express H using (1)-(4), and some trivial integration, as

$$H = \tilde{H} + N_0 \Psi$$

where Ψ has entries

$$\Psi_{i,j} = \frac{1}{\pi}\left[\frac{B^{1-2(i+j)}}{1-2(i+j)}\right], \qquad i, j = 0, \cdots, N/2. \tag{6}$$

It remains only to find N_0 so that

$$\tilde{H} = H - N_0 \Psi \tag{7}$$

is singular positive semidefinite (SPSD). If Ψ was an identity matrix (as in the Pisarenko case) this would become an eigenvalue problem. As it stands, it is what is known as a matrix *pencil* problem. There are dedicated algorithms for such problems [6]. We prefer to convert it to an eigenvalue problem as follows. Since Ψ is strictly positive definite, it can be factored, using the Cholesky decomposition [6], as $\Psi = \Lambda\Lambda^t$, where Λ is square and nonsingular. Multiplying both sides of (7) by Λ^{-1} and Λ^{-t} yields

$$\Lambda^{-t}\tilde{H}\Lambda^{-1} = \Lambda^{-t}H\Lambda^{-1} - N_0 I. \tag{8}$$

It follows from Sylvester's theorem [6] that the right-hand side of (8) is SPSD iff $\tilde{H}$ is SPSD. Subtracting the smallest eigenvalue from the diagonal of *any* matrix results in an SPSD matrix. Thus, the desired N_0 is simply the minimum eigenvalue of $\Lambda^{-t}H\Lambda^{-1}$.

We have derived the following analog harmonic retrieval scheme.

Harmonic Retrieval Algorithm: The following algorithm finds the frequencies, ω_i, and the power levels, a_i/π of $N/2$ analog sinusoids in broad-band noise.

1) Collect the power estimates p_i using the circuitry in Fig. 1.
2) Sample the p_i and form the Hankel matrix H as in (3).
3) Set N_0 = the minimum eigenvalue of $\Lambda^{-t}H\Lambda^{-1}$.
4) Set α = the zero eigenvector of $H - N_0\Psi$.
5) Set ω_i = the real roots of the polynomial $\omega^N(\Sigma_{i=0}^{N/2} \alpha_i \omega^{-2i})$.
6) Solve the system of equations in (5) for the a_i, with p_i replaced by $\tilde{H}_{j,i-j}$.

In a theorem given by Iohvidov, it is shown that any SPSD Hankel matrix possesses a unique extension [7]. It follows from this that there is correspondingly a unique power spectrum $S(\omega)$ which matches the given p_i. Thus, the parameters in (1) are uniquely defined. This is an important result since it guarantees that our algorithm will correctly reconstruct the true harmonics in the data (at least in the limiting case of exact power estimates).

If $\tilde{H}$ has multiple zero eigenvectors, then the polynomial in step 5) is not uniquely defined. Despite this nonuniqueness, the Iohvidov theorem ensures us that whatever eigenvector we choose will lead us to the correct (unique) result. It is shown in [7] that if the zero eigenvalue has multiplicity k, then there are exactly $N/2 + 1 - k$ nonzero terms in the summations in (5),[4] and consequently, there are exactly $N/2 + 1 - k$ sinusoids in $u(t)$. Thus, the rank of $\tilde{H}$ equals the number of sinusoids in the data.

IV. Further Questions and Issues

Below is a list of questions and issues which our work raises.

An adaptive updating of the spectral model parameters in (1) would be desirable. How does one implement such a scheme in a hybrid system?

What are the effects of "real world" limitations such as analog component imperfections, statistical fluctuations, and signal modeling inaccuracies?

One could replace the cascaded integrators with other more general filters, provided the resultant Ψ matrix remains nonsingular. (Cascaded all-pass filters, as employed in [5], will definitely not work here!) Does this opportunity to generalize have any merit?

The integrators can likewise be replaced by *digital* filters, thereby offering an alternative to the ordinary Pisarenko scheme for digital harmonic retrieval. Bruzzone and Kaveh [8] have explored the information preserving character of correlation estimates, and subsequently introduced the notion of the *relative information index* (RII). Do our power estimates have a higher or lower RII than correlation estimates? Which digital filters yield a high RII?

[4]In [7] a sum involving positive powers of ω is considered. This converts trivially to our problem by a simple change of variables.

References

[1] V. Pisarenko, "The retrieval of harmonics from a covariance function," *Geophys. J. Roy. Astronom. Soc.*, pp. 347-366, 1973. Also in *Modern Spectrum Analysis*, Reprint Series. New York: IEEE Press, 1986.

[2] D. Furhmann and B. Liu, "Rotational search methods for adaptive Pisarenko harmonic retrieval," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-34, pp. 1550-1565, Dec. 1986.

[3] V. Reddy, B. Egardt, and T. Kailath, "A least-squares-type algorithm for adaptive implementation of Pisarenko's harmonic retrieval method," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-30, pp. 399-405, June 1982.

[4] Y. Hu and S. Y. Kung, "A Toeplitz eigensystem solver," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-33, pp. 1264-1271, Oct. 1985.

[5] A. Oppenheim and D. Johnson, "Discrete representation of signals," *Proc. IEEE*, vol. 60, pp. 681-691, June 1972.

[6] G. Golub and C. Van Loan, *Matrix Computations*. Baltimore, MD: John Hopkins University Press, 1983.

[7] I. Iohvidov, *Hankel and Toeplitz Forms*. Boston, MA: Birkhauser, 1982.

[8] S. Bruzzone and M. Kaveh, "Information tradeoffs in using the sample autocorrelation function in ARMA parameter estimation," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-32, pp. 701-715, Aug. 1984.

On the Asymptotic Distribution of Exponentially Weighted Prediction Error Estimators

PETRE STOICA AND ARYE NEHORAI

***Abstract*—This correspondence establishes the distribution of the exponentially weighted prediction error estimators for a class of general (possibly nonlinear) discrete-time systems. An explicit formula for the covariance matrix of the estimation errors is provided. The distributional results presented hold for a large number N of data points and a weighting (or forgetting) factor λ close to one. The covariance matrices corresponding to $\{N_1 \gg 0; \lambda_1 = 1\}$ and to $\{N_2 \ggg 0; \lambda_2 = 1 - \delta$, where δ is a small positive number$\}$ are shown to be related in a simple way. Specifically, it is shown that they are equal iff $N_1 = 2/(1 - \lambda_2)$. This extends a result recently obtained by Porat in the special case of times-series autoregressive models.**

I. Main Results

Consider the following general prediction error (PE) model:

$$y(t) = g(u^{t-1}, y^{t-1}, \theta) + \epsilon(t, \theta), \qquad t = 1, 2, \cdots, \tag{1.1}$$

Manuscript received January 24, 1987; revised July 20, 1987. The work of A. Nehorai was supported in part by the National Science Foundation under Grant DCI-8604351.

P. Stoica is with the Department of Automatic Control, Polytechnic Institute of Bucharest, Splaiul Independentei 313, R-77 206 Bucharest, Romania.

A. Nehorai is with the Department of Electrical Engineering, Yale University, New Haven, CT 06520.

IEEE Log Number 8717433.

0096-3518/88/0100-0136$01.00 © 1988 IEEE

where

$y(t)$ = the output at time t
$u(t)$ = the input at time t
$u^{t-1} = \{u(t-1), u(t-2), \cdots\}$
$y^{t-1} = \{y(t-1), y(t-2), \cdots\}$
θ = the vector of unknown parameters
$\epsilon(t, \theta)$ = the prediction errors corresponding to θ, at time t.

The model (1.1) and the system which generates the data $\{u(t), y(t)\}$ are assumed to satisfy the following conditions.

$C1$: There exists a (true) parameter vector θ^* such that

$$e(t) = \epsilon(t, \theta^*), \qquad t = 1, 2, \cdots, \tag{1.2}$$

form a sequence of independent and identically distributed random variables with zero mean and variance denoted by σ [in other words, the system belongs to the model class (1.1)].

$C2$: The input and output signals $u(t)$ and $y(t)$ are stationary processes whose covariance functions decay exponentially to zero. Furthermore, there exists a compact subset D of the parameter space, which contains θ^* and is such that $\epsilon(t, \theta)$ and $\partial\epsilon(t, \theta)/\partial\theta$ are stationary processes with exponentially decaying covariances for all $\theta \in D$ (this assumption essentially means that the model equation (1.1) and its derivative with respect to θ are exponentially stable for $\theta \in D$).

$C3$: The data are scalar-valued Gaussian distributed variables (this assumption, which is introduced to simplify the calculations, could be relaxed with some effort).

The exponentially weighted PE estimate of θ^* is defined by (see, e.g., [2]-[5])

$$\hat{\theta}_\lambda = \arg\min_{\theta \in D} \sum_{t=1}^{N} \lambda^{N-t} \epsilon^2(t, \theta) \tag{1.3}$$

where λ lies in the interval $(0, 1]$ and N denotes the number of data samples. The reasons for introducing the weighting factor λ^{N-t} in (1.3) are well known. For example, in general we do not dispose of data prior to the moment $t = 0$. Thus, the first few prediction errors (or residuals) $\epsilon(t, \theta)$ $t = 1, 2, \cdots$, cannot be computed exactly and we may wish to weight them out in the PE criterion. A similar reason applies to the on-line (i.e., recursive in N) computation of $\hat{\theta}$ where again unknown initial conditions may affect the initial computations. Since motivating the form of the weights in (1.3) is not our main concern here, we refer the reader to [3] and [4] for more details.

The distributional properties of $\hat{\theta}_\lambda$ for large N and $\lambda = 1$ are well known (see e.g., [3]-[5]). The normalized estimation errors $\sqrt{N}(\hat{\theta}_1 - \theta^*)$ corresponding to $\lambda = 1$ are asymptotically (for $N \to \infty$) Gaussian distributed

$$\sqrt{N}(\hat{\theta}_1 - \theta^*) \xrightarrow[N\to\infty]{\text{law}} \mathfrak{N}(0, P_1) \tag{1.4a}$$

with the covariance matrix P_1 given by

$$P_1 = \sigma\left[E\psi(t)\,\psi^T(t)\right]^{-1} \tag{1.4b}$$

where

$$\psi(t) = \left.\frac{\partial\epsilon(t, \theta)}{\partial\theta}\right|_{\theta=\theta^*}. \tag{1.4c}$$

The distributional properties of $\hat{\theta}_\lambda$ in the case $\lambda < 1$ have only recently been studied. In [2] and [3] it has been shown that the estimate $\hat{\theta}_\lambda$ may be inconsistent (i.e., $\lim_{N\to\infty} \hat{\theta}_\lambda \neq \theta^*$) for $\lambda < 1$. In this correspondence we assume that λ is (very) close to one, which seems to be the case of interest for most applications. We show that the asymptotic bias $\lim_{N\to\infty} \hat{\theta}_\lambda - \theta^*$ goes to zero as $\lambda \to 1$, as is expected. Thus, $\hat{\theta}_\lambda$ is asymptotically (for $N \to \infty$ and $\lambda \to 1$) an unbiased estimate of θ^*. The asymptotic covariance matrix of the estimation errors $(\hat{\theta}_\lambda - \theta^*)$ has recently been derived by Porat [1] in the case of time-series autoregressive models, i.e., when the predictor function g in (1.1) has the following special form:

$$g(u^{t-1}, y^{t-1}, \theta) = \theta_1 y(t-1) + \cdots + \theta_n y(t-n), \qquad n < \infty.$$

In this correspondence we extend the result of [1] to models of the general form (1.1). We show that under conditions $C1$-$C3$, the normalized estimation errors $(\hat{\theta}_\lambda - \theta^*)/\sqrt{1-\lambda}$ are asymptotically (for $N \to \infty$ and $\lambda \to 1$) Gaussian distributed

$$(\hat{\theta}_\lambda - \theta^*)/\sqrt{1-\lambda} \xrightarrow[\substack{N\to\infty\\ \lambda\to 1}]{\text{law}} \mathfrak{N}(0, P_\lambda) \tag{1.5a}$$

where

$$P_\lambda = (\sigma/2)\left[E\psi(t)\,\psi^T(t)\right]^{-1}. \tag{1.5b}$$

Our analysis, despite its generality, is simpler than that of [1]. Furthermore, our approach makes it possible to bypass some (minor) flaws in the analysis of [1].

Finally, let us note that the basic difference between our analysis and the standard analysis of the nonweighted PE method ([3]-[5]) lies in the order in which λ and N tend to one and infinity, respectively. In [3]-[5], λ is first set to one and then N is let to tend to infinity. Here we first let N tend to infinity and next let λ tend to one.

II. Proof of (1.5)

Lemma A in the Appendix implies that for λ close to one we have

$$\lim_{N\to\infty} (1-\lambda) \sum_{t=1}^{N} \lambda^{N-t} \epsilon^2(t, \theta) = E\epsilon^2(t, \theta) + O(\sqrt{1-\lambda}), \qquad \text{for all } \theta \in D. \tag{2.1}$$

Since

$$\begin{aligned} E\epsilon^2(t, \theta) &= E\left[g(u^{t-1}, y^{t-1}, \theta^*) - g(u^{t-1}, y^{t-1}, \theta) + e(t)\right]^2 \\ &\geq Ee^2(t) = E\epsilon^2(t, \theta^*), \end{aligned} \tag{2.2}$$

it follows that $\theta = \theta^*$ minimizes the dominant term in (2.1). Thus, for large N and λ close to one, $\hat{\theta}_\lambda$ is close to θ^*. A simple Taylor series expansion of the gradient equation

$$(1-\lambda) \sum_{t=1}^{N} \lambda^{N-t} \left.\frac{\partial\epsilon(t, \theta)}{\partial\theta}\right|_{\theta=\hat{\theta}_\lambda} \cdot \epsilon(t, \hat{\theta}_\lambda) = 0 \tag{2.3}$$

around θ^* gives

$$\begin{aligned} 0 \simeq\ & (1-\lambda) \sum_{t=1}^{N} \lambda^{N-t} \psi(t)\, e(t) \\ & + \left\{ (1-\lambda) \sum_{t=1}^{N} \lambda^{N-t} \left[\psi(t)\,\psi^T(t) + \left.\frac{\partial^2\epsilon(t, \theta)}{\partial\theta^2}\right|_{\theta=\theta^*} \cdot e(t)\right] \right\} (\hat{\theta}_\lambda - \theta^*). \end{aligned} \tag{2.4}$$

Note from Lemma A that the terms neglected in (2.4) tend to zero faster than $(\hat{\theta}_\lambda - \theta^*)$ as $N \to \infty$ and $\lambda \to 1$. It also follows from Lemma A that under the conditions of this proof, the matrix multiplying $(\hat{\theta}_\lambda - \theta^*)$ in (2.4) can be replaced by its expected value without affecting the dominant term in (2.4). More precisely, from (A.2) we have

$$(1-\lambda) \sum_{t=1}^{N} \lambda^{N-t} \left[\psi(t)\,\psi^T(t) + \left.\frac{\partial^2\epsilon(t, \theta)}{\partial\theta^2}\right|_{\theta=\theta^*} e(t)\right] \xrightarrow[N\to\infty]{} E\psi(t)\,\psi^T(t) + O(\sqrt{1-\lambda}). \tag{2.5}$$

It follows from (2.4) and (2.5) that, for large N and λ close to one,

$$(\hat{\theta}_\lambda - \theta^*)/\sqrt{1-\lambda} \simeq -\left[E\psi(t)\,\psi^T(t)\right]^{-1} \cdot \left\{\sqrt{1-\lambda}\sum_{t=1}^{N}\lambda^{N-t}\psi(t)\,e(t)\right\}. \tag{2.6}$$

Standard results on the convergence in law [3], [6] and some central limit theorem [7] then imply that

$$(\hat{\theta}_\lambda - \theta^*)/\sqrt{1-\lambda} \xrightarrow[\substack{N\to\infty\\ \lambda\to 1}]{\text{law}} \mathfrak{N}(0, P_\lambda) \tag{2.7a}$$

where

$$P_\lambda = \left[E\psi(t)\,\psi^T(t)\right]^{-1}\tilde{P}\left[E\psi(t)\,\psi^T(t)\right]^{-1} \tag{2.7b}$$

and where

$$\tilde{P} = \lim_{\lambda\to 1}\lim_{N\to\infty}(1-\lambda)\,E\sum_{t=1}^{N}\sum_{s=1}^{N} \cdot \lambda^{2N-t-s}\psi(t)\,e(t)\,\psi^T(s)\,e(s). \tag{2.7c}$$

It remains to evaluate the covariance matrix $\tilde{P}$. Due to the Gaussian assumption and since $E\psi(t)\,e(s) = 0$ for $t \le s$, we can write

$$\begin{aligned}\tilde{P} &= \lim_{\lambda\to 1}\lim_{N\to\infty}(1-\lambda)\sum_{t=1}^{N}\sum_{s=1}^{N}\lambda^{2N-t-s} \\ &\quad\cdot\left\{\left[E\psi(t)\,e(t)\right]\left[E\psi^T(s)\,e(s)\right]\right. \\ &\quad+\left[E\psi(t)\,\psi^T(s)\right]\left[Ee(t)\,e(s)\right] \\ &\quad\left.+\left[E\psi(t)\,e(s)\right]\left[E\psi^T(s)\,e(t)\right]\right\} \\ &= \lim_{\lambda\to 1}\lim_{N\to\infty}(1-\lambda)\sum_{t=1}^{N}\lambda^{2(N-t)} \\ &\quad\cdot\sigma E\psi(t)\,\psi^T(t) = \frac{\sigma}{2}\left[E\psi(t)\,\psi^T(t)\right].\end{aligned} \tag{2.8}$$

Inserting the expression above of $\tilde{P}$ into (2.7) we get (1.5), and the proof is finished.

III. Conclusions

It follows from the distributional properties (1.4) and (1.5) that the covariance matrix of the unnormalized estimation errors $(\hat{\theta}_\lambda - \theta^*)$ is approximately given by

$$\frac{\sigma}{N_1}\left[E\psi(t)\,\psi^T(t)\right]^{-1}$$

$$\text{for } N = N_1 \gg 0 \quad \text{and } \lambda = \lambda_1 = 1$$

and by

$$\frac{\sigma(1-\lambda_2)}{2}\left[E\psi(t)\,\psi^T(t)\right]^{-1}$$

$$\text{for } N = N_2 \to \infty \text{ (or } N_2 \ggg 0)$$

$$\text{and } \lambda_2 = 1-\delta, \quad \delta = 0^+.$$

Note that the two covariance matrices above are equal iff $N_1 = 2/(1-\lambda_2)$. Thus, the exponentially weighted PE method with λ close to one may need processing of a very large data sample to achieve the same accuracy as the nonweighted PE method which processes $2/(1-\lambda)$ data samples only. Roughly speaking, this means that the exponentially weighted PE method applied to a long data sample "effectively" uses the "information" contained only in the most recent $2/(1-\lambda)$ data points. Stated in this way, our conclusion resembles a well-known rule of thumb which asserts that the prediction errors $\epsilon(t, \theta)$ for $t < (N - T)$, where $T = 1/(1-\lambda)$, have a negligible influence on the loss function in (1.1) [T is sometimes called "the memory time constant" of the criterion (1.1)], see, e.g., [4, p. 274]. Our analysis provides a quantitative interpretation of the above rule of thumb, which was previously used on more or less heuristical grounds. This extends the interpretation provided by Porat [1] for time-series autoregressive models, to more general models and experimental conditions.

Appendix
On Convergence of Exponentially Weighted Sample Covariances

In this appendix we state and prove a basic result for the analysis in the paper.

Lemma A: Let $\{v(t)\}$ and $\{w(t)\}$ be two Gaussian stationary sequences and define

$$S = (1-\lambda)\sum_{t=1}^{N}\lambda^{N-t}v(t)\,w(t) \tag{A.1}$$

where $\lambda \in (0, 1)$. The covariance functions $\gamma_k = Ev(t)\,v(t-k)$ and $\rho_k = Ew(t)\,w(t-k)$, as well as the cross-variance function $r_k = Ev(t)\,w(t-k)$, are assumed to decay exponentially to zero as $k \to \infty$. Then, for λ close to one, it holds that

$$S \xrightarrow[N\to\infty]{} r_0 + O(\sqrt{1-\lambda}) \tag{A.2}$$

where $O(\delta)$ denotes a zero-mean random variable whose standard deviation goes to zero at least as δ when δ tends to zero.

Proof: A simple calculation shows that

$$ES = (1-\lambda)\,r_0\sum_{t=1}^{N}\lambda^{N-t} \xrightarrow[N\to\infty]{} r_0. \tag{A.3}$$

Thus, to prove (A.2), it remains to show that $E(s - r_0)^2$ goes to zero as $(1-\lambda)$ when $N \to \infty$ and $\lambda \to 1$.

Using the Gaussian assumption, we can write

$$\begin{aligned}ES^2 &= (1-\lambda)^2\sum_{t=1}^{N}\sum_{s=1}^{N}\lambda^{2N-t-s}\,Ev(t)\,w(t)\,v(s)\,w(s) \\ &= (1-\lambda)^2\sum_{t=1}^{N}\sum_{s=1}^{N}\lambda^{2N-t-s} \\ &\quad\cdot\left[r_0^2 + \gamma_{t-s}\rho_{t-s} + r_{t-s}r_{s-t}\right] \\ &\triangleq T_1 + T_2 + T_3.\end{aligned} \tag{A.4}$$

It is easy to see that

$$\lim_{N\to\infty} T_1 = r_0^2. \tag{A.5}$$

Next consider T_2. The assumption that r_k and ρ_k decay exponentially to zero when k approaches infinity implies that

$$|\gamma_{t-s}\rho_{t-s}| \le c\mu^{|t-s|}$$

for some constant c and $\mu < 1$. It follows that

$$\begin{aligned}|T_2| &\le c(1-\lambda)^2\sum_{t=1}^{N}\sum_{s=1}^{N}\lambda^{2N-t-s}\mu^{|t-s|} \\ &= c(1-\lambda)^2\sum_{k=0}^{N-1}\sum_{p=0}^{N-1}\lambda^{k+p}\mu^{|k-p|}\end{aligned}$$

and therefore that

$$\begin{aligned}\lim_{N\to\infty}|T_2| &\le c(1-\lambda)^2\sum_{k=0}^{\infty}\left\{\sum_{p=0}^{k-1}(\lambda\mu)^k(\lambda/\mu)^p\right. \\ &\quad\left.+\sum_{p=k}^{\infty}(\lambda/\mu)^k(\lambda\mu)^p\right\} \\ &= c(1-\lambda)^2\sum_{k=0}^{\infty}\left|(\lambda\mu)^k\frac{(\lambda/\mu)^k-1}{(\lambda/\mu)-1}\right. \\ &\quad\left.+(\lambda/\mu)^k(\lambda\mu)^k\frac{1}{1-\lambda\mu}\right| \\ &\le \text{const}\cdot(1-\lambda)^2\sum_{k=0}^{\infty}\left[2\lambda^{2k}-(\lambda\mu)^k\right]\end{aligned}$$

$$= \text{const} \cdot (1-\lambda)^2 \left[\frac{2}{1-\lambda^2} - \frac{1}{1-\lambda\mu} \right]$$

$$\leq \text{const} \cdot \frac{(1-\lambda)^2}{1-\lambda^2} \leq \text{const} \cdot (1-\lambda). \quad (A.6)$$

Similarly we can show that

$$\lim_{N \to \infty} |T_3| \leq \text{const} \cdot (1-\lambda). \quad (A.7)$$

The assertion of the lemma follows from (A.3)-(A.7).

ACKNOWLEDGMENT

The authors acknowledge useful discussions with Prof. L. Ljung on the subject matter of this correspondence.

REFERENCES

[1] B. Porat, "Second-order equivalence of rectangular and exponential windows in least-squares estimation of Gaussian autoregressive processes," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-33, pp. 1209-1212, Oct. 1985.
[2] M. Zarrop, "Variable forgetting factors in parameter estimation," *Automatica*, vol. 19, pp. 295-298, 1983.
[3] T. Söderström and P. Stoica, *System Identification*. London, England: Prentice-Hall International, 1988, to appear.
[4] L. Ljung and T. Söderström, *Theory and Practice of Recursive Identification*. Cambridge, MA: M.I.T. Press, 1983.
[5] L. Ljung, "On the consistency of prediction error identification methods," in *System Identification—Advances and Case Studies*, R. K. Mehra and D. G. Lainiotis, Eds. New York: Academic, 1976.
[6] H. Cramér, *Mathematical Methods in Statistics*. Princeton, NJ: Princeton University Press, 1946.
[7] L. Ljung, "Some limit results for functionals of stochastic processes," Dep. Elec. Eng., Linköping Univ., Linköping, Sweden, Rep. LiTH-ISY-I-0167, 1977.

Some Improvements on the Synchronized-Overlap-Add Method of Time Scale Modification for Use in Real-Time Speech Compression and Noise Filtering

JAMES L. WAYMAN AND DENNIS L. WILSON

***Abstract*—The synchronized-overlap-add (SOLA) method of time scale modification as developed by Roucos and Wilgus [1] has proven to be a useful technique for speech compression both alone and when cascaded with other methods. We present here a modification of the method which, while costing approximately 3 percent in compression ratio, increases the computational speed of the method by 30 percent and results in much higher quality transmission. We also show how the method can be used as a correlation filter for removing white noise from speech.**

I. INTRODUCTION

The synchronous-overlap-add (SOLA) method of time scale modification (TSM) is a recent development in speech compression [1], [2] and shows great promise because of its simplicity, allowing potentially real-time implementation, its high-quality upon reexpansion, preserving the original pitch and formant structure, and its independence of other (LPC, FFT, VQ) methods, allowing it to be cascaded with these methods for achieving very high compression ratios. The method proves not only to be robust in the presence of white noise, but because it is a correlation method, to actually improve the signal-to-noise ratio of a corrupted input signal.

Manuscript received February 7, 1987; revised August 10, 1987.
The authors are with the Ford Aerospace and Communications Corporation, Palo Alto, CA 94303.
IEEE Log Number 8717677.

The SOLA method begins with overlapping frames of time domain data. The frame length is N, the amount of new data per frame is S_a, and then the amount of overlap is $N - S_a$. Compression can be achieved by "accordioning" these frames (shifting to increase the amount of overlap) and averaging such that the amount of new data per frame is S_s (where $S_s < S_a$). For expansion, this shifting is done to decrease the amount of overlap ($S_s > S_a$). The ratio of S_s/S_a is the modification factor α. For compression, $\alpha < 1$, while for expansion, $\alpha > 1$.

If the accordioning is of the same amount for all frames, the resulting speech is of poor quality. The contribution of the SOLA method is to vary the degree of accordioning such that the amount of new data per frame averages S_s, but is allowed to vary by an amount k_m at any particular overlapping where m is the frame number. The k_m, which could have as easily been called ΔS_s, may take on any value between $\pm N/2$. It is chosen such that the cross correlation between the new frame and the averaged values of the previous frames is maximized. That is, k_m is the value of k that maximizes

$$R_m(k) = \frac{\sum_{j=0}^{L-1} y(mS_s + k + j)\, x(mS_a + j)}{\left[\sum_{j=0}^{L-1} y^2(mS_s + k + j) \sum_{j=0}^{L-1} x^2(mS_a + j) \right]^{1/2}}$$
$$-N/2 \leq k \leq N/2 \quad (1)$$

as given in [2] where L is the length of overlap between the new signal samples $x(mS_a + j)$ and the composite vector y formed by averaging previous overlapped vectors.

The vector y is updated by each new vector x once k_m is found by the formula

$$y(mS_s + k_m + j) = (1 - f(j))\, y(mS_s + k_m + j) + f(j)\, x(mS_a + j) \quad \text{for } 0 \leq j \leq L_m - 1 \quad (2)$$

and

$$y(mS_s + k_m + j) = x(mS_a + j) \quad \text{for } L_m \leq j \leq N - 1 \quad (3)$$

where $f(j)$ is a weighting function to be discussed and L_m is the length of the overlap of the two vectors x and y for the particular k_m involved. The algorithm is initialized by setting the first $y(j) = x(j)$, $j = 1 \cdots N$. This implies that k_1 will always be $-S_s$ (for nontrivial signals) because the maximum cross correlation will occur when each $y(mS_s + k + j) = x(mS_a + j)$ for $j = 1 \cdots N$, and this happens the very first time through the algorithm when $k = -S_s$.

The compression ratio is asymptotically S_s/S_a, the amount of new data in each output frame divided by the amount of new data in each overlapped input frame.

II. IMPROVED METHOD OF REEXPANSION

The problem of reexpanding the compressed signal then emerges. In [1] and [2], reexpansion was accomplished by executing the SOLA algorithm in the expansion mode on the compressed data, with the value of α during expansion equal to the inverse of the value used during compression. We have noticed in our experimentation that it is expansion that causes the "reverberant" quality of the reexpanded speech noted in [2]. To minimize this form of signal degradation, we have developed another expansion method that not only improves the quality of the reexpanded signal, but is faster to compute than SOLA expansion. In our method, the value of the variable compression factor k_m is sent along with each S_s samples of the compressed signal. Sending this value, of course,