POL.TECHNIC INSTITUTE OF BROOKLYN

DEPARTMENT OF ELECTRICAL ENGINEERING

FINAL STATUS REPORT

NASA GRANT NGR 33-006-040

A STUDY OF DIGITAL TECHNIQUES FOR SIGNAL PROCESSING

July 1, 1968 - January 31, 1969

Principal Investigators

Mischa Schwartz, Professor

Robert R. Boorstyn, Assistant Professor

Summary

Work during the initial half of the grant period, as previously reported in the semi-annual status report of June 30, 1968, centered on three areas in the general field of digital processing of signals. These included:

1. Recursive techniques for digital signal processing,
2. Data smoothing and compression,
3. Computer simulation of low error rate communication systems.

In this report we focus on the following four areas:

1. Continuation of recursive techniques,
2. Continuation of optimal adaptive control for data compression systems,
3. Adaptive equalizers for digital transmission,
4. Signal zero-crossings as information carriers in communication systems.

As noted in the previous report, work on recursive techniques extended prior work on the detection of binary signals in additive noise by allowing the inclusion of colored noise with numerator dynamics into the previous digital formulation. In this report we continue the discussion of the detection of binary signals in colored noise with numerator dynamics and show that the resultant detector called for consists of a Kalman filter followed by recursive generation of the likelihood ratio. A paper based on this work was presented at the 1969 International Symposium on Information Theory [1], and is being submitted for publication.

One practical difficulty with recursive signal detection, as noted previously, is that derivatives of samples are required in its implementation. We report here on work done in determining the effect of approximating

necessary differential operations by sample differences. The work described here will be presented at the PIB-MRI International Symposium on Computer Processing in Communications, April 9, 1969, and will appear in the published proceedings of that symposium. A copy of the paper submitted is included as an appendix to this report.

In the area of data smoothing and compression we concentrate here on buffer design at the transmitter to handle the adaptive nature of the smoothed data flow. This is a continuation of work reported on in the semi-annual report.

The third area reported on here is that of digital processing for automatic adaptation of communication systems. Specifically we report here on work carried out relating to the digital implementation of adaptive equalizers for minimizing intersymbol interference in multi-level digital transmission. Two complementary studies are reported on. One has to do with methods of speeding up known adaptation procedures (or algorithms). The other has to do with the design of equalizers for nonlinear channel distortion. In addition to this work on adaptation, we have begun broadening its scope to include the area of adaptive antenna array processing. It is hoped to summarize preliminary results for this problem in the first report under the continuation of this grant.

The final area reported on here concerns that of using signal-crossings as information carriers in communication systems. This work was also presented at the 1969 International Symposium on Information Theory [2], and extends work first done for a Master's thesis at the Polytechnic [3].

1. Recursive Techniques for Digital Signal Processing

a. In a recent paper [4] Pickholtz and Boorstyn described a recursive approach to signal detection. The scheme was based on the following: The received signal was converted into a vector Markov process which was then sampled. The recursive structure of the digital processor followed readily. Of concern here are two aspects of this problem. First, in order to form a vector Markov process derivatives of the incoming signal are usually required. Investigations have been conducted into replacing these differentiation operations with approximating digital operations, such as differences. These studies, including simulation results, indicate that it is possible to replace deriviatives with differences without adversely affecting performance. Details appear below.

Secondly, the previous paper considered a special type of noise--that generated by a linear differential equation driven by white noise. A more general noise description would include numerator dynamics. Work has been initiated extending the recursive approach in this direction.

The essential part of the recursive receiver is to convert the incoming signal plus noise [r(t) = s(t) + y(t)] into a vector Markov process in such a manner that information is not destroyed. If this is done by a linear processor then the output of this device is $\underline{\rho}(t) = \underline{\sigma}(t) + \underline{\eta}(t)$ where the noise component $\underline{\eta}(t)$ is to be Markov. Furthermore we insist that r(t) be recoverable from $\underline{\rho}(t)$. Because of the linearity we need only consider the noise term. In the original work $\underline{\eta}(t)$ consisted of the derivatives of y(t) as well as y(t) itself and satisfied both of the above requirements.

We now consider y(t) to be generated by the following differential equation

$$\frac{dy^n}{dt^n} + \sum_{k=0}^{n-1} \alpha_K(t) \frac{dy^k}{dt^k} = \sum_{\ell=0}^{n-1} \beta_K(t) \frac{dw^\ell}{dt^\ell}$$

where w(t) is white Gaussian noise. It is possible to find a state vector $\underline{x}(t)$ for this system such that the first component $x_1(t) = y(t)$. This vector is the solution of

$$\dot{\underline{x}}(t) = A(t)\, \underline{x}(t) + \underline{b}(t)\, w(t)$$

$$y(t) = \underline{c}^T\, \underline{x}(t)$$

where $\underline{c} = \begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$

Although $\underline{x}(t)$ is Markov it cannot be obtained from y(t) alone--either w(t) or $\underline{x}(t_0)$ is needed in addition (neither is available). We consider next the best mean-square estimate of $\underline{x}(t)$ given the input y(s), s $\leq$ t. Thus

$$\hat{\underline{x}}(t) = E[\underline{x}(t) | y(s),\ s \leq t].$$

Since $\hat{x}_1(t) = y(t)$, y(t) can be recovered from $\hat{\underline{x}}(t)$--it is reversible. Furthermore because of the Gaussian assumption $\hat{\underline{x}}(t)$ is obtained by a linear operation on y(t). We have shown that $\hat{\underline{x}}(t)$ is Markov and thus letting $\underline{\eta}(t) = \underline{x}(t)$ satisfies our requirements.

To solve for $\hat{\underline{x}}(t)$ we convert the above problem into a more tractable one by differentiating y(t) to obtain

$$\dot{y}(t) - \underline{c}^T A(t)\underline{x}(t) + \underline{c}^T \underline{b}(t) W(t).$$

Now Kalman-Bucy techniques can be used to obtain

$$\dot{\hat{\underline{x}}}(t) = A(t)\hat{\underline{x}}(t) + \underline{z}(t)[\dot{y}(t) - \underline{c}^T A(t)\hat{\underline{x}}(t)]$$

where $\underline{z}(t) = [M(t)A(t)^T\underline{c} + \underline{b}(t)N_0(t)\underline{b}(t)^T\underline{c}]\ [\underline{c}^T\underline{b}(t)N_0(t)\underline{b}(t)^T\underline{c}]^{-1}$.

$M(t)$ is the solution of a matrix Riccatti equation.

The differential equation for $\hat{\underline{x}}(t)$ driven by $\dot{y}(t)$ above is just the linear processor used to convert $r(t)$ into $\underline{\rho}(t)$ when it is driven by $\dot{r}(t)$. When driven by $\dot{s}(t)$, $\underline{\sigma}(t)$ will be produced. If these vector waveforms are now sampled the following recursive form for the log of the likelihoood ratio results:

$$\theta_n - \theta_{n-1} = 2(\underline{\sigma}_n - B_n \underline{\sigma}_{n-1})^T K_n^{-1} (\underline{\rho}_n - B_n \underline{\rho}_{n-1})$$

where B_n and K_n are found from statistics of the noise.

The proposed form of the detector--a Kalman filter followed by recursive generation of the likelihood ratio--has several features: it is the only known solution for non-stationary noise; it is a causal and real-time operation; and, being recursive, these operations require a minimum of computational effort and memory capacity.

Further insight can be obtained by allowing the time between samples to shrink to zero which results in a continuous processor. The result of this limiting procedure is a differential equation for $\theta(t)$,

$$\dot{\theta}(t) = \alpha[\dot{s}(t) - \sigma_2(t)]\,[\dot{r}(t) - \rho_2(t)]$$

where $\sigma_2(t)$ and $\rho_2(t)$ are the second components of the vectors $\underline{\sigma}(t)$ and $\underline{\rho}(t)$, respectively. The significance of these expressions is that $\dot{y}(t) - \eta_2(t)$ can be shown to be an innovations process [5] and the subsequent operations are equivalent to performing a causal, and causally invertible, prewhitening operation [6].

The work derived above has been presented at the 1969 International

Symposium on Information Theory [1] and is being submitted for publication. This investigation is continuing.

b. The recursive approach to signal detection reduces considerably the computational effort when discrete samples are used. The problem is that the derivatives of the samples are required. These derivatives are not available directly from the sampled process and must be approximated.

Two simple types of derivative approximation have been treated and are summarized here. One replaces the derivatives, which are random variables, with their expected values. The other method approximates the derivatives by difference equations.

Equations for the signal-to-noise ratio using these derivative approximation techniques were derived. The results of these sub-optimum procedures may then be compared with the optimum procedure which assumes all derivatives are completely known.

A computer program for finding the optimum and sub-optimum signal to noise ratio for various noise processes and signals was developed. The results of this program for several signals of interest are included here in order to show the effectiveness of the derivative approximation methods.

The work described in this section will be presented at the PIB-MRI Symposium on Computer Processing in Communications, New York, April 9, 1969. The corresponding detailed paper will appear in the proceedings of that symposium and is included as an appendix to this report.

Two derivative approximation techniques were examined. The first estimates the derivatives by replacing them with their expected value. The second uses a difference equation of the sample values. In each case equations were derived which give the signal-to-noise ratio of these sub-optimum processes in recursive form. While the results have the same structure as the optimum process, they are much more complicated.

The results of this work indicate that near-optimal results can be obtained using either derivative approximation method. Of the two methods, the difference equation method yields more predictable results. In general, the difference equation technique yields results which appear to converge to the optimum as samples increase. The expectation method, however, often shows deterioration as samples increase. The optimum number of samples for this method varies considerably with the signal and noise characteristics.

As would be expected, the optimum as well as sub-optimum processes yield best results when the samples are taken at points which maximize the difference between the two binary transmitted signals. In these cases the sub-optimum processes give near optimal results, even for few samples. Those cases which depend on sample points where the two signals are equal show rather poor performance. It is also significant that good results are often obtained in cases where the derivatives approximated by the difference equation technique could not possibly be meaningful due to the large sample interval. These observations lead to the conclusion that errors in derivative approximation do not adversely affect the signal-to-noise ratio in those cases where meaningful information is available at the sample point in distinguishing the two signals. Where this information content is not available both optimum as well as sub-optimum processes give poor results.

The method of replacing derivatives with their expected values usually works as well as, if not better than, the difference equation method for two or three samples. The results often deteriorate, however, as samples increase. The cause of this is not obvious, but it may be reasoned that increased samples often add more error than information. Obviously this method does not yield approximations which converge to the true derivative value as the samples increase. However, since this method is extremely easy to implement, it should have application in those cases where good results are attainable.

The difference equation method gives good results in all of the test cases. As samples increase, results appear to converge to the optimum. The signal-to-noise ratio from this method does not always increase monotonically as the number of samples increase. This effect seems to be of a minor nature, and probably occurs only in areas where the sample interval is too long for the difference equation to give valid results.

The relative performance of the sub-optimum processes is often poorest for narrow band noise. This is probably due to the fact that where the noise is highly correlated, there is significant information in the derivatives of the sample state vector. In this case the errors in derivative approximation result in a signal-to-noise ratio well below optimum. It seems logical to conclude that accurate derivative estimation is more important for narrow band noise than for wide band n noise.

Although the results obtained are only for second order systems, some conclusions may be extended to higher order systems. It appears that the importance of accurately approximating the derivatives varies inversely with the order of the derivative. When considering higher order systems, it should be noted that the approximation of derivatives by difference equations becomes more inaccurate as the order of the derivative increases. It is quite likely that the errors introduced may offset the advantage of modeling the noise process by a high order process. Better results may be obtained by approximating the noise process with one of lower order.

It is concluded that the derivative approximation methods investigated in this report are capable of near optimal results. It has also been shown that the relative performance of these approximation methods are dependent upon the choice of signal. The fact that simple derivative approximations, in general, yield very good results is quite significant since any digital processing system using the proposed detection scheme would require derivative approximation.

2. Optimal Adaptive Control for Data Compression Systems

As noted in the prior semi-annual report the object here is to determine an optimal controller to minimize the mean-squared error between discrete input data, x_n, and reconstructed discrete output data, y_n. The controller-buffer system was modeled as a discrete Markov process and a method of solution adopted using an iterative dynamic programming algorithm based on the work of Howard [7]. Once the statistics of the input process, x_n, and the compressor algorithm have been specified, an optimum controller can be determined using this technique.

During the period reported on, a new method was developed to determine the transition probabilities, $p_{ij}^{(k)}$, and the expected immediate losses, $q_i^{(k)}$, from buffer states i to j during major cycle time T, when using aperture value k. These quantities are the inputs to Howard's policy iteration routine which yields the value of the minimum mean squared error and the optimum policy which causes this error.

Three sample problems were solved by hand using the equations derived from this new method for uniformly distributed independent data. A computer program was also written to simulate the system being analyzed. This program was run using a uniformly distributed random number generator (having independent samples) as the simulated input and operating with the optimum policy determined by the hand calculations. In all three cases the mean squared error obtained from the simulation was within a 1% sampling error from the theoretical minimum mean squared error. As a further check the program was run for non-optimum policies which yielded larger mean squared errors.

The remaining work for the uniformly distributed case will be to solve the equations for a sufficient number of problems to yield tabular or graphical data for the design curves. Due to the complexity of the equations, the iterative solution has been programmed in FORTRAN on a digital computer and is persently being debugged. The results of the three hand calculated sample problems are being used as a standard to help debug this program. After the program has been debugged and run for a sufficient number of problems, these results will be spot checked with the simulation program.

It is then proposed to derive the equations for uniformly distributed Markov data. A Markov data model is presently being investigated. This model should have the properties of containing a parameter that can vary the correlation coefficient of adjacent samples between zero and unity, and be in a form that can be readily simulated on a digital computer.

3. Adaptive Equalizers for Digital Transmission

A great deal of effort in the past few years has gone into the development of adaptive equalizers to minimize the intersymbol interference introduced in the transmission of multilevel digital data through (unknown) dispersive channels. This is of significant technological importance both in the transmission of high-speed digital data over telephone lines and through time-varying dispersive media such as the ionosphere, troposphere, etc. The mathematics of the problem are such as to also permit the results to be extended to such other areas as adaptive antenna systems.

The adaptive equalizers initially developed and reported on in the literature were of the tapped-delay line, transversal filter configuration [8]. More recently emphasis has been placed on all-digital configurations using integrated circuitry. More generally one may represent the processor (the equalizer) as a non-recursive digital filter and search for means for optimally adjusting the filter coefficients. The equalizer problem then becomes equivalent to that of developing a special-purpose computer (whether hard-wired or in software form) that processes the incoming signal samples to minimize distortion due to intersymbol interference. As signal waveshapes or channel characteristics change the processor automatically adapts to the new conditions.

In most previous work a gradient search technique with fixed step size was assumed in formulating the algorithms for adjusting the filter coefficients. This method can be rather slow in converging to the appropriate optimum point, putting a limit on the rate at which the system adapts to changing conditions. We have explored various methods of speeding up the search procedure during this past reporting period. One particularly promising approach investigated is that of using a gradient search but adjusting the step size of each iteration to ensure the smallest possible error after a specified number (say M) of iterations. (By error we actually mean error vector with dimensionality N corresponding to the number of filter coefficients to be adjusted. It is then the norm of the error vector that is to be minimized.)

Interestingly the problem when posed in this way provides as its solution step sizes given by the reciprocal of the zeroes of polynomials related to the classical Chebyshev polynomials. These polynomials involve the smallest and largest eigenvalues of the matrix involving lagged products of the incoming signal samples. They are hence in turn related to the (unknown) channel characteristics.

An analysis of the convergence rate of this algorithm indicates it is substantially faster than that for the fixed step size gradient search. The ratio of the norms of the Chebyshev error vector to the fixed step size error vector may be shown to decrease exponentially with the iteration number M. Computer simulation for various types of channels has verified these results.

Additional analysis of this Chebyshev-related algorithm has focused on the effects of additive noise on the convergence. A rather loose upper bound on the variance of the error vector, due to noise, and obtained by a randomly-generated step size argument, indicates the variance is roughly the same as that for fixed step size. There is thus no deterioration due to noise using this type of gradient search. Extensive computer simulation has been begun to verify the variance bound. Further work is continuing to tighten the bound, to analyze several specific types of channels, and to compare in detail the Chebyshev approach with both fixed step size and optimum gradient search procedure.

Nonlinear Distortion and Adaptive Equalizers

Most previous work on adaptive equalization, including the work described above has focused on minimizing intersymbol interference due to channel amplitude and phase distortion. We have now initiated a study of the effects of nonlinear channel distortion on the equalizer problem.

Specifically two approaches are to be initially investigated: 1. Assuming a predetermined structure for the equalizer investigate both its ultimate capabilities in minimizing intersymbol interference and various algorithms for automatically adapting to the desired characteristics; 2. Using some optimum criterion--minimum probability of error or least mean-squared error--attempt to obtain an optimum form for the structure of the equalizer.

As is true with most nonlinear problems general results will presumably be difficult to obtain. However, by restricting ourselves to particularly simple yet meaningful cases we hope to gain insight into the problem. Specifically, the types of nonlinearity to be investigated, for which some preliminary analysis has already been completed, include quadratic and cubic distortion terms. The size of the equalizer (number of filter coefficients) will also be restricted to a manageable number, and intersymbol interference will be assumed limited to adjacent bounds only.

Some thought indicates that any adaptive equalizer working with signals having come through a nonlinear channel must contain nonlinear terms as well. As a particularly simple predetermined structure for such an adaptive equalizer we have chosen to investigate one suggested by a Volterra series expansion. Such expansion is commonly used to represent a nonlinear input-output relationship between an input signal x(t) and the output y(t). The resultant functional power series may be considered to be the generalization of a transfer function to nonlinear systems. The first term in the series represents a general linear filter, and so suggests a linear non-recursive digital filter as its implementation. The second term turns out to be given by the linear sum of second-order products of the delayed input samples, etc.

Some limited computer simulation of such an ad hoc equalizer scheme indicates that it does in fact provide improved equalization with nonlinear distortion present over a purely linear equalizer of the same size. Its ultimate capabilities as well as comparison to more "optimum" structures remain to be investigated, however.

4. Signal Zero-Crossings as Information Carriers in Communication Systems [2],[3]

The distribution of zero-crossings has received little attention as a means of signal transmission. However, the intelligibility of clipped speech is well known from Licklider's work [9]. This phenomenon suggests the idea that zero-crossings carry much information about the signal. Another fact is the striking equality between the number of degrees of freedom per unit of time of a bandlimited signal and the maximum density of zero-crossings of such a signal.

Titchmarsh [10] has shown that a signal band limited to (-W,+W) is completely defined by its zero-crossings if the density of these points is 2W. Our purpose is to use a computer to recover the signal from its zero-crossings. Thus we require that they be uniformly distributed in the following ~~sense: it is possible to find a division of~~ the time axis in intervals 1/2W wide such that there is one and only one zero-crossing in each. This feature is necessary for a computer implementation since only a finite number of data can be processed at the same time; this leads to the recovery of the signal in an interval from the zeros it contains (with a resulting error); uniform distribution of the zeros insures uniform goodness of fit of the approximation in successive intervals.

A signal with these two properties will be called an optimum signal. In the following we investigate an algorithm to recover an optimum signal from its zeros. It is usual for a communication engineer to expand a signal truncated in time in a Fourier series. We thus come to an algorithm which follows this philosophy. Let

$$\hat{x}(t) = \sum_{k=-N}^{+N} c_{N-k} e^{jk\omega_0 t} \qquad (\omega_0 = \pi/T; \; |t| \leq T)$$

be the approximation with the same zero-crossings as the signal x(t) in the interval (-T, +T). Since the signal is optimum the interval contains 4WT zeros, therefore if we know the zero-crossing locations and the energy of the signal in the interval we have enough information to compute 4WT + 1 coefficients in the above equation; thus we let N = 2WT. From the identity of the zero-crossings of x(t) and $\hat{x}(t)$ we can write

$$\prod_{i=1}^{2N} (q - q_i) = \sum_{k=-N}^{+N} \frac{c_{N-k}}{c_0} q^{N+k}$$

where $q = \exp(j\omega_0 t)$ and q_i is the value taken by q at the i-th zero-crossings. From this we derive the 2N equations

$$\sum_{i=1}^{2N} q_i = -\frac{c_1}{c_0}$$

$$\sum_{i=1,\,j>i}^{2N} q_i q_j = +\frac{c_2}{c_0}$$

$$\sum_{i=1,\,k>j>i}^{2N} q_i q_j q_k = -\frac{c_3}{c_0}$$

etc.

and from the identity of the energies:

$$\sum_{k=-N}^{+N} |c_{N-k}|^2 = \frac{1}{2T} \int_{-T}^{+T} x^2(t)dt = \frac{E_T}{2T}$$

It is easily seen that this algorithm does not in general lead to the classical equality of Fourier series theory

$$c_{N-k} = c^*_{N+k}$$

and therefore $\hat{x}(t)$ is usually a complex function. Taking the real part of $\hat{x}(t)$ as an approximation, the signal-to-mean-square-error ratio for various optimum signals and intervals of time containing 6 zero-crossings is of the order of 20 db.

For the class of optimum signals the output signal-to-noise-ratio $(\frac{S}{N})_o$ of a communication system consisting of a clipper, the channel, a second clipper and the computer can be evaluated for high channel (input) signal-to-noise ratio $(\frac{S}{N})_i$ and large bandwidth. We get

$$(\frac{S}{N})_o = \frac{2}{\pi} (\frac{S}{N})_i e^{(\frac{S}{N})_i}$$

i.e, essentially an exponential behavior.

For the transmission of a general bandlimited signal x(t) we first transform the signal into an optimum signal and transmit on two separate channels the zero crossings of the optimum signal and a simple coded sequence enabling the original signal to be recovered. In this case the output signal-to-noise ratio is still given by an exponential function of the channel signal-to-noise ratio.

Bibliography

1. R.R. Boorstyn and R.L. Pickholtz, Recursive Detection with Numerator Dynamics, International Symposium on Information Theory, Ellenville, N.Y., January 1969.

2. A.L. Fawe and R.R. Boorstyn, Investigation of Signal Zero-Crossings as Information Carriers in Communications Systems, International Symposium on Information Theory, Ellenville, N. Y., January 1969.

3. A.L. Fawe, The Use of Zero-Crossings in Signal Processing, Masters Report, Polytechnic Institute of Brooklyn, N. Y., June 1968.

4. R.L. Pickholtz and R.R. Boorstyn, A Recursive Approach to Signal Detection, IEEE Trans. Information Theory, Vol. IT-14, pp.445-450, May 1968.

5. T. Kailath, An Innovations Approach to Least Squares Estimation Part 1: Linear Filtering in Additive White Noise; Part 2 (with Paul Frost): Linear Smoothing in Additive White Noise, Trans. Automatic Control, Vol. AC-13, pp. 646-660, December 1968.

6. R.A. Geesey and T. Kailath, Application of the Canonical Representation of Second Order Processes to Estimation and Detection in Colored Noise, MRI Symposium on Computer Processing in Communications, Polytechnic Institute of Brooklyn, N. Y., April 1969.

7. R.A. Howard, "Dynamic Programming and Markov Processes", MIT Press, Cambridge, Mass., 1960.

8. R.W. Lucky, J. Salz, and E.J. Weldon, "Principles of Data Communication", McGraw-Hill, 1968.

9. J.C.R. Licklider and I. Pollack, Effects of Differentiation, Integration, and Infinite Peak Clipping upon the Intelligibility of Speech, J A.S.A., Vol. 20, pp. 42-51, January 1948.

10. E.C. Titchmarsh, The Zeros of Certain Integral Functions, Proc.London Math.Soc., Vol. 25, pp. 283-302, 1926.