expanding gives

$$S(\omega)^2 A(\omega) = \frac{G_A^2}{A^*(\omega)} \quad (3)$$

But

$$\begin{aligned} |S(\omega)|^2 &\leftrightarrow R_S(n) && : \text{ autocorrelations} \\ A(\omega) &\leftrightarrow a(n) && : \text{ predictors} \\ \frac{G_A^2}{A^*(\omega)} &\leftrightarrow G_A h(-n) && : \text{ scaled, all-pole response} \end{aligned}$$

In the time domain, equation (3) implies

$$R(n) * a(n) = G_A h(-n), \quad h(0) = G_A \quad (4)$$

Since the impulse response is causal, values for the predictors, $\{a(n)\}$ and gain G_A are computed from the well-known autocorrelation analysis equations:

$$\begin{aligned} \sum_{k=0}^M a(k)R(n-k) &= 0 && n > 0 \\ \sum_{k=0}^M a(k)R(k) &= G_A^2 && n = 0 \end{aligned} \quad (5)$$

Pole-Zero Spectral Matching

Consider now matching a given power spectrum $|S(\omega)|^2$ by the spectrum of a rational filter $H(z)$, having both poles and zeros:

$$H(z) = \frac{B(z)}{A(z)} = \frac{\sum_{i=0}^L b(i)z^{-i}}{\sum_{i=0}^M a(i)z^{-i}} \quad (6)$$

Applying the inverse filtering interpretation requires that the filter coefficients $\{a(i)\}$ and $\{b(i)\}$ be computed such that

$$|S(\omega)|^2 \frac{|A(\omega)|^2}{|B(\omega)|^2} = 1 \quad (7)$$

Expanding equation (7) and examining the time domain relation gives

$$|S(\omega)|^2 A(\omega) = B(\omega) \frac{B^*(\omega)}{A^*(\omega)} = B(\omega)H^*(\omega) \quad (8)$$

Thus

$$R(n)*a(n) = b(n)*h(-n) \quad (9)$$

or

$$R(n) = -\sum_{k=1}^M a(k)R(n-k) + \sum_{k=0}^L b(k)h(k-n) \quad \text{all } n \quad (10)$$

For $n > L$ the presence of the zeros in equation (10) is removed by the causality of $h(n)$, thus

$$R(n) = -\sum_{k=1}^M a(k)R(n-k) \quad n > L \quad (11)$$

The relation given in equation (11) is referred to as the Generalized Linear Prediction normal equations, which are solved to obtain the all-pole predictors $\{a(k)\}$.

Pole Removal By Convolution

Before estimating the zeros of $H(z)$, the effect of the approximated poles as defined by the filter $A(z)$ is removed from $|S(\omega)|^2$ by convolution.

Define

$$|D(\omega)|^2 = |S(\omega)|^2 |A(\omega)|^2 \quad (12)$$

then

$$R_D(n) = R_S(n) * R_A(n) \quad (13)$$

where

$$R_S(n) \leftrightarrow |S(\omega)|^2$$

$$R_A(n) \leftrightarrow |A(\omega)|^2$$

$$R_A(n) = \sum_{k=0}^M a(k)a(k+n)$$

In a somewhat simplified sense $|S(\omega)|^2$ can be described as being composed of its spectral poles and zeros. With the removal of the all-pole match to $|S(\omega)|^2$ by $1/|A(\omega)|^2$, the spectrum $|D(\omega)|^2$ can be considered as the residual spectrum containing spectral zero information. These spectral zeros are to be matched by the all-zero filter, $B(z)$.

All-Zero Approximation to $|D(\omega)|^2$

The all-zero approximation is done in two steps, first $|D(\omega)|^2$ is approximated by a high order all-pole filter $G_Q/Q(z)$. Second, after matching $|D(\omega)|^2$ by $G_Q^2/|Q(\omega)|^2$, the all-zero filter $B(z)$ is computed by matching $|B(\omega)|^2$ to the approximated inverse of $|D(\omega)|^2$ given by $|Q(\omega)|^2/G_Q^2$.

To obtain the coefficients of $Q(z)$, the inverse filtering relation is applied. Thus

$$|D(\omega)|^2 \frac{|Q(\omega)|^2}{G_Q^2} = 1$$

Using the same procedure as with standard Linear Prediction, the analysis equations reduce to

$$\sum_{k=0}^N R_D(n-k)q(k) = 0$$

$$\sum_{k=0}^N R_D(k)q(k) = G_Q^2$$

Finally, the all-zero predictors $\{b_k\}$ are computed using the relation

$$|Q(\omega)|^2 |B(\omega)|^2 = G_B^2$$

The analysis equation is given by

$$\sum_{k=0}^L b(k)R_Q(n-k) = 0 \quad n > 0$$

$$\sum_{k=0}^L b(k)R_Q(k) = G_B^2 \quad n = 0$$

where

$$R_Q(k) = \sum_{n=0}^N q(k)q(n+k)$$

Discussion

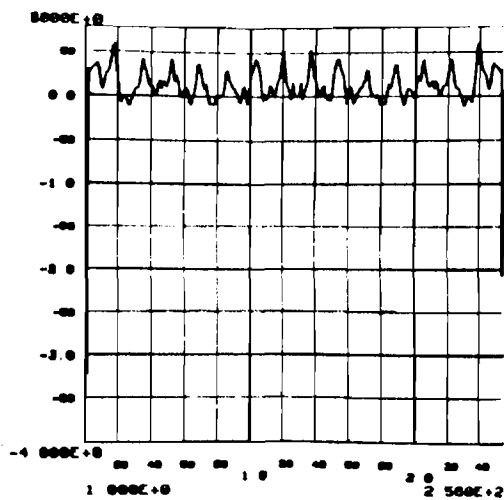
As is shown in the analysis block diagram (Figure 1), the analysis method uses standard linear prediction to estimate the zeros and so-called generalized linear prediction to estimate the poles. Unfortunately, with this generalized method, stability of the filter cannot be guaranteed. There are at least two options for guaranteeing stability, one, estimate the pole-predictors with standard linear prediction analysis or two, initially match the pole-zero filter to the inverted spectrum, then use the reciprocal of the resulting filter.

Some of the advantages of this method for spectral matching are that the analysis is carried out in the time domain requiring only the solution of linear equations. More important, since linear prediction can be used, not only can both poles and zero be optimally coded using reflection coefficients, but the filter can be implemented by simply cascading a normalized lattice filter with an inverted normalized lattice filter.

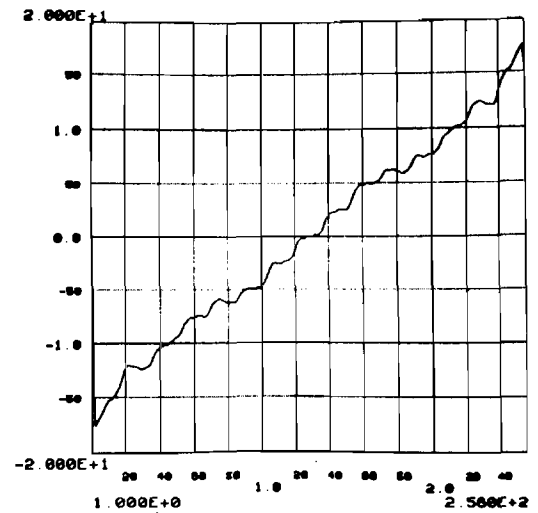


FIGURE 5

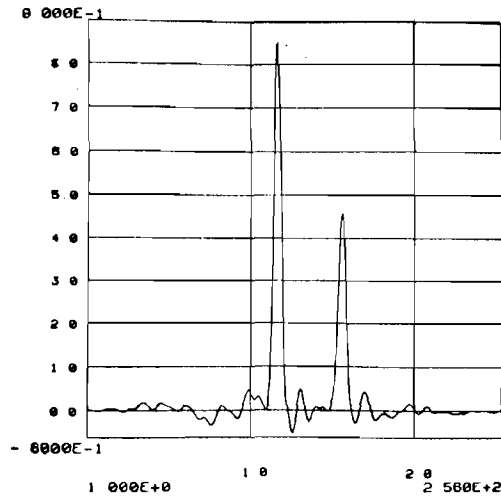
Speckle image of Arcturus made on the Mayall four-meter telescope at Kitt Peak. Delays imposed on the wavefront by its passage through the atmosphere cause speckles whose size is characteristic of the diffraction limit of the telescope.



(a)



(b)



(c)

FIGURE 4

Phase estimation using Knox's method.

- (a) Phase differences are small compared with 2π preventing any ambiguity in the averaging process.
- (b) Reconstructed phase.
- (c) Object intensity distribution estimated by computing the inverse Fourier transform of 3(c) and 4(b).

SECTION 3

Short Time Spectrum Acoustic Processing

Michael Wayne Callahan

The frequency domain representation of a time signal afforded by the Fourier transform is a powerful tool in acoustic signal processing. The usefulness of this representation is rooted in the mechanisms of sound production and perception. Many sources of sound exhibit normal modes or natural frequencies of vibration, and can be described concisely in the frequency domain. The human auditory system performs frequency analysis early in the hearing process, so perception is often best described by frequency domain parameters.

The published technical report UTEC-CSc-76-209, March 1976, "Acoustic Signal Processing Based on the Short Time Spectrum" by M. Callahan, U. of U. Computer Science Dept., investigates a new approach to acoustic signal processing based on the short-time Fourier transform, a two-dimensional representation which shows the time and frequency structure of sounds. This representation is appropriate for signals such as speech and music, where the natural frequencies of the source change and timing of these changes is important to perception. The principal advantage of this approach is that the signal processing domain is

similar to the perceptual domain, so that signal modifications can be related to perceptual criteria.

The mathematical basis for this type of processing is developed, and four examples are described: removal of broadband background noise, isolation of perceptually important speech features, dynamic range compression and expansion, and removal of locally periodic interfering signals.

SECTION 4

Log Spectral Estimation for Stationary
and Nonstationary Processes

Robert B. Ingebretsen

This research is concerned with two log spectral estimators in the context of both stationary and nonstationary signals. They differ because in one, smoothing is realized before the logarithmic transformation, while the other is smoothed in the logarithmic domain. It is shown that for stationary signals the two estimators are similar, differing in expected value by only a universal constant. The first estimator, however, is smoother. For nonstationary signals, the estimators are biased by different amounts dependent upon the nonstationarity. The difference between the estimators is shown to be a sensitive test for nonstationarity. The estimators are used in the analysis and implementation of two solutions to the problem of blind deconvolution. It is found that the methods are equivalent for stationary signals, but differ markedly for nonstationary signals in the presence of stationary background noise. Recommendations are made for the practical digital implementation of the log spectral estimators. This research has been published in the technical report UTEC-CSc-75-118, August 1976, "Log Spectral

Estimation for Stationary and Nonstationary Processes" by R.
Ingebretsen, U. of U. Computer Science Dept.

SECTION 5

Word Recognition in Continuous Speech

Using Linear Prediction Analysis

Richard W. Christiansen

A promising method of automatic word recognition in continuous speech, recently designated as word spotting, has been demonstrated. The method uses error residual ratios from LPC (Linear Predictive Coding) vocoder analysis for waveform comparison and a dynamic programming procedure for time registration between the incoming speech and a template of the key word. Using a similarity threshold, the incoming speech is compared with several templates to account for variability in spectral shape. This system can work in real time using a real time vocoder.

The multiple templates are used in such a way that a small number of templates, three or four, is made to look like several hundred or more. This is accomplished by dynamically constructing a composite template from parts of each single template as part of the processing of the incoming speech. Thus a particular composite template is constructed for each word being recognized.

An accuracy of 99 percent with no false alarms was achieved using 205 key words, five different speakers, and approximately ten minutes of speech text. Performance in the presence of additive white Gaussian noise of approximately 11dB signal-to-noise ratio was 66 percent. When the speech was processed to account for the noise, results improved to 85 percent to 90 percent accuracy. Finally a digit recognition experiment was performed using over 1200 digits spoken by ten different people with a resultant accuracy of 97 percent. This research has been published in the technical report UTEC-CSc-76-226, August 1976, "Word Recognition in Continuous Speech using Linear Prediction Analysis" by R. Christiansen, U. of U. Computer Science Dept.

SECTION 6

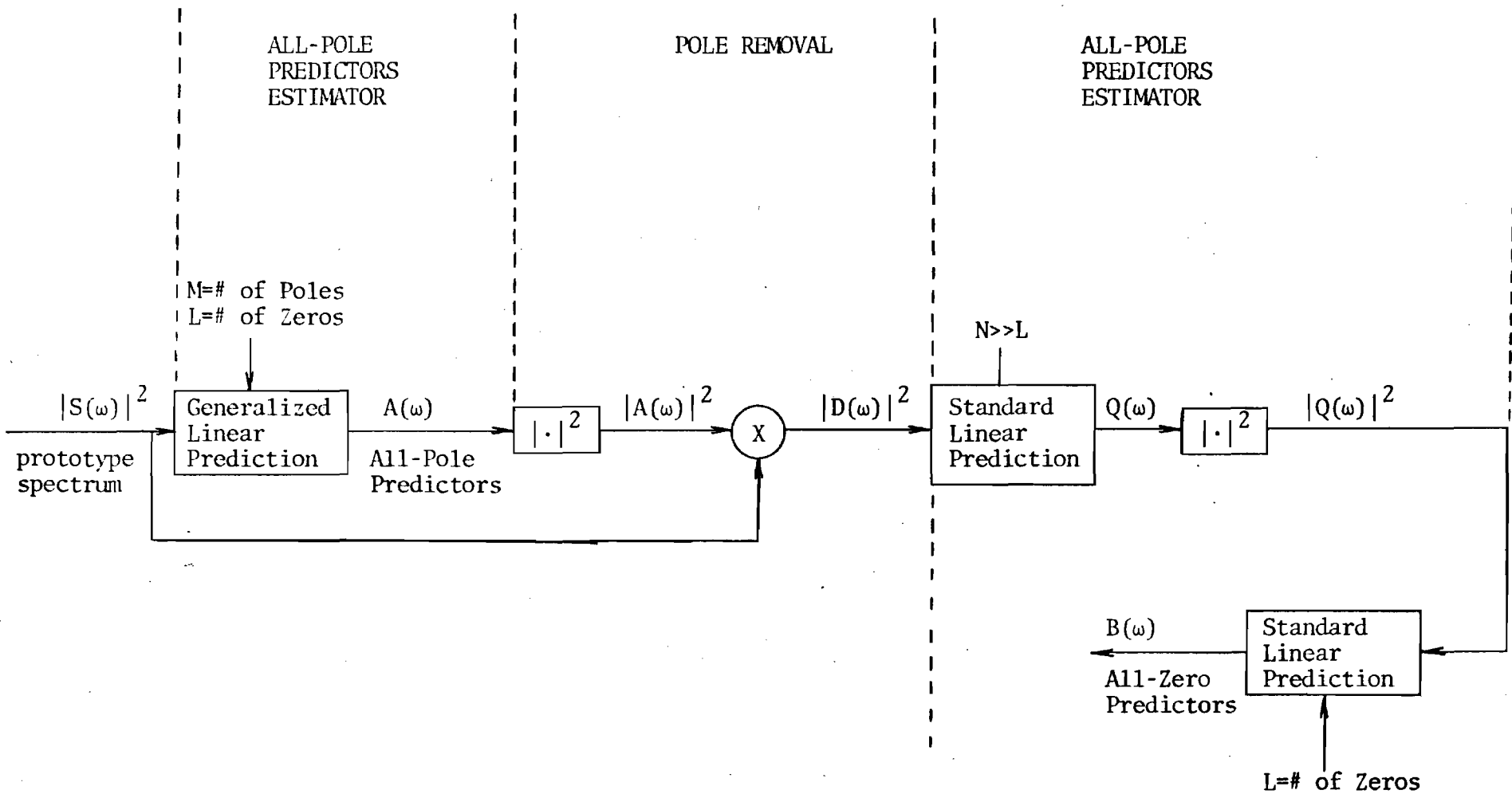
Dynamic Filtering of Degraded Speech
Using Autocorrelation Prediction

M. Ali Atashroo

Tracy Petersen

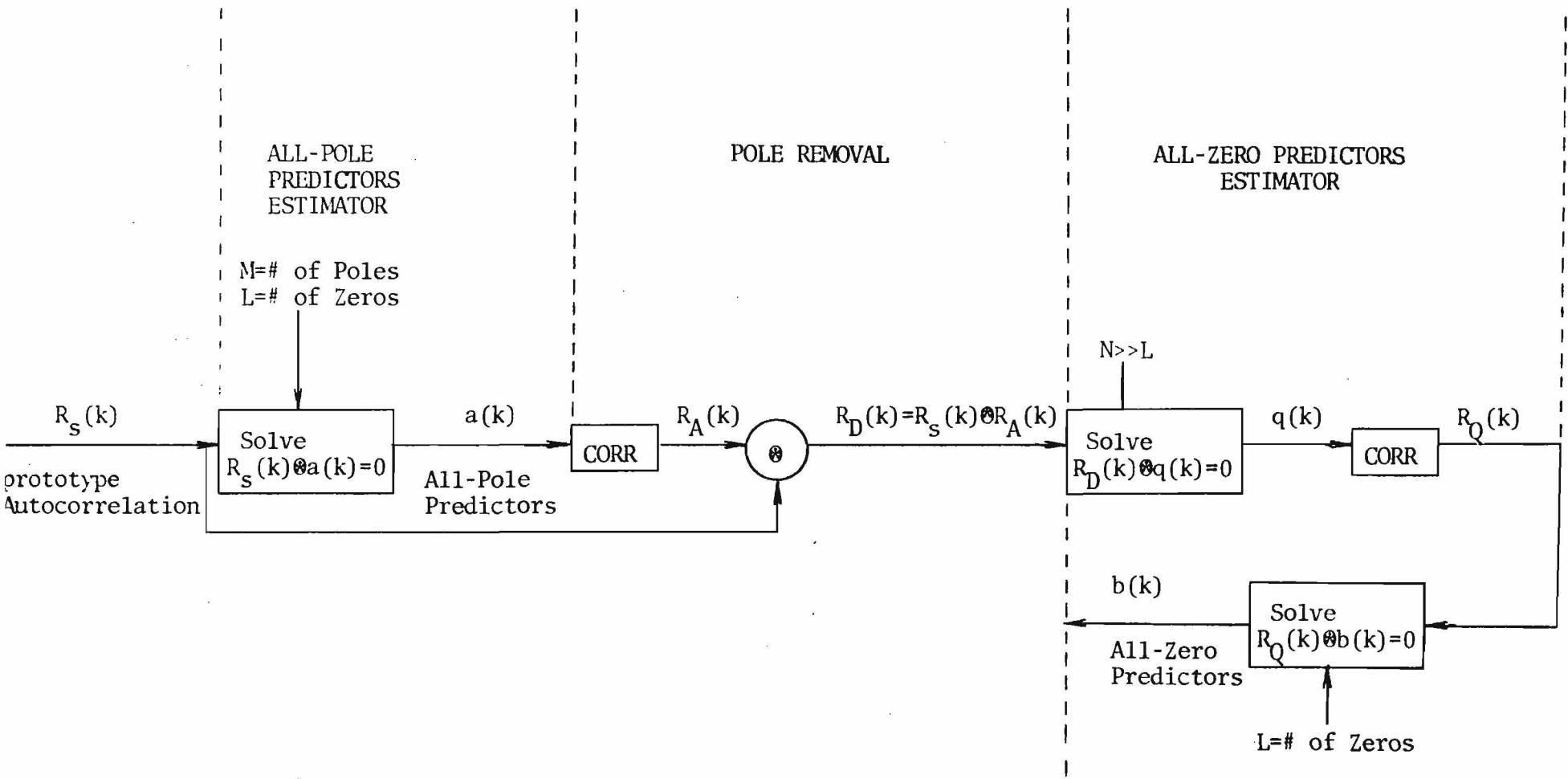
Abstract

A dynamic smoothed wiener filter is estimated from the degraded speech. Then, this dynamic smoothed wiener filter is approximated to desired accuracy with a time-varying rational filter using autocorrelation prediction [1]. This time-varying rational filter is specified with two sets of parcor parameters, and is realized as a cascade of two lattice filters. Finally, the degraded speech is filtered by this realization. This work has been expanded and published in technical report UTEC-CSc-76-271, 1976, "Pole Zero Modeling and its Applications to Speech Processing", U. of U. Computer Science Dept.



ANALYSIS BLOCK DIAGRAM

FIGURE 1 a



AUTOCORRELATION PREDICTION
BLOCK DIAGRAM.

FIGURE 1 b

Introduction

The speech is degraded with additive stationary colored noise. Assuming the speech itself is stationary during short intervals ($T = 30$ msec), a smoothed wiener filter $W(kT, \omega)$ is estimated for each of these intervals. The spectrum of a rational filter $H(kT, \omega)$ is then matched to the $1/|W(kT, \omega)|^2$ using Autocorrelation Prediction [1]. Finally, the rational filter $1/H(kT, \omega)$ is used to filter the corresponding interval of the degraded speech.

Estimation of the Smoothed Wiener Filter

The smoothed wiener filter for each interval T is estimated from the following equation.

$$|W(kT, \omega)| = 1 - \frac{\hat{\phi}_N(\omega)}{\hat{\phi}_{S+N}(kT, \omega)} \quad (1)$$

where, $\hat{\phi}_N(\omega)$ is an estimate for the smoothed spectrum of the stationary noise and $\hat{\phi}_{S+N}(kT, \omega)$ is an estimate for the smoothed spectrum of the degraded speech during the k th interval. $\hat{\phi}_N(\omega)$ is obtained by averaging the short-time Linear Prediction (LP) spectrum of the noise - the silent portions of the degraded speech during the whole passage. $\hat{\phi}_{S+N}(kT, \omega)$ is obtained by averaging the short-time LP spectrum of the degraded speech during the k th interval.

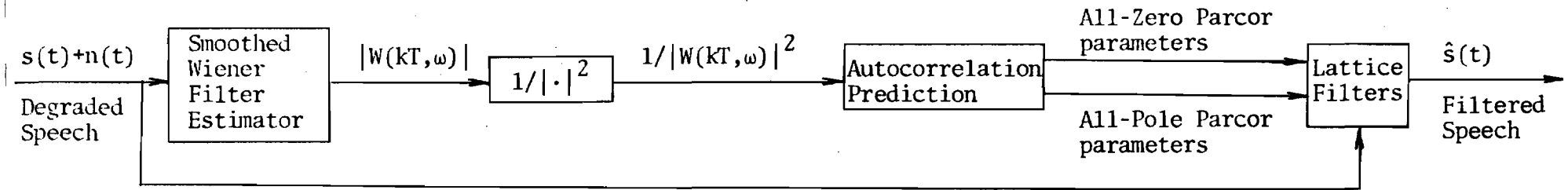
Spectral Matching of Smoothed Wiener Filter

Equation 1 shows that the smoothed wiener filter has flat pass bands for those frequencies where the speech dominates the noise, and has deep stop bands at those frequencies where the noise dominates the speech. In the case where speech dominates the noise at most of the frequencies, this suggests to match the valleys of the smoothed wiener spectrum first. This can be accomplished by matching the spectrum of a rational filter $H(kT, \omega)$ to the $1/|W(kT, \omega)|^2$ using Autocorrelation Prediction. Then, the rational filter $1/H(kT, \omega)$ is used to filter the k th interval of the degraded speech. In this way, the computation and the number of parameters of the matching rational filter $H(kT, \omega)$ is reduced considerably from an all-pole match to $|W(kT, \omega)|^2$.

While the stability may not be guaranteed for the rational filter $H(kT, \omega)$, it is guaranteed for $1/H(kT, \omega)$ [1].

Implementation and Results

The rational filter $1/H(kT, \omega)$ is identified by two sets of the parcor parameters obtained from the Autocorrelation Prediction. The $1/H(kT, \omega)$ is realized as the cascade of two lattice filters - inverse and all-pole. To improve the filtering process both sets of the parcor parameters are interpolated for each sample of the input degraded speech.



DYNAMIC FILTERING BLOCK DIAGRAM

REFERENCES

- [1] M. Ali Atashroo, S.F. Boll, "Pole-Zero Modeling Using Autocorrelation Prediction", presented at 1976 Arden House Workshop.

SECTION 7

Linear Predictive Coding With a Glottal Waveform Model

William Done

Speech analysis using the linear prediction technique, based on the approximation of the n th speech sample $s(n)$ as a summation of the N previous, linearly-weighted samples, achieves good spectral matching to the poles in the speech spectrum. Zeros of the spectrum are not matched as well.

To better model the zeros, it is assumed that the error signal

$$\begin{aligned} e(n) &= s(n) - \hat{s}(n) \\ &= s(n) - \sum_{i=1}^N a(i)s(n-i) \end{aligned}$$

could be modelled as the effect of the glottal pulse on the zeros of the vocal tract. That is

$$\begin{aligned} s(n) &= \sum_{i=1}^N a(i)s(n-i) + e(n) \\ &= \sum_{i=1}^N a(i)s(n-i) + \sum_{j=0}^M b(j)g(n-j), \end{aligned}$$

where the $b(j)$ are the zero coefficients and $g(n)$ is an assumed glottal waveform model. The analysis procedure is summarized as follows: determine the $\{a(i)\}$ sequence using the covariance technique, calculate the error signal, approximate the error signal by calculating the $\{b(j)\}$ sequence so that $\sum_{j=0}^M b(j)g(n-j)$ approximates $e(n)$ in the

least squares sense. Synthesis uses the coefficient sequences $\{a(i)\}$ and $\{b(j)\}$ and the assumed glottal waveform to calculate

$$\hat{s}(n) = \sum_{i=1}^N a(i)\hat{s}(n-i) + \sum_{j=0}^M b(j)g(n-j)$$

Models for $g(n)$ were based on results by Rosenberg [1] and Holmes [2] - a polynomial-based model and a triangular model. Square and triangular doublets were also used.

Initially, the vocoder structure for this system was a fixed-frame size for analysis and synthesis. This led to amplitude modulation of the speech envelope and poor quality synthetic speech. Changing to a pitch synchronous system for analysis and synthesis improved that aspect of the synthesis, but it was evident that the synthetic speech was still degraded by spectrum-matching errors. Experimental data showed that even with the best glottal waveform model found, the triangular doublet, the synthesis quality was fairly insensitive to the number of zeros for a moderate number of zeros (≤ 20).

Because the goals of research included improvement of speech quality while maintaining reasonable bit rates, this project was terminated.

REFERENCES

- [1] Rosenberg, A.E., "Effect of Glottal Pulse Shape on the Quality of Natural Vowels", Journal of the Acoustical Society of America, Vol. 49 (1971), pp. 583-590.

- [2] Holmes, John N., "The Influence of Glottal Waveform on the Naturalness of Speech from a Parallel Formant Synthesizer", IEEE Transactions on Audio and Electroacoustics, Vol. AU-21, No. 3 (June 1973), pp. 298-305.

SECTION 8

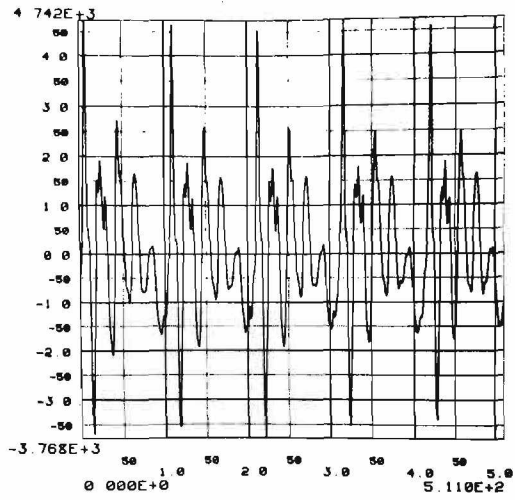
Speech Synthesis Using the Complex Cepstrum

The phase of vowels is computed, unwrapped and incorporated into speech analysis - synthesis technique. This task is to create synthetic speeches containing the phase information of the original human speech. The careful comparison between "phased" and "phaseless" synthetic speeches could reveal certain insight to the effect of phase on the vocoder.

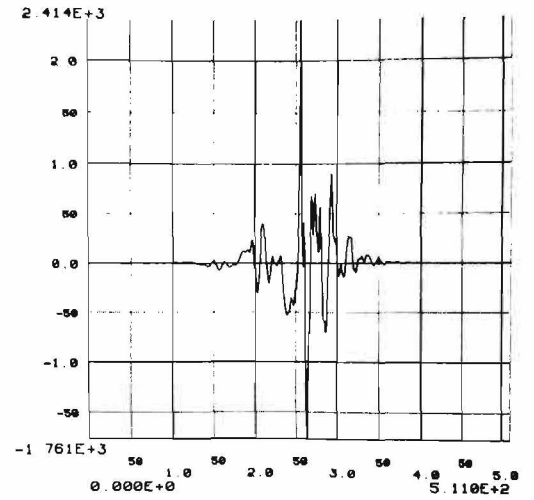
The log magnitude is obtained by taking the DFT of the analysis data buffer as in an ordinary real cepstrum homomorphic vocoder. However, to avoid the difficulty of unwrapping the phase of a periodic waveform, the unwrapped phase is obtained by unwrapping the phase of an isolated wavelet which serves as a prototype wavelet for the adjacent waveform within this data buffer. This unwrapped phase can be visualized as if it were obtained through the process of taking the DFT of this data buffer, which includes several pitch periods, unwrapping and then smoothing.

As far as analysis of speech is concerned, the log magnitude and the "smoothed" phase is then taken as the real and the imaginary part to calculate the complex cepstrum. The impulse response results easily after cepstrum filtering.

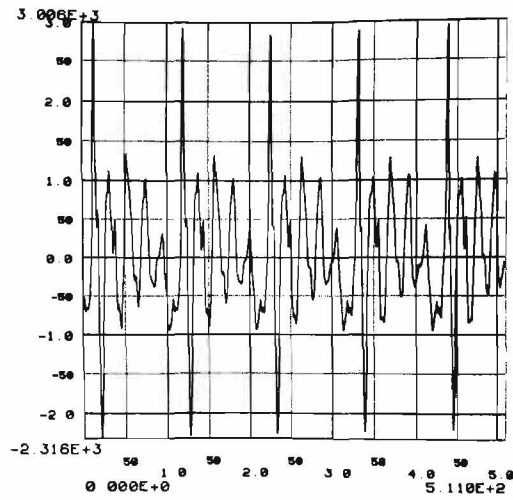
Figure 1 gives an example of an original vowel waveform, and the corresponding synthetic impulse response and synthetic speech obtained by this method.



(a)



(b)



(c)

FIGURE 1

Speech Synthesis Using the Complex Cepstrum

- (a) Original Vowel
- (b) Synthetic Impulse Response
- (c) Synthetic Speech

Image Understanding

Martin Newell

The work in image understanding is progressing on all four tasks outlined in the previous semi-annual report. However, no single area has progressed to the stage where significant results are available. However, no fundamental difficulties have been uncovered so that at the end of the next six months we expect to have all the necessary pieces of the system working together on the PDP-11/45 facility.

PUBLICATIONS AND PRESENTATIONS

- [1] Boll, S.F., Ferretti, E., Petersen, T.L. "Improving Synthetic Speech Quality Using Binaural Reverberation." Proceedings of the IEEE, Philadelphia, Pa., April 14, 1976.

- [2] Petersen, T.L. "Dynamic Sound Processing." Conference of the ACM, Anaheim, CA., February 12, 1976.

- [3] Atashroo, M. Ali, Petersen, T.L. "Dynamic Filtering of Degraded Speech Using Autocorrelation Prediction." Arden House Signal Processing Workshop, Harrimon, N.Y., February 23, 1976.