

# Adaptive interpolation of discrete-time signals that can be modeled as autoregressive processes

**Citation for published version (APA):**

Janssen, A. J. E. M., Veldhuis, R. N. J., & Vries, L. B. (1986). Adaptive interpolation of discrete-time signals that can be modeled as autoregressive processes. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 34(2), 317-330. <https://doi.org/10.1109/TASSP.1986.1164824>

**DOI:**

[10.1109/TASSP.1986.1164824](https://doi.org/10.1109/TASSP.1986.1164824)

**Document status and date:**

Published: 01/01/1986

**Document Version:**

Publisher's PDF, also known as Version of Record (includes final page, issue and volume numbers)

**Please check the document version of this publication:**

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

**General rights**

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

[www.tue.nl/taverne](http://www.tue.nl/taverne)

**Take down policy**

If you believe that this document breaches copyright please contact us at:

[openaccess@tue.nl](mailto:openaccess@tue.nl)

providing details and we will investigate your claim.

# Adaptive Interpolation of Discrete-Time Signals That Can Be Modeled as Autoregressive Processes

A. J. E. M. JANSSEN, RAYMOND N. J. VELDHUIS, AND LODEWIJK B. VRIES

**Abstract**—This paper presents an adaptive algorithm for the restoration of lost sample values in discrete-time signals that can locally be described by means of autoregressive processes. The only restrictions are that the positions of the unknown samples should be known and that they should be embedded in a sufficiently large neighborhood of known samples. The estimates of the unknown samples are obtained by minimizing the sum of squares of the residual errors that involve estimates of the autoregressive parameters. A statistical analysis shows that, for a burst of lost samples, the expected quadratic interpolation error per sample converges to the signal variance when the burst length tends to infinity. The method is in fact the first step of an iterative algorithm, in which in each iteration step the current estimates of the missing samples are used to compute the new estimates. Furthermore, the feasibility of implementation in hardware for real-time use is established. The method has been tested on artificially generated autoregressive processes as well as on digitized music and speech signals.

## I. INTRODUCTION

**T**HIS paper treats the problem of restoring (or interpolating) unknown or lost sample values in a discrete-time signal. An algorithm is presented that is capable of restoring satisfactorily unknown samples with known positions occurring in bursts and more general patterns. Examples of both cases are shown in Fig. 1. To restore the unknown samples, the algorithm uses the information contained in the known neighboring samples.

Until rather recently, the problem of estimating unknown sample values in discrete-time signals in real time could only be solved by relatively simple, nonadaptive methods, such as Lagrange-type curve fitting. These methods are not well suited for signals primarily containing harmonic components, especially not when the number of samples in the periods of the harmonic components is less than the number of unknown samples. For instance, linear interpolation gives already audible interpolation errors for bursts in digital audio signals of length 5. Because of the progress made in the field of chip design, one can now contemplate more complicated real-time restoration methods that may also involve some signal model. An example of such a method, where no model is assumed, can be found in [1]. This (adaptive) method deals with restoring discrete-time signals of which every  $n$ th sample is unknown. Examples of methods that interpolate under certain model assumptions have been given

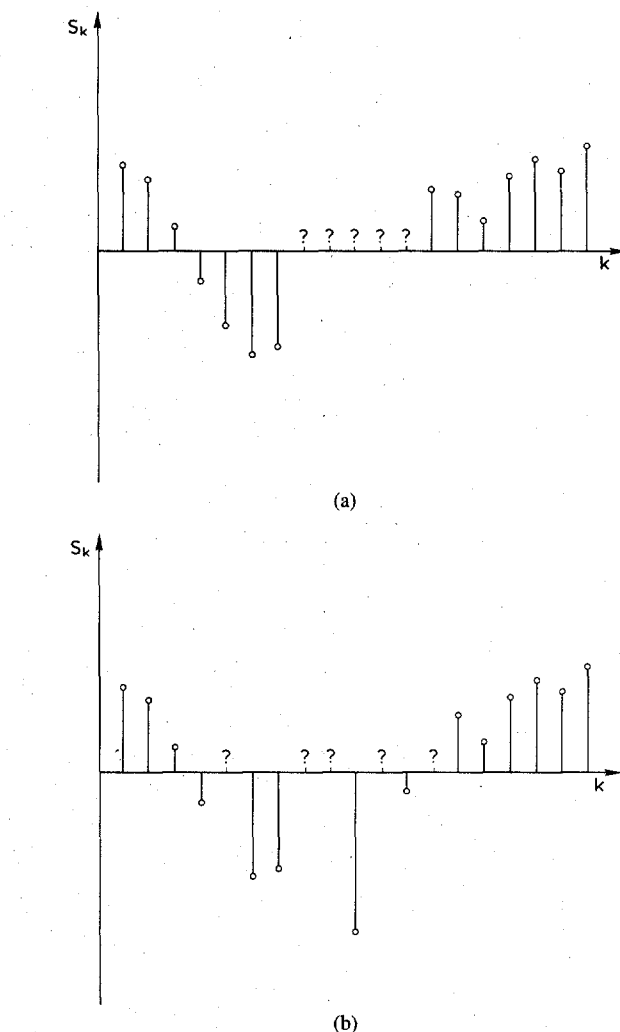


Fig. 1. (a) Sequence containing a burst of unknown samples. (b) Sequence containing a random pattern of unknown samples.

in [2, Section II] and in [3]. Reference [2], in which the assumed model is an autoregressive process, deals with the restoration of a single unknown sample by minimizing the expected quadratic interpolation error. The assumed model in [3] is band-limitedness of the signals to a baseband which is a fraction of the sample frequency. In this nonadaptive method, analyzed in [3] for burst errors, the restoration is done in such a way that the restored signal has minimal energy outside the prescribed baseband. Unfortunately, the latter method is very sensitive to the presence of noise and of out-of-band components in the sig-

Manuscript received March 2, 1984; revised September 9, 1985.  
The authors are with the Philips Research Laboratories, 5600 JA Eindhoven, The Netherlands.  
IEEE Log Number 8406769.

nal, even when the number of lost samples is small and the spectral energy of the signal is well within the baseband.

In this paper, the same point of view as in [2, Section II] is taken for the restoration of more general patterns of unknown samples than single ones or bursts. That is, it is assumed that the signals to be interpolated can be modeled as autoregressive (AR) processes of finite order. The restoration is done in such a way that the restored signal fits the assumed model as well as possible.

The method is adaptive in the sense that, from a finite segment of data, one first has to estimate the AR parameters. Once these are known, the unknown samples can be obtained as the solutions of a system of linear equations. In fact, the AR parameters as well as the unknown samples could be obtained in one step by minimizing some function involving both AR parameters and unknown samples. However, this function contains fourth-order terms and minimizing it is a nontrivial problem. Here a suboptimal approach is adopted, where first the parameters are estimated from the incomplete data and next the unknown samples. This can be considered to be the first step of a rapidly converging iterative minimization procedure. This procedure will be discussed in an appendix.

The choice of the autoregressive process as a model for the signal can be motivated by the fact that many signals that are encountered in practice can be modeled in this way. Therefore, it is expected that the interpolation method presented here can be applied successfully in many practical situations. For instance, as will be demonstrated further on, good results are obtained for the interpolation of digitized music and speech signals.

The organization of this paper is as follows. In Section II, the interpolation method is presented and a statistical analysis is given. The interpolation error is analyzed under the assumption that the AR parameters are known; this analysis is detailed for the case that Gaussian probability density functions are assumed. In Section III, efficient methods for the approximate calculation of AR parameters and unknown samples are given. Also, the numerical properties of certain parts of the algorithm are discussed. In Section IV, some results are presented. Here a comparison of performance to other methods is given. Section V presents some conclusions. Finally, the paper contains a number of appendixes to which proofs not relevant to the main text and much of the mathematics are deferred.

## II. PRESENTATION AND ANALYSIS OF THE INTERPOLATION METHOD

In this section, it is assumed that  $s_k$ ,  $k = -\infty, \dots, \infty$ , is a realization of a stationary autoregressive process  $\tilde{s}_k$ ,  $k = -\infty, \dots, \infty$  (the tilde indicates that a variable is a stochastic variable). This means that there exist a finite positive integer  $p$ , the prediction order, numbers  $a_0, a_1, \dots, a_p$ ,  $a_0 = 1$ , the prediction coefficients, and a zero-mean white noise process  $\tilde{e}_k$ ,  $k = -\infty, \dots, \infty$ , the excitation noise, with variance  $\sigma_e^2$ , such that

$$a_0 \tilde{s}_k + a_1 \tilde{s}_{k-1} + \dots + a_p \tilde{s}_{k-p} = \tilde{e}_k, \quad k = -\infty, \dots, \infty. \quad (\text{II.1})$$

For notational convenience, it shall be agreed that  $a_k = 0$  for  $k < 0$  or  $k > p$ . The AR spectrum  $S(\theta)$  of  $\tilde{s}_k$ ,  $k = -\infty, \dots, \infty$ , is given by

$$S(\theta) = \frac{\sigma_e^2}{\left| \sum_{l=0}^p a_l \exp(-j\theta l) \right|^2} = \frac{\sigma_e^2}{\sum_{l=-p}^p b_l \exp(-j\theta l)} \quad (\text{II.2})$$

where

$$b_l = \sum_{k=0}^p a_k a_{k+l}. \quad (\text{II.3})$$

In Section II-A, the algorithm for estimating the AR parameters and the unknown samples from a finite sequence of samples is presented. A statistical analysis of the interpolation error is given in II-B.

### A. Presentation of the Interpolation Method

The available data consist of a segment  $s_k$ ,  $k = 0, \dots, N-1$ , of a realization of an AR process  $\tilde{s}_k$ ,  $k = -\infty, \dots, \infty$ . It is assumed throughout that the unknown samples occur at the known time instants  $t(1), \dots, t(m)$ , where  $0 < p \leq t(1) < \dots < t(m) \leq N-p-1$ . The problem is to estimate the values of the unknown samples  $s_{t(1)}, \dots, s_{t(m)}$  and the AR parameters  $p, a_1, \dots, a_p$  and  $\sigma_e^2$  from the available data in such a way that the restored segment fits the assumed model as well as possible in a quadratic sense. That is, the restoration is such that the sum of the squares of the residual error  $e_p, \dots, e_{N-1}$  is minimal.

Although methods to estimate the order of an autoregressive process have been reported [4], it has been decided, if  $p$  is unknown, to choose  $p$  as a function of the number  $m$  of unknown samples. The rather arbitrary relation  $p = 3m + 2$  has proved to give good interpolation results. For notational convenience, the vector notation  $\mathbf{a} = [a_1, \dots, a_p]^T$ ,  $\mathbf{x} = [s_{t(1)}, \dots, s_{t(m)}]^T$  (the superscript  $T$  denotes vector or matrix transposition) shall be adopted. The estimation of  $\mathbf{a}$  and  $\mathbf{x}$  is expressed as a minimization problem, where the estimates  $\hat{\mathbf{a}}$  for  $\mathbf{a}$  and  $\hat{\mathbf{x}}$  for  $\mathbf{x}$  are chosen such that

$$Q(\mathbf{a}, \mathbf{x}) = \sum_{k=p}^{N-1} \left| \sum_{l=0}^p a_l s_{k-l} \right|^2 = \sum_{k=p}^{N-1} |e_k|^2 \quad (\text{II.4})$$

is minimal as a function of  $\mathbf{a}$  and  $\mathbf{x}$ . Once  $\hat{\mathbf{a}}$  and  $\hat{\mathbf{x}}$  have been determined,  $\sigma_e^2$  is estimated by

$$\hat{\sigma}_e^2 = \frac{1}{N-p-m} Q(\hat{\mathbf{a}}, \hat{\mathbf{x}}). \quad (\text{II.5})$$

The particular choice for minimizing  $Q(\mathbf{a}, \mathbf{x})$  to obtain estimates for  $\mathbf{a}$  and  $\mathbf{x}$  is motivated by the following two facts. First, if  $\mathbf{s} = [s_0, \dots, s_{N-1}]^T$  and  $\mathbf{u} = [s_0, \dots, s_{p-1}]^T$  then, under the hypothesis that the sample values have a Gaussian probability density function, minimizing

$Q(\mathbf{a}, \mathbf{x})$  with respect to  $\mathbf{a}$  turns out to be the same as maximizing the log likelihood function

$$L(\mathbf{a}, \sigma_e^2) = \log(p_{\hat{s}|\hat{u}}(s|\mathbf{u}, \mathbf{a}, \sigma_e^2)) \quad (\text{II.6})$$

as a function of  $\mathbf{a}$  and  $\sigma_e^2$ . This is a common procedure to estimate  $\mathbf{a}$  and  $\sigma_e^2$ . Second, also under the hypothesis that the sample values have a Gaussian probability density function, minimizing  $Q(\mathbf{a}, \mathbf{x})$  with respect to  $\mathbf{x}$  is the same as finding a minimum variance estimate for  $\mathbf{x}$ , for known  $p$  and  $\mathbf{a}$ . Both claims shall be proved in Appendix A.

Since  $Q(\mathbf{a}, \mathbf{x})$  involves fourth-order terms, such as  $a_i^2 s_{i(m)}^2$ , the minimization with respect to  $\mathbf{a}$  and  $\mathbf{x}$  is a non-trivial problem. Fortunately, one can often assume that the number  $m$  of unknown samples is small compared to the segment length  $N$ . A suboptimal approach can then be found as follows. One chooses an initial estimate  $\hat{\mathbf{x}}^{(0)}$ , for instance,  $\hat{\mathbf{x}}^{(0)} = \mathbf{0}$ , for the vector  $\mathbf{x}$  of the unknown samples. Next, one minimizes  $Q(\mathbf{a}, \hat{\mathbf{x}}^{(0)})$  as a function of  $\mathbf{a}$  to obtain an estimate  $\hat{\mathbf{a}}$ . Finally, one minimizes  $Q(\hat{\mathbf{a}}, \mathbf{x})$  as a function of  $\mathbf{x}$  to obtain an estimate  $\hat{\mathbf{x}}$  for the unknown samples.

Both minimizations are feasible, since  $Q(\mathbf{a}, \mathbf{x})$  is a quadratic form in both  $\mathbf{a} \in \mathbb{R}^p$  and  $\mathbf{x} \in \mathbb{R}^m$ . In fact, it can be shown that

$$Q(\mathbf{a}, \mathbf{x}) = \mathbf{a}^T C(\mathbf{x}) \mathbf{a} + 2\mathbf{a}^T \mathbf{c}(\mathbf{x}) + c_{00}(\mathbf{x}). \quad (\text{II.7})$$

Here

$$\begin{aligned} C(\mathbf{x}) &= (c_{ij}(\mathbf{x}))_{i,j=1,\dots,p}, \\ \mathbf{c}(\mathbf{x}) &= [c_{01}(\mathbf{x}), \dots, c_{0p}(\mathbf{x})]^T, \end{aligned} \quad (\text{II.8})$$

where

$$c_{ij}(\mathbf{x}) = \sum_{k=p}^{N-1} s_{k-i} s_{k-j}, \quad i, j = 0, 1, \dots, p. \quad (\text{II.9})$$

Hence,  $C(\mathbf{x})$  is the  $p \times p$ -autocovariance matrix, estimated from  $s_k, k = 0, \dots, N-1$ . At the same time, it can be shown that

$$Q(\mathbf{a}, \mathbf{x}) = \mathbf{x}^T B(\mathbf{a}) \mathbf{x} + 2\mathbf{x}^T \mathbf{z}(\mathbf{a}) + D(\mathbf{a}). \quad (\text{II.10})$$

Here

$$\begin{aligned} B(\mathbf{a}) &= (b_{l(i)} - b_{l(j)})_{i,j=1,\dots,m}, \\ \mathbf{z}(\mathbf{a}) &= [z_1(\mathbf{a}), \dots, z_m(\mathbf{a})]^T, \end{aligned} \quad (\text{II.11})$$

$b_l, l = -p, \dots, p$ , has been defined in (II.3), and

$$z_i(\mathbf{a}) = \sum_{k=-p}^p b_k s_{i(i)-k}, \quad i = 1, \dots, m, \quad (\text{II.12})$$

and  $D(\mathbf{a}) \in \mathbb{R}$  depends on  $\mathbf{a}$  and the known samples only. Hence,  $\hat{\mathbf{a}}$  and  $\hat{\mathbf{x}}$  are given by

$$C(\hat{\mathbf{x}}^{(0)})\hat{\mathbf{a}} = -\mathbf{c}(\hat{\mathbf{x}}^{(0)}), \quad (\text{II.13})$$

and

$$B(\hat{\mathbf{a}})\hat{\mathbf{x}} = -\mathbf{z}(\hat{\mathbf{a}}), \quad (\text{II.14})$$

respectively. The above method for calculating prediction coefficients from a sequence of samples is known as the

autocovariance method [5]. On substitution of (II.13) into (II.5), it easily follows that

$$\hat{\sigma}_e^2 = \frac{1}{N-p-m} (c_{00}(\hat{\mathbf{x}}) + \hat{\mathbf{a}}^T \mathbf{c}(\hat{\mathbf{x}})). \quad (\text{II.15})$$

It should be noted at this point that the interpolation method just described can be considered as the first step of an iterative algorithm in which, in every step, new prediction coefficients  $\hat{\mathbf{a}}$  are estimated as in (II.13) by using, instead of  $\hat{\mathbf{x}}^{(0)}$ , the previously estimated vector of sample values  $\hat{\mathbf{x}}$  obtained in (II.14). The prediction coefficients can be used again to obtain new estimates for the unknown samples and so on. It is clear that in this way  $Q(\mathbf{a}, \mathbf{x})$  decreases to some nonnegative number. One may hope that the sequence thus obtained converges to a point where  $Q(\mathbf{a}, \mathbf{x})$  attains its global minimum. Unfortunately, it seems very hard to prove any definite result in this direction. However, it can be shown that this iterative minimization procedure closely resembles a maximum likelihood parameter estimation algorithm, well known in statistics: the EM algorithm [6]–[8]. The iterative version of the interpolation algorithm and its resemblance to the EM algorithm are discussed in Appendix B.

The interpolating vector  $\hat{\mathbf{x}}$  of (II.14) can also be obtained as the solution to a minimization problem in the frequency domain. Denote by  $\hat{S}(\theta)$  the AR spectrum obtained by substituting  $\hat{\sigma}_e^2$  of (II.15) and  $\hat{\mathbf{a}}$  of (II.13) into the expression for  $S(\theta)$  in (II.2). In Appendix A, it is proved that  $\hat{\mathbf{x}}$  of (II.14) minimizes, as a function of  $\mathbf{x} \in \mathbb{R}^m$ , the integral

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{S(\theta; \mathbf{x})}{\hat{S}(\theta)} d\theta. \quad (\text{II.16})$$

Here

$$\begin{aligned} S(\theta; \mathbf{x}) &= \frac{1}{N-2p} \left| \sum_{k=p}^{N-p-1} s_k \exp(-j\theta k) \right|^2, \\ -\pi &\leq \theta \leq \pi. \end{aligned} \quad (\text{II.17})$$

Intuitively, by minimizing (II.16) with respect to the values of the unknown samples, one forces the restored signal to have little (much) spectral energy in those regions in the frequency domain where the estimated spectral energy is small (large). This brings out a relation with the interpolation method of [3], where the restoration is such that the spectral energy of the restored signal is concentrated as much as possible in the assumed baseband of the original signal. It should be noted that the integral in (II.16) can be related with the work of Itakura and Saito [9] on distortion measures for spectral densities.

### B. Statistical Analysis of the Interpolation Error

In this subsection, some statistical properties of the interpolation error are discussed. It is assumed that  $p, \mathbf{a}$ , and  $\sigma_e^2$  are known. Since, in practice, these parameters are estimated from the data, this assumption may be a simplification from reality. However, it has the advantage that the results take a pleasant form.

The interpolation error is defined as the stochastic vector  $\tilde{\mathbf{d}}$ ,

$$\tilde{\mathbf{d}} = \hat{\mathbf{x}} - \tilde{\mathbf{x}} = \hat{\mathbf{x}} + (\mathbf{B}(\mathbf{a}))^{-1} \tilde{\mathbf{z}}(\mathbf{a}). \quad (\text{II.18})$$

Note that the realization  $\mathbf{z}(\mathbf{a})$  of (II.14) is replaced by a stochastic vector  $\tilde{\mathbf{z}}(\mathbf{a})$ . It follows easily from (II.18), and from the fact that  $E[\tilde{s}_k] = 0$ , that  $E[\tilde{\mathbf{d}}] = \mathbf{0}$  and that the estimator  $\hat{\mathbf{x}}$  is unbiased. The (stochastic) relative quadratic interpolation error per sample  $\tilde{e}$  is defined by

$$\tilde{e} = \frac{\tilde{\mathbf{d}}^T \tilde{\mathbf{d}}}{m E[\tilde{s}_k^2]}. \quad (\text{II.19})$$

To evaluate the expectation  $E[\tilde{e}]$  of  $\tilde{e}$ , it is noted that

$$\tilde{\mathbf{d}} = (\mathbf{B}(\mathbf{a}))^{-1} (\tilde{\mathbf{z}}(\mathbf{a}) + \mathbf{B}(\mathbf{a})\tilde{\mathbf{x}}) =: (\mathbf{B}(\mathbf{a}))^{-1} \tilde{\mathbf{w}}, \quad (\text{II.20})$$

and that for  $i = 1, \dots, m$ ,

$$\tilde{w}_i = \sum_{k=-p}^p b_k \tilde{s}_{i(i)-k} = \sum_{l=0}^p a_l \tilde{e}_{i(i)+l} \quad (\text{II.21})$$

as follows straightforwardly from the definitions in (II.3), (II.11), and (II.12). Thus,

$$E[\tilde{\mathbf{d}}\tilde{\mathbf{d}}^T] = (\mathbf{B}(\mathbf{a}))^{-1} E[\tilde{\mathbf{w}}\tilde{\mathbf{w}}^T] (\mathbf{B}(\mathbf{a}))^{-1}. \quad (\text{II.22})$$

Since  $(E[\tilde{\mathbf{w}}\tilde{\mathbf{w}}^T])_{ij} = E[\tilde{w}_i\tilde{w}_j] = \sigma_e^2 b_{i(i)-i(j)}$ , one has that  $E[\tilde{\mathbf{w}}\tilde{\mathbf{w}}^T] = \sigma_e^2 \mathbf{B}(\mathbf{a})$  and that

$$E[\tilde{\mathbf{d}}\tilde{\mathbf{d}}^T] = \sigma_e^2 (\mathbf{B}(\mathbf{a}))^{-1}. \quad (\text{II.23})$$

Finally,  $E[\tilde{e}]$  is given by

$$E[\tilde{e}] = \frac{\sigma_e^2}{m E[\tilde{s}_k^2]} \text{trace} ((\mathbf{B}(\mathbf{a}))^{-1}). \quad (\text{II.24})$$

For the expected relative quadratic interpolation error of the  $i$ th unknown sample, one has

$$E[\tilde{d}_i^2] = \sigma_e^2 ((\mathbf{B}(\mathbf{a}))^{-1})_{ii}, \quad i = 1, \dots, m. \quad (\text{II.25})$$

The case of a burst of  $m$  consecutive unknown samples deserves somewhat more attention than the general case. Then the matrix  $\mathbf{B}(\mathbf{a})$  is Toeplitz and therefore has some properties that facilitate a further analysis of the interpolation error. Toeplitz matrices are persymmetric: an  $n \times n$ -matrix  $\mathbf{M}$  is persymmetric if  $M_{ij} = M_{n+1-j, n+1-i}$ ,  $i, j = 1, \dots, n$ . It is a property of persymmetric matrices that their inverses are also persymmetric. If  $\mathbf{B}(\mathbf{a})$  is Toeplitz then  $(\mathbf{B}(\mathbf{a}))^{-1}$  is persymmetric, and

$$E[\tilde{d}_i^2] = E[\tilde{d}_{m+1-i}^2], \quad i = 1, \dots, m. \quad (\text{II.26})$$

Extensive observations for the case of a burst of  $m$  unknown samples have revealed that the  $((\mathbf{B}(\mathbf{a}))^{-1})_{ii}$ ,  $i = 1, \dots, m$ , seem to depend quadratically on  $i$  for  $m$  not too large, and that the  $((\mathbf{B}(\mathbf{a}))^{-1})_{ii}$  tend to have their maximum for  $i \cong m/2$ , i.e., in the middle of the burst. Hence, much of the error energy is usually concentrated in the middle of the burst.

In case of a burst, the asymptotic behavior of  $E[\tilde{e}]$  as  $m \rightarrow \infty$  can be determined by applying the Szegö limit theorem [10]. From (II.24), one has

$$E[\tilde{e}] = \frac{\sigma_e^2}{m E[\tilde{s}_k^2]} \sum_{i=1}^m \lambda_i^{-1} \quad (\text{II.27})$$

where  $\lambda_i$  is the  $i$ th eigenvalue of  $\mathbf{B}(\mathbf{a})$ . According to the Szegö limit theorem, one has, for any function  $F$ , continuous on the set  $\{\sum_{k=-p}^p b_k \exp(-j\theta k) \mid |\theta| < \pi\}$ ,

$$\lim_{m \rightarrow \infty} \frac{1}{m} \sum_{i=1}^m F(\lambda_i) = \frac{1}{2\pi} \int_{-\pi}^{\pi} F\left(\sum_{k=-p}^p b_k \exp(-j\theta k)\right) d\theta. \quad (\text{II.28})$$

Taking  $F(a) = a^{-1}$ , one finds by using (II.2),

$$\begin{aligned} \lim_{m \rightarrow \infty} \frac{1}{m} \sum_{i=1}^m \lambda_i^{-1} &= \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{1}{\sum_{k=-p}^p b_k \exp(-j\theta k)} d\theta \\ &= \frac{1}{\sigma_e^2} \frac{1}{2\pi} \int_{-\pi}^{\pi} S(\theta) d\theta \\ &= \frac{E[\tilde{s}_k^2]}{\sigma_e^2}. \end{aligned} \quad (\text{II.29})$$

Hence,

$$\lim_{m \rightarrow \infty} E[\tilde{e}] = 1. \quad (\text{II.30})$$

This shows that for long bursts of consecutive samples, the quadratic interpolation error per sample approaches the signal energy per sample.

The result (II.30) derived for the burst case is also useful for finding a bound on the interpolation error in the general case. Indeed, the matrix  $\mathbf{B}(\mathbf{a}) = (b_{i(i)-i(j)})_{i,j=1,\dots,m}$  is a principal submatrix of the  $(t(m) - t(1) + 1) \times (t(m) - t(1) + 1)$  Toeplitz matrix  $\mathbf{B}'(\mathbf{a}) = (b_{(i-l)k}, l=1, \dots, t(m)-t(1)+1)$ . Denoting the first  $m$  eigenvalues of  $\mathbf{B}(\mathbf{a})$  and  $\mathbf{B}'(\mathbf{a})$  in increasing order by  $\lambda_1, \dots, \lambda_m$  and  $\lambda'_1, \dots, \lambda'_m$ , one has by [11, Section 3.5, Theorem 5.6] that  $0 < \lambda'_i < \lambda_i$ ,  $i = 1, \dots, m$ . Hence,

$$\begin{aligned} \text{trace} ((\mathbf{B}(\mathbf{a}))^{-1}) &= \sum_{i=1}^m \lambda_i^{-1} < \sum_{i=1}^m \lambda'_i^{-1} \\ &< \text{trace} ((\mathbf{B}'(\mathbf{a}))^{-1}), \end{aligned} \quad (\text{II.31})$$

and it follows that  $E[\tilde{e}]$  is asymptotically bounded by  $\lim_{m \rightarrow \infty} \sup m^{-1}(t(m) - t(1) + 1)$ . Although this bound is not as good as for the burst case, the interpolation error in the case of  $m$  randomly positioned unknown samples usually turns out to be smaller than in the case of a burst of length  $m$ .

The interpolation error can be analyzed in some more detail if  $\tilde{e}_k$  has a Gaussian probability density function. It then follows that  $\tilde{\mathbf{d}}$  has a probability density function

$$p_{\tilde{\mathbf{d}}}(\mathbf{d}) = \frac{|\mathbf{B}(\mathbf{a})|^{1/2}}{(2\pi)^{m/2} \sigma_e^m} \exp\left(-\frac{\mathbf{d}^T \mathbf{B}(\mathbf{a}) \mathbf{d}}{2\sigma_e^2}\right). \quad (\text{II.32})$$

It is a rather tedious but straightforward exercise to calculate the variance,  $\text{var}(\bar{e}) = E[(\bar{d}^T \bar{d} - E[\bar{d}^T \bar{d}])^2]$ , of  $\bar{e}$ :

$$\text{var}(\bar{e}) = \frac{\sigma_e^4}{m E[\bar{s}_k^2]^2} \sum_{i=1}^m \lambda_i^{-2}. \quad (\text{II.33})$$

In the case of a burst of unknown samples of length  $m$ , one can use the Szegő limit theorem (II.28) with  $F(a) = a^{-2}$ . For large  $m$ , one finds

$$\text{var}(\bar{e}) = 2 \frac{\frac{1}{2\pi} \int_{-\pi}^{\pi} |S(\theta)|^2 d\theta}{\left| \frac{1}{2\pi} \int_{-\pi}^{\pi} S(\theta) d\theta \right|^2}. \quad (\text{II.34})$$

It can be observed that  $\text{var}(\bar{e})$  is larger if the signal spectrum  $S(\theta)$  is more peaky.

### III. COMPUTATIONAL ASPECTS OF THE INTERPOLATION ALGORITHM

In this section, the computational aspects of the calculation of  $\hat{a}$  in (II.13) and  $\hat{x}$  in (II.14) are considered. It should be noted that a linear system needs to be solved for the calculation of both  $\hat{a}$  and  $\hat{x}$ . If  $p$  is chosen  $3m + 2$ , which is done when  $p$  is unknown, the need for efficiency is more urgent for the calculation of  $\hat{a}$  than for the calculation of  $\hat{x}$ .

#### A. Calculation of the Prediction Coefficients

The calculation of  $\hat{a}$  in (II.13) is, in fact, a well-known problem. It is often referred to as the autocovariance method and is discussed in great detail for instance in [5]. In this reference also, an efficient algorithm is given for solving  $\hat{a}$  from (II.13) in  $O(p^2)$  operations. It is the experience of the authors that an approximate calculation of  $\hat{a}$  also gives satisfactory interpolation results, so that other methods to calculate  $\hat{a}$  can be applied too. An example of a different method to calculate  $\hat{a}$  is the so-called autocorrelation method, where, instead of the system (II.13), the system  $R\hat{a} = -r$  is solved. Here  $R = ((r(i - j))_{i,j=1,\dots,p})$  is the  $p \times p$ -autocorrelation matrix,  $r = [r(1), \dots, r(p)]^T$  and  $r(j) = 1/N \sum_{k=0}^{N-|j|-1} s_k s_{k+|j|}$ ,  $j = -p, \dots, p$ , is a (biased) estimate for the  $j$ th autocorrelation lag of  $s_k$ ,  $k = -\infty, \dots, \infty$ . The system obtained in this way can also be solved in  $O(p^2)$  operations by the Levinson-Durbin algorithm [12]. As can be seen from the results presented in Section IV, in most cases there are no significant differences between the interpolation results obtained via the autocovariance method or via the autocorrelation method. In [5], a summary of the various methods to estimate  $a$  from a sequence of samples is given. The numerical stability of some of these methods is discussed in [13].

#### B. Calculation of the Unknown Samples

For the calculation of  $\hat{x}$  in (II.14), it makes sense to analyze the matrix  $B(\hat{a})$  defined in (II.11) and (II.3) in some detail. It can be seen from (II.11) that  $B(\hat{a})$  has constant values  $b_0$  on its main diagonal. Furthermore, the ma-

trix  $B(\hat{a})$  is positive definite, as can be seen from the expression

$$\sum_{i=1}^m \sum_{j=1}^m (B(\hat{a}))_{ij} v_i v_j = \sum_k \left| \sum_{i=1}^m \hat{a}_{k+t(i)} v_i \right|^2 \quad (\text{III.1})$$

which follows on inserting (II.11) and (II.3) into the left-hand side of (III.1). Indeed, when  $i'$  is the largest index with  $v_{i'} \neq 0$ , the term in the right-hand sum of (III.1) with  $k = -t(i')$  equals  $v_{i'}^2$ , as  $\hat{a}_l = 0$  for  $l < 0$ ,  $a_0 = 1$  and  $v_i = 0$  for  $i > i'$ . Hence, if  $v$  has nonzero elements, the right-hand sum of (III.1) consists of nonnegative terms of which at least one is positive. This shows that  $B(\hat{a})$  is positive definite.

The fact that  $B(\hat{a})$  is positive definite allows one to use Cholesky decomposition [14] of  $B(\hat{a})$  for solving  $\hat{x}$  from (II.14) in  $O(m^3)$  operations. In case of a burst of unknown samples,  $B(\hat{a})$  is Toeplitz and (II.14) can be solved in  $O(m^2)$  operations by the Levinson algorithm [15]. Even in the case of a more general pattern of unknown samples,  $B(\hat{a})$  is related to a Toeplitz matrix, so that the system in (II.14) can be solved more efficiently by using generalized Levinson algorithms [16]. However, this requires rather involved mathematics and does not lead to a less complicated hardware implementation, since the generalized Levinson algorithm to be used strongly depends on the pattern of unknown samples. For these reasons, in this paper, only the solution of  $\hat{x}$  from (II.14) by using Cholesky decomposition is considered.

In a Cholesky decomposition, the matrix  $B(\hat{a})$  is decomposed as a product

$$B(\hat{a}) = LL^T \quad (\text{III.2})$$

or as a product

$$B(\hat{a}) = \tilde{L}D\tilde{L}^T. \quad (\text{III.3})$$

In (III.2),  $L$  is a lower triangular  $m \times m$ -matrix, in (III.3)  $\tilde{L}$  is a lower triangular  $m \times m$ -matrix with constant values 1 on its main diagonal,  $D$  is a diagonal  $m \times m$ -matrix with  $D_{ii} = L_{ii}^2$ ,  $i = 1, \dots, m$ . The systems  $B(\hat{a})\hat{x} = LL^T\hat{x} = -z(\hat{a})$  and  $B(\hat{a})\hat{x} = \tilde{L}D\tilde{L}^T = -z(\hat{a})$  are now solved by subsequently solving by back substitution  $y$  and  $\tilde{y}$  from  $Ly = -z(\hat{a})$  and from  $\tilde{L}\tilde{y} = -z(\hat{a})$ , respectively, and  $\hat{x}$  from  $L^T\hat{x} = y$  and  $\tilde{L}^T\hat{x} = D^{-1}\tilde{y}$ , respectively. Both forms of Cholesky decomposition take  $O(m^3)$  operations. A drawback of the decomposition in (III.2) is that it requires the calculation of square roots. On the other hand, as is shown further on, the elements of  $L$  in (III.2) satisfy bounds that are more convenient if one has a fixed point implementation in mind.

For the elements of the matrices  $L$  and  $D$ , one has the following results:

$$1 \leq L_{jj} = D_{jj}^{1/2} \leq b_0^{1/2}, \quad j = 1, \dots, m, \quad (\text{III.4})$$

$$\sum_{i=1}^m L_{ij}^2 = b_0, \quad j = 1, \dots, m, \quad (\text{III.5})$$

so that

$$|L_{ij}| \leq (b_0 - 1)^{1/2}, \quad i = 1, \dots, j-1, \\ j = 1, \dots, m. \quad (\text{III.6})$$

On substitution of  $L_{ij} = \tilde{L}_{ij} D_{jj}^{1/2}$  into (III.6) and by using (III.4), one obtains

$$|\tilde{L}_{ij}| \leq (b_0 - 1)^{1/2}, \quad i, 1, \dots, j-1, \\ j = 1, \dots, m. \quad (\text{III.7})$$

The bounds in (III.5) and (III.6) and the right-hand bound of (III.4) can be derived by using results of [17, Section 7] and by the fact that  $(B(\hat{a}))_{jj} = b_0, j = 1, \dots, m$ . The left-hand bound in (III.4) was not known to the authors and will be derived in Appendix C.

In a fixed point implementation it is more convenient to solve the system  $B'(\hat{a})\hat{x} = -z'(\hat{a})$ , where  $B'(\hat{a}) = B(\hat{a})/b_0$  and  $z'(\hat{a}) = z(\hat{a})/b_0$ , than the system in (II.14), because the absolute values of the elements  $B'(\hat{a})$  are all bounded by 1. Then  $B'(\hat{a}) = L'L'^T = \tilde{L}D'\tilde{L}'^T$ , where  $L' = L/b_0$  and  $D' = D/b_0$ . On substituting this into (III.4), (III.5), and (III.6) one obtains

$$1/b_0^{1/2} \leq L'_{jj} = D'_{jj}{}^{1/2} \leq 1, \quad j = 1, \dots, m, \quad (\text{III.8})$$

$$\sum_{i=1}^m L'_{ij}{}^2 = b_0, \quad j = 1, \dots, m, \quad (\text{III.9})$$

or

$$|L'_{ij}| \leq 1, \quad i, j = 1, \dots, m. \quad (\text{III.10})$$

Now the  $L'L'^T$  decomposition of  $B'(\hat{a})$  has the advantage over the  $\tilde{L}D'\tilde{L}'^T$  decomposition that the absolute values of all elements of  $L'$  are bounded by 1 and that all fixed point multiplications can be performed without prescaling.

The lower bound in (III.8) is important because the elements  $L'_{jj}, j = 1, \dots, m$ , are used as divisors in the process of back substitution and accuracy will be lost if they are too small. It is the experience of the authors that, for digitized music,  $b_0$  usually has rather modest values, say,  $b_0 < 4$ , so that the  $L'_{jj}$  do not become too small.

#### IV. RESULTS

In this section, the performance of the adaptive interpolation method discussed in this paper is considered for the following test signals.

1) Artificially generated realizations of an autoregressive process of 10th-order with a peaky spectrum. Table I and Fig. 2 show the prediction coefficients and the AR spectrum. Ten statistically independent sequences of 512 samples each have been used. The excitation noise sequences are uncorrelated pseudorandom sequences with a Gaussian probability density function with zero mean and unit variance. The patterns of the unknown samples were bursts of lengths  $m = 1, 4, 16, 50$ .

2) Artificially generated realizations of an autoregressive process of 10th-order with a smooth spectrum. Table I and Fig. 2 show the prediction coefficients and the AR

TABLE I  
PREDICTION COEFFICIENTS OF THE AUTOREGRESSIVE PROCESSES USED AS TEST SIGNALS IN THIS SECTION

i	prediction coefficients	
	peaky spectrum $a_i$	smooth spectrum $a_i$
0	0.100000000000000+01	0.100000000000000+01
1	-0.84538931513857d+00	-0.71858091786779d+00
2	0.76158831819825d+00	0.55024755989824d+00
3	0.12977849626292d+00	0.79700219017466d-01
4	0.30387326386341d-02	0.15862374294460d-02
5	0.85694798229335d-01	0.38023237227971d-01
6	0.14042707569132d+00	0.52962003577618d-01
7	0.87115349165802d-01	0.27927184980177d-01
8	-0.42344306979864d-01	-0.11538422441359d-01
9	-0.23427477005693d+00	-0.54262006831786d-01
10	-0.30186789185306d-01	-0.59430061398161d-02

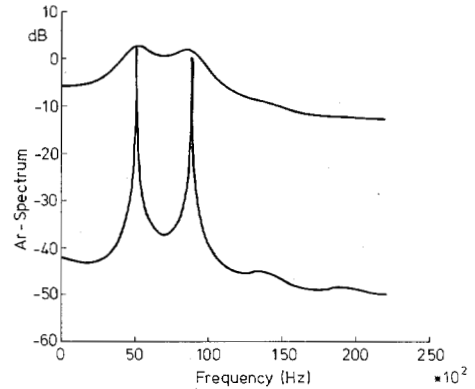


Fig. 2. AR spectra of test signals 1 and 2.

spectrum. The ten statistically independent sequences of 512 samples that have been used are generated in the same manner as the test signals described under 1). The patterns of the unknown samples were bursts of lengths  $m = 1, 4, 16, 50$ .

3) Multiple sinusoids. A sequence of 512 samples given by

$$s_n = 100 \sin(0.23\pi n + 0.3\pi) \\ + 60 \sin(0.4\pi n + 0.3\pi) \quad (\text{IV.1})$$

has been used. The patterns of the unknown samples were bursts of lengths  $m = 4, 8, 14, 16$ .

4) Digital audio signals. Bursts of 4, 6, or 16 unknown samples, occurring at a rate of  $10 \text{ s}^{-1}$  in a fragment of 36 s taken from a Compact Disc® recording of Beethoven's Violin Concert have been interpolated. The sample frequency of the signal is 44 100 Hz, so that a burst of 16 samples has a duration of 0.36 ms.

5) Digitized speech signals. Bursts of 100 unknown samples, occurring at a rate of  $10 \text{ s}^{-1}$  in 10 English sentences of male and female speech, have been interpolated. The sample frequency of the signal is 8000 Hz, so that the bursts have a duration of 12.5 ms.

6) Artificially generated realizations of an autoregressive process corrupted by pseudorandom white noise. To the sequences described under 1), pseudorandom white noise with zero mean has been added. Signal-to-noise ratios of 40 and 20 dB are considered. The pattern of unknown samples was a burst of length 16.

7) Sinusoids corrupted by pseudorandom white noise.

TABLE II

AVERAGE INTERPOLATION ERRORS AND EXPECTED INTERPOLATION ERRORS WITH KNOWN COEFFICIENTS AND AFTER 1 AND 3 ITERATIONS FOR 10 REALIZATIONS OF AN AUTOREGRESSIVE PROCESS OF ORDER 10 WITH A PEAKY SPECTRUM. THE COEFFICIENTS OF THE AUTOREGRESSIVE PROCESS ARE GIVEN IN TABLE I

m	p	N	f	c <sub>1</sub>	c <sub>3</sub>	r <sub>1</sub>	r <sub>3</sub>	E[ $\bar{e}$ ]	var( $\bar{e}$ )
1	10		0.79E-02	0.90E-02	0.72E-02	0.90E-02	0.79E-02	0.34E-02	0.23E-04
4	10	512	0.12E-01	0.11E-01	0.13E-01	0.13E-01	0.12E-01	0.21E-02	0.10E-03
16	10	512	0.26E-01	0.27E-01	0.26E-01	0.28E-01	0.26E-01	0.12E-01	0.67E-03
50	10	512	0.60E-01	0.61E-01	0.56E-01	0.62E-01	0.57E-01	0.24E-01	0.60E-02

TABLE III

AVERAGE INTERPOLATION ERRORS AND EXPECTED INTERPOLATION ERRORS WITH KNOWN COEFFICIENTS AND AFTER 1 AND 3 ITERATIONS FOR 10 REALIZATIONS OF AN AUTOREGRESSIVE PROCESS OF ORDER 10 WITH A SMOOTH SPECTRUM. THE COEFFICIENTS OF THE AUTOREGRESSIVE PROCESS ARE GIVEN IN TABLE I

m	p	N	f	c <sub>1</sub>	c <sub>3</sub>	r <sub>1</sub>	r <sub>3</sub>	E[ $\bar{e}$ ]	var( $\bar{e}$ )
1	10		0.31E+00	0.34E+00	0.32E+00	0.34E+00	0.32E+00	0.27E+00	0.15E+00
4	10	512	0.50E+00	0.62E+00	0.62E+00	0.62E+00	0.61E+00	0.51E+00	0.76E+00
16	10	512	0.99E+00	0.10E+01	0.10E+00	0.10E+01	0.10E+01	0.84E+00	0.29E+01
50	10	512	0.11E+01	0.11E+01	0.11E+01	0.11E+01	0.11E+01	0.95E+00	0.39E+01

TABLE IV

INTERPOLATION ERRORS AFTER 1 AND 3 ITERATIONS FOR A SUM OF 2 SINE WAVES. IN THE CASE c<sub>3</sub>, p = 10 THE PREDICTION COEFFICIENTS CANNOT BE CALCULATED BECAUSE THE MATRIX C( $\hat{x}$ ) IN (II.13) IS SINGULAR. THE SINUSOIDS ARE GIVEN BY (IV.1)

m	p	N	c <sub>1</sub>	c <sub>3</sub>	r <sub>1</sub>	r <sub>3</sub>
16	4	512	0.64E-02	0.28E-28	0.78E-01	0.16E-01
16	10	512	0.16E-06	XXXXXXXX	0.30E-05	0.17E-06

To the sequences described under test signal 3, pseudo-random white noise has been added. Signal-to-noise ratios of 40 and 20 dB are considered. The patterns of the unknown samples were bursts of lengths m = 4, 8, 14, 16.

The test signals 1-7 have been interpolated with the aid of the two following versions of the adaptive interpolation method.

1) The adaptive interpolation method that uses the autocovariance method to obtain the prediction coefficients (cf. Section III). This method is denoted by c<sub>i</sub>, where the subscript i denotes the number of iterations.

2) The adaptive interpolation method that uses the autocorrelation method to obtain the prediction coefficients (cf. Section III). This method is denoted by r<sub>i</sub>, where the subscript i denotes the number of iterations.

For all test signals, the performances of the adaptive interpolation methods are judged by means of the relative quadratic interpolation error e

$$e = \frac{\frac{1}{m} \sum_{i=1}^m (\hat{s}_{i(i)} - s_{i(i)})^2}{\frac{1}{N} \sum_{i=1}^N s_i^2} \quad (IV.2)$$

This is the realization of the stochastic relative quadratic interpolation error  $\bar{e}$ , defined in (II.19). The (averaged) value of e is presented for the test signals in Tables II-XI. For the test signals 1 and 2, the values of E[ $\bar{e}$ ] and of var( $\bar{e}$ ), given in (II.24) and (II.33), respectively, are

TABLE V

AVERAGE INTERPOLATION ERRORS AFTER 1 AND 3 ITERATIONS FOR 10 SHORT (N = 64) REALIZATIONS OF AN AUTOREGRESSIVE PROCESS OF ORDER 10 WITH A PEAKY SPECTRUM. THE COEFFICIENTS OF THE AUTOREGRESSIVE PROCESS ARE GIVEN IN TABLE I

m	p	N	c <sub>1</sub>	c <sub>3</sub>	r <sub>1</sub>	r <sub>3</sub>
16	10	64	0.42E-01	0.23E-01	0.63E-01	0.25E-01

TABLE VI

INTERPOLATION ERRORS AFTER 1 AND 3 ITERATIONS FOR A SHORT SEQUENCE OF 64 SAMPLES OF A SUM OF 2 SINUSOIDS. THE SINUSOIDS ARE GIVEN BY (IV.1)

m	p	N	c <sub>1</sub>	c <sub>3</sub>	m	p	N	r <sub>1</sub>	r <sub>3</sub>
16	4	64	0.52E+00	0.58E-20	16	10	64	0.56E-01	0.70E-03

TABLE VII

AVERAGE INTERPOLATION ERRORS FOR VARIOUS SIGNAL-TO-NOISE RATIOS AND ORDERS OF PREDICTION WITH KNOWN COEFFICIENTS AND AFTER 1 ITERATION FOR 10 REALIZATIONS OF AN AUTOREGRESSIVE PROCESS OF ORDER 10 WITH A PEAKY SPECTRUM. THE COEFFICIENTS OF THE AUTOREGRESSIVE PROCESS ARE GIVEN IN TABLE I

SNR	m	p	N	f	c <sub>1</sub>	r <sub>1</sub>
40dB	16	10	512	0.26E-01	0.27E-01	0.28E-01
40dB	16	20	512		0.28E-01	0.30E-01
20dB	16	10	512	0.42E-01	0.42E-01	0.43E-01
20dB	16	20	512		0.45E-01	0.46E-01

TABLE VIII

INTERPOLATION ERRORS FOR VARIOUS SIGNAL-TO-NOISE RATIOS AND ORDERS OF PREDICTION FOR A SUM OF SINUSOIDS. THE SINUSOIDS ARE GIVEN BY (IV.1)

SNR	m	p	N	c <sub>1</sub>	c <sub>3</sub>	r <sub>1</sub>	r <sub>3</sub>
40dB	16	4	512	0.84E-02	0.40E-03	0.87E-01	0.21E-01
40dB	16	10	512	0.12E-03	0.11E-03	0.12E-03	0.12E-03
20dB	16	4	512	0.38E+00	0.33E+00	0.44E+00	0.40E+00
20dB	16	10	512	0.10E-01	0.11E-01	0.10E-01	0.10E-01

TABLE IX

INTERPOLATION ERRORS AND EXPECTED INTERPOLATION ERRORS OBTAINED WITH THE ALGORITHM FOR THE INTERPOLATION OF BAND-LIMITED SIGNALS FOR VARIOUS SIGNAL-TO-NOISE RATIOS AND BURSTS LENGTHS FOR A SUM OF SINUSOIDS. THE SINE WAVES ARE GIVEN BY (IV.1)

m	SNR	infinite		40dB		20dB	
		b <sub>0,27</sub>	E[ $\bar{e}$ ]	b <sub>0,27</sub>	E[ $\bar{e}$ ]	b <sub>0,27</sub>	E[ $\bar{e}$ ]
4		0.53E-13	0.00E+00	0.32E-01	0.25E-01	0.14E+01	0.25E+01
8		0.85E-06	0.00E+00	0.16E+02	0.84E+02	0.21E+04	0.84E+04
14		0.86E-02	0.00E+00	0.36E+03	0.50E+04	0.15E+05	0.50E+06

TABLE X

AVERAGE INTERPOLATION ERRORS OBTAINED WITH THE ALGORITHM FOR THE INTERPOLATION OF BAND-LIMITED SIGNALS AND AVERAGE INTERPOLATION ERRORS OBTAINED AFTER 1 ITERATION FOR A PREFILTERED FRAGMENT OF BEETHOVEN'S VIOLIN CONCERT FOR VARIOUS BURST LENGTHS

m	trace(H)	b <sub>0,27</sub>	m	p	N	r <sub>1</sub>
4	0.10E+04	0.10E-03	4	14	128	0.67E-02
6	0.89E+05	0.88E-02	6	20	192	0.17E-01

included in Tables I and II. Diagrams of some typical interpolation results are presented in Figs. 3-14, together with the original signals, in which the correct values of the unknown samples have been substituted. In the diagrams, the original signal is marked by a "1," the interpolation result is marked by a "2," the positions of the



**TABLE XI**  
AVERAGE INTERPOLATION ERRORS OBTAINED WITH THE ALGORITHM FOR THE INTERPOLATION OF BAND-LIMITED SIGNALS AND AVERAGE INTERPOLATION ERRORS OBTAINED AFTER 1 ITERATION FOR A NONPREFILTERED FRAGMENT OF BEETHOVEN'S VIOLIN CONCERT FOR VARIOUS BURSTS LENGTHS

$m$	trace(H)	$b_{0.27}$	$m$	$p$	$N$	$r_1$
4	0.10E+04	0.35E-01	4	14	128	0.98E-02
6	0.89E+05	0.21E+01	6	20	192	0.16E-01
			16	50	512	0.19E-01

**TABLE XII**  
AVERAGE INTERPOLATION ERROR OBTAINED WITH THE ADAPTIVE INTERPOLATION ALGORITHM FOR 10 ENGLISH SENTENCES PRONOUNCED BY A MALE AND A FEMALE VOICE

$m$	$p$	$N$	$r_1$
100	50	512	0.56E+00

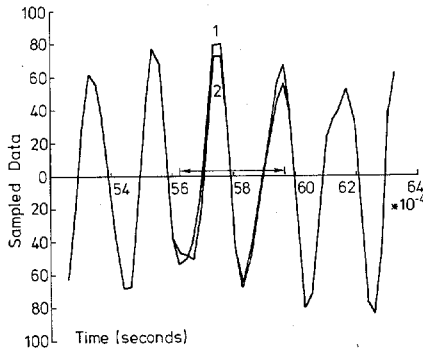


Fig. 3. Interpolation result and original signal for an autoregressive process with a peaky spectrum,  $m = 16$ ,  $p = 10$ ,  $N = 512$ , interpolation method  $c_1$ . Interpolation error  $e = 0.23E - 01$ .

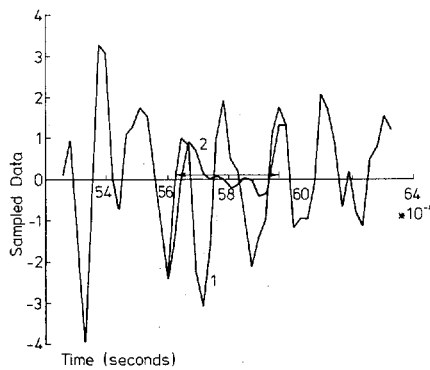


Fig. 4. Interpolation result and original signal for an autoregressive process with a smooth spectrum,  $m = 16$ ,  $p = 10$ ,  $N = 512$ , interpolation method  $c_1$ . Interpolation error  $e = 0.10E + 01$ .

unknown samples are indicated on the time axis. Besides the tables and the diagrams, the performances of the adaptive interpolation method on the music signals and the speech signals are also evaluated by listening tests.

For the test signals 1 and 2, the interpolation results are compared to those obtained by using the true prediction coefficients. In the tables, this method is denoted by  $f$ , where  $f$  stands for fixed coefficients. In the case of one single unknown sample, this nonadaptive interpolation method amounts to the method presented in [2].

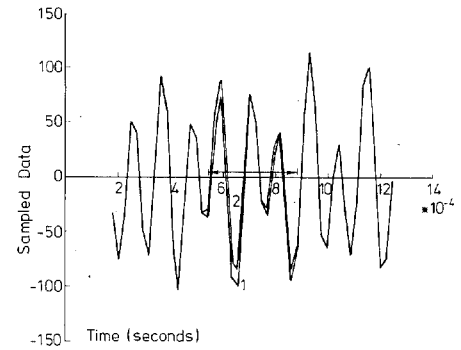


Fig. 5. Interpolation result and original signal for an autoregressive process with a peaky spectrum,  $m = 16$ ,  $p = 10$ ,  $N = 64$ , interpolation method  $c_1$ . Interpolation error  $e = 0.38E - 01$ .

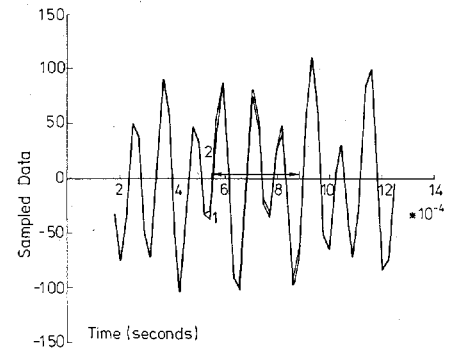


Fig. 6. Interpolation result and original signal for an autoregressive process with a peaky spectrum after 3 iterations,  $m = 16$ ,  $p = 10$ ,  $N = 64$ , interpolation method  $c_3$ . Interpolation error  $e = 0.12E - 01$ .

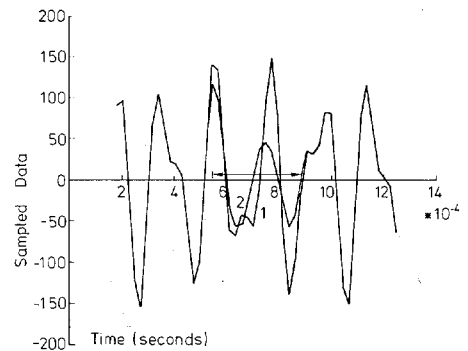


Fig. 7. Interpolation result and original signal for a sum of 2 sine waves,  $m = 16$ ,  $p = 10$ ,  $N = 64$ , interpolation method  $c_1$ . Interpolation error  $e = 0.52E + 00$ .

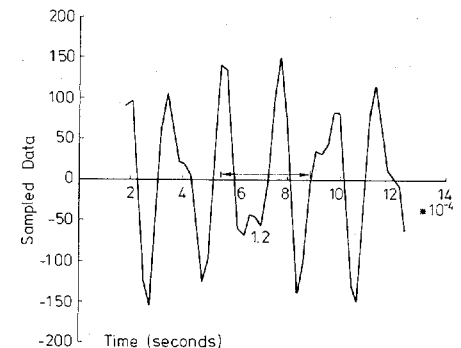


Fig. 8. Interpolation result and original signal for a sum of 2 sinusoids after 3 iterations,  $m = 16$ ,  $p = 10$ ,  $N = 64$ , interpolation method  $c_3$ . Interpolation error  $e = 0.58E - 02$ .

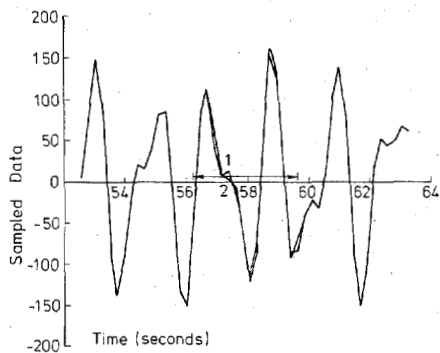


Fig. 9. Interpolation result and original signal for a sum of 2 sine waves, corrupted by white noise, SNR = 20 dB,  $m = 16$ ,  $p = 10$ ,  $N = 512$ , interpolation method  $c_1$ . Interpolation error  $e = 0.10E - 01$ .

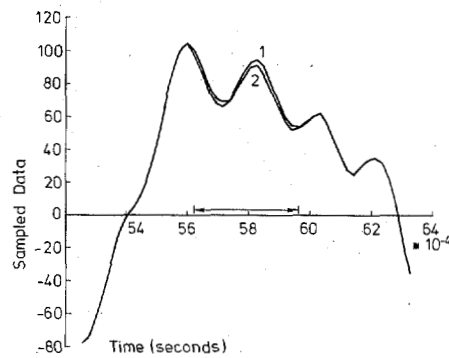


Fig. 13. Interpolation result and original signal for a music signal,  $m = 16$ ,  $p = 50$ ,  $N = 512$ , interpolation method  $r_1$ . Interpolation error  $e = 0.22E - 01$ .

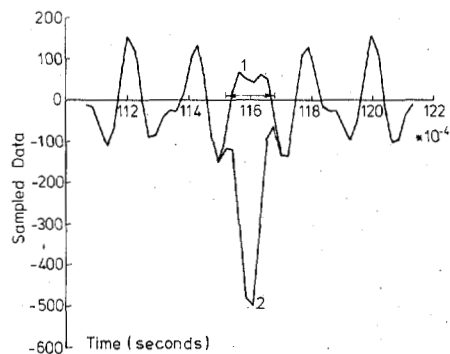


Fig. 10. Interpolation result and original signal of the interpolation method for band-limited signals, for a sum of 2 sinusoids, corrupted by white noise, SNR = 40 dB,  $m = 16$ , interpolation method  $b_{0.27}$ . Interpolation error  $e = 0.16E + 02$ .

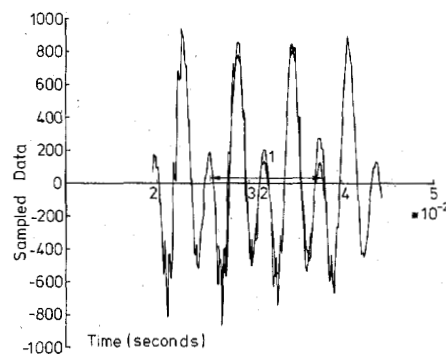


Fig. 14. Interpolation result and original signal for a voiced fragment of mate speech,  $m = 100$ ,  $p = 50$ ,  $N = 512$ , interpolation method  $r_1$ . Interpolation error  $e = 0.70E - 01$ .

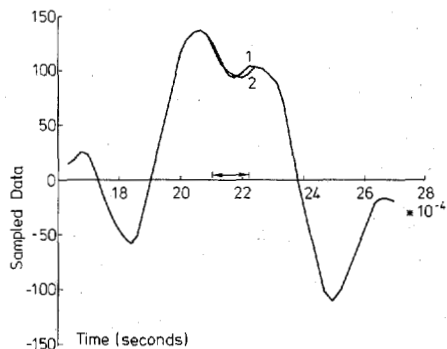


Fig. 11. Interpolation result and original signal for a music signal,  $m = 6$ ,  $p = 20$ ,  $N = 192$ , interpolation method  $r_1$ . Interpolation error  $e = 0.59E - 02$ .

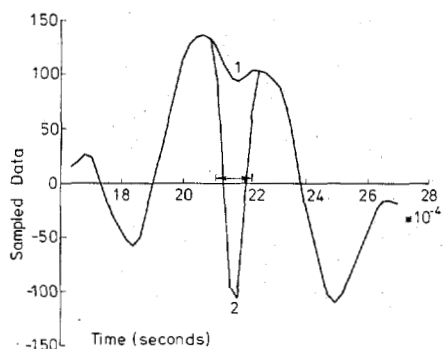


Fig. 12. Interpolation result and original signal of the interpolation method for band-limited signals, for a music signal  $m = 6$ , interpolation method  $b_{0.27}$ . Interpolation error  $e = 0.28E + 01$ .

For the (noisy) sinusoids and the music signals, a comparison is made to an interpolation method for band-limited signals, presented in [3]. For this method, it is required that the signals are band-limited to a baseband extending to  $\alpha$ ,  $0 < \alpha < 1/2$ , times the sample frequency. In the tables, this method is denoted by  $b_\alpha$ . Here  $\alpha = 0.27$ , and the noise-free sinusoids considered are limited to this band. For this comparison also, a prefiltered version of the music signal, limited to the baseband, is used.

It must be remarked that the interpolation method for band-limited signals is rather sensitive to the presence of out-of-band components in the signal. For instance, the expected value of the relative quadratic interpolation error  $\bar{e}$  of (II.19) for a band-limited signal corrupted by white noise is given by

$$E[\bar{e}] = \rho^{-1} \text{trace}(H) \tag{IV.3}$$

where  $\rho$  is the signal-to-noise ratio, and  $H$  is an  $m \times m$ -matrix given by

$$H_{ij} = \delta_{ij} - \frac{\sin(2\pi\alpha(t(i) - t(j)))}{\pi(t(i) - t(j))}, \tag{IV.4}$$

$$i, j = 1, \dots, m.$$

The values of  $E[\bar{e}]$  or, if the signal-to-noise ratio is unknown, of  $\text{trace}(H)$ , are included in the tables. In [3] it is shown that for bursts of unknown samples,  $\text{trace}(H)$  increases roughly as  $\exp(\pi m \alpha)$ , so that even small bursts

in signals with fairly high signal-to-noise ratios cannot be interpolated successfully.

The tables and figures show interpolation errors for various segment lengths and prediction orders. The true prediction orders of the artificially generated autoregressive processes and of the multiple sinusoids are known in advance. For the autoregressive processes, the prediction order is 10; for the multiple sinusoids, the prediction order is twice the number of sinusoids in the signal, in this case  $p = 4$ . For these signals, the true prediction order is used in most cases, a higher prediction order is sometimes tried to achieve improvement in the interpolation quality. For the music and the speech signal,  $p = \min \{3m + 2, 50\}$  is chosen. This rather arbitrary choice gives good interpolation results. The pattern of the unknown samples is always a burst. It has turned out that, as a rule, general patterns of unknown samples are usually interpolated with a smaller interpolation error than bursts with the same number of unknown samples.

The tables and figures give rise to the following remarks. From Tables II and III, it is seen that the interpolation errors for either adaptive method do not differ significantly from the interpolation errors for the interpolation method that uses the true prediction coefficients. It seems that the estimation of the prediction coefficients from the incomplete data does not influence the quality of the interpolation. The deviation of the average interpolation errors from the expected interpolation errors in Tables II and III is explained by the high variance of the interpolation error. It is also seen from the Tables II and III that iterative use of the adaptive interpolation methods does not give a significant improvement. However, if the segment length  $N$  is smaller, iteration does give an improvement, as can be seen from Tables V and VI. Here results close to that of Tables II and III are obtained after 3 iterations. In general, the interpolation errors for autoregressive processes with a peaky spectrum are substantially smaller than for processes with a smooth spectrum.

For sinusoids  $\sigma_e^2 = 0$ , so that, theoretically, the interpolation error is also zero. Indeed, Table IV shows very small interpolation errors for methods  $c_1$  and  $c_3$ . The poorer results for  $r_1$ ,  $r_3$ ,  $p = 4$  can be explained by the fact that the autocorrelation method uses a biased estimate for the autocorrelation function. This has less influence on the result if  $p$  is chosen higher. If the autocovariance method is used to estimate the prediction coefficients,  $p$  must not be chosen too high. For after more than one iteration, the autocovariance matrix will become nearly singular and the prediction coefficients can no longer be calculated straightforwardly by solving the system (II.13). As can be seen from the Tables II, III, VII, and VIII for signals other than multiple sinusoids, there are no significant differences between the interpolation results obtained by using the autocovariance or the autocorrelation method.

With a decreasing signal-to-noise ratio, the interpolation results of the adaptive interpolation methods deteriorate slightly. However, they still do not differ significantly from the results that can be obtained if the true

prediction coefficients were used. This can be seen from Tables VII and VIII. The interpolation method for band-limited signals gives poor results for noisy sinusoids, as can be seen from Table IX, where the quadratic interpolation error becomes several times larger than the signal energy.

For the music signal, the relative quadratic interpolation errors for the adaptive interpolation methods are of the same orders of magnitude as those for the autoregressive processes with a peaky spectrum. The high value for the relative quadratic interpolation error for the speech signal in Table XII can be explained as follows. In popular speech models [18], speech is assumed to consist of voiced parts, where the speech signal is highly periodic, and unvoiced parts, where the speech can be modeled as an autoregressive process of order approximately 12. In the voiced case, the speech spectrum contains many sharp equidistant peaks, and the interpolation results are similar to those obtained with autoregressive signals that have a peaky spectrum. In the unvoiced case, the speech spectrum is rather flat, and the interpolation results are similar to those obtained with autoregressive signals with a smooth spectrum. As can be seen from Table III and Fig. 4, these results are rather poor, especially if the bursts are large. The relative quadratic interpolation error in Table XII is averaged over 20 sentences and the high value is caused by the presence of unvoiced fragments. However, the poor interpolation results for unvoiced fragments do not cause any audible disturbance in the interpolated speech. Fig. 14 shows a typical interpolation result for voiced speech.

Listening tests have revealed that the interpolation errors in these test signals and in many other signals are practically inaudible. After increasing the burst length from 16 to 50, the interpolation results are still quite good for most music signals, although some interpolation errors become audible. For the speech signals, bursts can be restored up to 100 unknown samples without audible errors. It may seem curious that the method still works for bursts of these lengths (which represent time intervals of durations up to 12.5 ms), since the length  $N$  of the segment used to estimate the prediction coefficients represents a time interval of more than 60 ms which is generally too long for a speech signal to be assumed stationary. However, some speech sounds, for instance vowels, can be assumed stationary for several hundreds of milliseconds, and for these the method performs well. Other speech sounds, especially the plosive sounds, /b/, /d/, /g/, /p/, /t/, and /k/, can only be assumed stationary for a few milliseconds and cannot be interpolated correctly. Still, the errors made here do not seem to reduce the subjective interpolation quality, possibly because of masking effects.

Comparing the adaptive interpolation method to the interpolation method for band-limited signals, it is seen that the latter method performs better if the burst length  $m$  is small, say,  $m \leq 6$ , and if no out-of-band components are present in the signal. If these requirements are not met, it gives very poor results, as can be seen from Table XI

and Figs. 10 and 12. Usually, the errors are pulse shaped and may well exceed the peak values of the signal. The adaptive interpolation method performs significantly better in the presence of noise or for large bursts.

V. CONCLUSIONS

In this paper, an adaptive method has been presented for the interpolation of general patterns of unknown samples occurring in discrete-time signals that can be modeled as autoregressive processes. It has been demonstrated that this method gives satisfactory results for digital audio signals and digitized speech. Roughly speaking, the method amounts to trying to minimize, as a function of the unknown samples and the unknown prediction coefficients, a sum of squares of residual errors involving the unknown samples, the prediction coefficients, and the known samples from a sufficiently large neighborhood. The method can be used noniteratively as well as iteratively. In the noniterative case, one minimization step with respect to the prediction coefficients and, subsequently, one with respect to the unknown samples are performed. In the iterative case, the subsequent minimizations are performed repeatedly, the current estimates of the unknown samples being employed in each iteration step. Iteration gives an improvement in interpolation quality if a relatively small segment of data is available.

The interpolation method has been shown to have a sound mathematical foundation. Also, it can be related to several well-known estimation methods in statistics and linear filtering. Furthermore, when applied to signals satisfying the model assumption, the expected quadratic interpolation error per sample is bounded asymptotically by the signal energy.

The method has been tested on artificially generated autoregressive processes, sinusoids, digital audio signals, and digitized speech signals. The performance has been judged both objectively and subjectively. It has been observed that the interpolation method is capable of restoring satisfactorily at least 16 consecutive unknown samples in an audio signal sampled at 44 100 Hz, corresponding to a time interval of 0.36 ms, and in a speech signal, sampled at 8000 Hz up to 100 consecutive unknown samples, corresponding to a time interval of 12.5 ms.

It has been shown that the various minimizations can be carried out by efficiently solving, in a stable manner, certain systems of linear equations. This indicates that the interpolation method is suitable for a fixed point implementation in an integrated circuit. However, in that case, the number of unknown samples should not be too high (up to 16, say).

APPENDIX A  
ANALYSIS OF  $Q(a, x)$

In this appendix, the claims of Section II are proved that, if the sample values have a Gaussian probability density function,

1) minimizing  $Q(a, x)$  for known  $x$  as a function of  $a$  leads to a maximum likelihood estimate for  $a$ , and

2) minimizing  $Q(a, x)$  for known  $a$  as a function of  $x$  leads to a minimum variance estimate for  $a$ .

Furthermore, it is shown that the integral in (II.6) attains its minimum as a function of  $x$  for the same value as  $Q(a, x)$  does.

A. Maximum Likelihood Estimation of  $a$  for Known  $x$

It is assumed that the  $\tilde{e}_k$  are independent and have probability density functions

$$p_{\tilde{e}_k}(e) = \frac{1}{\sigma_e \sqrt{2\pi}} \exp\left(-\frac{e^2}{2\sigma_e^2}\right), \quad k = -\infty, \dots, \infty. \tag{A.1}$$

The log likelihood function that is usually taken to get maximum likelihood estimates for  $\sigma_e^2, a$  from a sequence  $s = [s_0, \dots, s_{N-1}]^T$  of data is  $\log(p_{\tilde{s}}(s|\sigma_e^2, a))$ , the logarithm of the probability density function of  $\tilde{s}$ . The log likelihood function in (II.6) differs slightly from this. However, it can be shown that for large  $N$  (compared to  $p$ ) one may approximate the more commonly used one by the one in (II.6) [19].

To express  $L(\sigma_e^2, a)$  in terms of  $Q(a, x)$ , one observes that

$$\begin{aligned} p_{\tilde{s}|\tilde{u}}(s|\mathbf{u}, \sigma_e^2, a) &= p_{s_p, \dots, s_{N-1}|\tilde{u}}(s_p, \dots, s_{N-1}|\mathbf{u}, \sigma_e^2, a) \\ &= p_{s_{p+1}, \dots, s_{N-1}|s_0, \dots, s_p} \\ &\quad \times (s_{p+1}, \dots, s_{N-1}|s_0, \dots, s_p, \sigma_e^2, a) \\ &\quad \times p_{s_p|\tilde{u}}(s_p|\mathbf{u}, \sigma_e^2, a). \end{aligned} \tag{A.2}$$

Furthermore,

$$\begin{aligned} p_{s_p|\tilde{u}}(s_p|\mathbf{u}, \sigma_e^2, a) &= p_{\tilde{e}_p}\left(\sum_{l=0}^p a_l s_{p-l}\right) \\ &= \frac{1}{\sigma_e \sqrt{2\pi}} \exp\left(-\frac{1}{2\sigma_e^2} \left|\sum_{l=0}^p a_l s_{p-l}\right|^2\right). \end{aligned} \tag{A.3}$$

By repeatedly applying the above reasoning, one finds that

$$\begin{aligned} p_{\tilde{s}|\tilde{u}}(s|\mathbf{u}, \sigma_e^2, a) &= \left(\frac{1}{\sigma_e \sqrt{2\pi}}\right)^{N-p} \\ &\quad \cdot \exp\left(-\frac{1}{\sigma_e^2} Q(a, x)\right). \end{aligned} \tag{A.4}$$

Therefore,

$$L(\sigma_e^2, a) = -(N-p) \log(\sigma_e \sqrt{2\pi}) - Q(a, x)/(2\sigma_e^2). \tag{A.5}$$

Maximizing  $L(\sigma_e^2, a)$  as a function of  $a$  is the same as minimizing  $Q(a, x)$  as a function of  $a$ . This proves the claim. Furthermore,  $\sigma_e^2$  can be estimated by maximizing  $L(\sigma_e^2, a)$  as a function of  $\sigma_e^2$ . This gives the estimate in (II.5) if  $m = 0$  is taken.

B. Minimum Variance Estimation of the Unknown Samples

It is shown that, under hypothesis (A.1), finding the minimum variance estimate for  $x$ , given  $a$  and  $s_k, k = 0,$

$\dots, N-1, k \neq t(1), \dots, t(m)$  is the same as minimizing  $Q(\mathbf{a}, \mathbf{x})$  as a function of  $\mathbf{x}$ . To this end, one can use the well-known fact from statistical estimation theory that the minimum variance estimator  $\hat{\mathbf{w}}_0$  of a stochastic vector  $\mathbf{w}$ , given a stochastic vector  $\mathbf{v}$  equals  $E[\mathbf{w}|\mathbf{v}]$ , the expectation under condition  $\mathbf{v}$ . Hence, for the minimum variance estimator  $\hat{\mathbf{x}}_0$  of  $\mathbf{x}$ , given the known samples one has

$$\hat{\mathbf{x}}_0 = E[\mathbf{x}|\mathbf{q}] \quad (\text{A.6})$$

where the known samples are arranged in a vector  $\mathbf{q}$ . To evaluate this, one needs  $p_{\mathbf{x}|\mathbf{q}}(\mathbf{x}|\mathbf{q})$ . It is straightforward to show that

$$p_{\mathbf{x}|\mathbf{q}}(\mathbf{x}|\mathbf{q}) = p_{\mathbf{s}|\mathbf{u}}(\mathbf{s}|\mathbf{u})/p_{\mathbf{q}|\mathbf{u}}(\mathbf{q}|\mathbf{u}). \quad (\text{A.7})$$

By using (A.4), one can express the right-hand side of (A.7) in terms of  $Q(\mathbf{a}, \mathbf{x})$ . More specifically, one has for  $\hat{\mathbf{x}}_0$

$$\hat{\mathbf{x}}_0 = D \int_{\mathcal{R}^m} \mathbf{x} \exp(-Q(\mathbf{a}, \mathbf{x})/(2\sigma_e^2)) d\mathbf{x}, \quad (\text{A.8})$$

where  $D$  is such that

$$D \int_{\mathcal{R}^m} \mathbf{x} \exp(-Q(\mathbf{a}, \mathbf{x})/(2\sigma_e^2)) d\mathbf{x} = 1. \quad (\text{A.9})$$

It follows from a standard fact about Gaussian integrals that  $Q(\mathbf{a}, \mathbf{x})$ , a quadratic form in  $\mathbf{x}$ , is minimized by  $\hat{\mathbf{x}}_0$  in (A.8). This proves the claim.

Observe also that  $\hat{\mathbf{x}}_0$  maximizes  $p_{\mathbf{x}|\mathbf{q}}(\mathbf{x}|\mathbf{q})$  as a function of  $\mathbf{x}$ . This shows that  $\hat{\mathbf{x}}_0$  is also a maximum *a posteriori* estimates for  $\mathbf{x}$ . This follows further from (A.7) and (A.4).

A further result is that  $\hat{\mathbf{x}}_0 = -(B(\mathbf{a}))^{-1} \mathbf{z}(\mathbf{a})$  is also the best linear minimum variance estimator. This result holds without any assumptions on the form of the probability density function of the excitation noise. Indeed, it is a well-known fact that the best linear minimum variance estimator is completely determined by the covariances  $E[\tilde{s}_j \tilde{s}_k]$ ,  $j, k = 0, \dots, N-1$ . Furthermore, under the hypothesis of Gaussian probability density functions, the minimum variance estimator  $\hat{\mathbf{x}}_0$  happens to depend linearly on the known samples, as is seen from (II.14), (II.11), and (II.12). Thus,  $\hat{\mathbf{x}}_0$  is certainly the best linear minimum variance estimator, whether or not Gaussian probability density functions have been assumed.

### C. A Spectral Interpretation

It is shown that for fixed  $\mathbf{a}$ , the integral (II.16) attains its minimum for the same  $\mathbf{x}$  as  $Q(\mathbf{a}, \mathbf{x})$  does. It is useful to note that

$$Q(\mathbf{a}, \mathbf{x}) = \sum_{k,l=p}^{N-p-1} s_k s_l b_{k-l} + E \quad (\text{A.10})$$

where  $E$  involves only samples with indexes  $l, k < p$  or  $l, k > N-p-1$ . The integral (II.16) can be written as

$$\frac{1}{2\pi\sigma_e^2} \int_{-\pi}^{\pi} \sum_{l=-p}^p b_l \exp(-j\theta l) \hat{S}(x; \theta) d\theta. \quad (\text{A.11})$$

On inserting the definition (II.17) into (A.11) and performing the integration, one finds for the integral in (II.16)

$$\frac{1}{N-2p} \frac{1}{\sigma_e^2} \sum_{k,l=p}^{N-p-1} s_k s_l b_{k-l} = aQ(\hat{\mathbf{a}}, \mathbf{x}) + b \quad (\text{A.12})$$

for properly chosen constants  $a > 0, b \in \mathbb{R}$ . This proves that the  $\hat{\mathbf{x}}$  of (II.14) minimizes the integral in (II.16) as a function of  $\hat{\mathbf{x}}$ .

## APPENDIX B

### SOME NOTES ON THE ITERATIVE VERSION OF THE INTERPOLATION METHOD

#### A. Convergence Properties

Iterating the interpolation method comes down to constructing two sequences  $\hat{\mathbf{a}}(k) \in \mathbb{R}^p$  and  $\hat{\mathbf{x}}(k) \in \mathbb{R}^m$  of vectors of prediction coefficients and sample estimates, respectively. Here  $\hat{\mathbf{a}}(1) = \hat{\mathbf{a}}$  and  $\hat{\mathbf{x}}(1) = \hat{\mathbf{x}}$ ,  $\hat{\mathbf{a}}$  and  $\hat{\mathbf{x}}$  as in (II.13) and (II.14). In the  $k$ th step,  $\hat{\mathbf{a}}(k)$  and  $\hat{\mathbf{x}}(k)$  are obtained by minimizing  $Q(\mathbf{a}, \hat{\mathbf{x}}(k-1))$  with respect to  $\mathbf{a}$  and  $Q(\hat{\mathbf{a}}(k), \mathbf{x})$  with respect to  $\mathbf{x}$ , respectively. That is,

$$Q(\hat{\mathbf{a}}(k), \hat{\mathbf{x}}(k-1)) = \min_{\mathbf{a} \in \mathbb{R}^p} Q(\mathbf{a}, \hat{\mathbf{x}}(k-1)) \quad (\text{B.1})$$

$$Q(\hat{\mathbf{a}}(k), \hat{\mathbf{x}}(k)) = \min_{\mathbf{x} \in \mathbb{R}^m} Q(\hat{\mathbf{a}}(k), \mathbf{x}). \quad (\text{B.2})$$

It was found that iterating the interpolation method can improve the results if the number of available samples is relatively small. Although the iterative method turns out to converge very rapidly in practice (cf. Section IV for more details), it does not seem easy to prove satisfactory convergence results. It can be shown that, when the sequences  $\hat{\mathbf{a}}(k), \hat{\mathbf{x}}(k)$  converge, the limit point  $\mathbf{a}', \mathbf{x}'$  is a stationary point. However,  $Q(\mathbf{a}, \mathbf{x})$  may have several such points. For the asymptotic speed of convergence, the Hessian  $H'$  of  $Q(\mathbf{a}, \mathbf{x})$  at  $\mathbf{a}', \mathbf{x}'$  is relevant. Letting

$$H' = \begin{vmatrix} A_1 & A_2 \\ A_2^T & A_3 \end{vmatrix} \quad (\text{B.3})$$

where  $A_1 = (\partial^2/\partial \mathbf{a}^2) Q(\mathbf{a}', \mathbf{x}')$ ,  $A_2 = (\partial^2/\partial \mathbf{a} \partial \mathbf{x}) Q(\mathbf{a}', \mathbf{x}')$  and  $A_3 = (\partial^2/\partial \mathbf{x}^2) Q(\mathbf{a}', \mathbf{x}')$ , one can check that

$$\begin{aligned} A_1 &= 2C(\mathbf{x}'), \\ A_2 &= 2(e_{i+t(j)} + f_{-i+t(j)})_{i=1, \dots, p, j=1, \dots, m}, \\ A_3 &= 2B(\mathbf{a}'), \end{aligned} \quad (\text{B.4})$$

where, with  $s_{t(i)} = x'_i, i = 1, \dots, m$ ,

$$\begin{aligned} e_k &= \sum_{l=0}^p a'_l s_{k-l} \\ f_k &= \sum_{l=0}^p a'_l s_{k+l}. \end{aligned} \quad (\text{B.5})$$

Since

$$\begin{aligned} Q(\mathbf{a}, \mathbf{x}) &\cong Q(\mathbf{a}', \mathbf{x}') + 1/2(\mathbf{a} - \mathbf{a}')^T A_1 (\mathbf{a} - \mathbf{a}') \\ &\quad + (\mathbf{a} - \mathbf{a}')^T A_2 (\mathbf{x} - \mathbf{x}') \\ &\quad + 1/2(\mathbf{x} - \mathbf{x}')^T A_3 (\mathbf{x} - \mathbf{x}'), \end{aligned} \quad (\text{B.6})$$

it follows that

$$\hat{x}(k + 1) \cong x' + A_3^{-1}A_2^T A_1^{-1}A_2(\hat{x}(k) - x'). \quad (\text{B.7})$$

The speed of convergence of the iterative method is determined by the spectral properties of  $D = A_3^{-1}A_2^T A_1^{-1}A_2$ . A condition guaranteeing linear convergence is that the absolute values of the eigenvalues of  $D^T D$  are all less than 1, but it does not seem easy to check on this condition.

### B. Relation to the EM Algorithm

The assumption is that the observed  $s = [s_0, \dots, s_{N-1}]^T$  is a vector of realizations of a stationary Gaussian autoregressive process of known order  $p$  and unknown prediction coefficients  $a \in \mathbb{R}^p$  and  $\sigma_e^2$ . Consider the log likelihood function

$$L(\sigma_e^2, a) = -(N - p) \cdot \log(\sigma_e \sqrt{2\pi}) - Q(a, x)/(2\sigma_e^2). \quad (\text{B.9})$$

The EM algorithm aims at finding estimates for parameters and complete data from incomplete data by maximizing the (log) likelihood function. It can be described for the present situation as follows. Denote for  $(\sigma_e^2, a) \in \mathbb{R} \times \mathbb{R}^p$

$$W(\sigma_e^2, a^* | \sigma_e^2, a) = E[L(\sigma_e^2, a^*) | q, \sigma_e^2, a], \quad (\text{B.10})$$

with  $q$  as in (A.6). Starting with initial estimates  $\hat{\sigma}_e^2(0)$ ,  $\hat{a}(0)$ , one constructs sequences  $\hat{\sigma}_e^2(k)$ ,  $\hat{a}(k)$ ,  $k = 1, 2, \dots$ , by choosing in the  $k$ th step  $\hat{\sigma}_e^2(k)$ ,  $\hat{a}(k)$  in such a way that  $W(\sigma_e^2, a | \hat{\sigma}_e^2(k-1), \hat{a}(k-1))$  is maximal at  $(\sigma_e^2, a) = (\hat{\sigma}_e^2(k), \hat{a}(k))$ . Heuristically, one would like to maximize  $L(\sigma_e^2, a)$ , but this is impossible since one does not know  $s$  completely.

To show the connection with the iterative interpolation method, it is necessary to evaluate (B.10). The conditional expectation in (B.10) refers to the conditional probability density

$$p_{\bar{x}|q}(x|q) = \frac{|B(a)|^{1/2}}{(2\pi)^{m/2} \sigma_e^m} \exp\left(-\frac{(x - \bar{x})^T B(a) (x - \bar{x})}{2\sigma_e^2}\right). \quad (\text{B.11})$$

It follows that

$$W(\sigma_e^2, a^* | \sigma_e^2, a) = -(N - p) \log(\sigma_e^* \sqrt{2\pi}) - E[Q(a^*, x) | q, \sigma_e^2, a] / (2\sigma_e^{*2}). \quad (\text{B.12})$$

It is a tedious but straightforward calculation to show that

$$E[Q(a^*, x) | q, \sigma_e^2, a] = \sigma_e^2 \text{trace}((B(a))^{-1} B(a^*)) + Q(a^*, x), \quad (\text{B.13})$$

so that one has

$$W(\sigma_e^{*2}, a^* | \sigma_e^2, a) = -(N - p) \log(\sigma_e^* \sqrt{2\pi}) - 1/2(\sigma_e/\sigma_e^*)^2 \text{trace}((B(a))^{-1} B(a^*)) - Q(a^*, x). \quad (\text{B.14})$$

Maximizing  $W(\sigma_e^{*2}, a^* | \sigma_e^2, a)$  is the same as minimizing the right-hand side of (B.13). Hence, the difference between the EM algorithm and the iterative version of the adaptive interpolation method is reflected by the first term in the right-hand side of (B.13). It is noted that minimizing  $Q(a^*, x)$  with respect to  $a^*$  is much easier than minimizing the right-hand side of (B.13), since it is not likely that a manageable form for the solution of the latter problem exists.

### APPENDIX C DECOMPOSITION OF $B(a)$

In this appendix, the left-hand inequality of (III.4) is proved. First, remark that

$$B(a) = A^T A, \quad (\text{C.1})$$

where  $A = [a_1, \dots, a_m]$  is a  $(t(m) - t(1) + p + 1) \times m$ -matrix, defined by

$$A_{ij} = (a_j)_i = a_{t(j)-i-t(1)+p+1} \quad (\text{C.2})$$

the  $a_i$  being the prediction coefficients of (II.1). Note that  $a_i = 0$  for  $i < 0$  or  $i > p$ . Since  $A$  has full rank,  $A$  can be decomposed as a product  $A = QR$ , where  $Q$  is a  $(t(m) - t(1) + p + 1) \times m$ -matrix, consisting of  $m$  orthogonal columns and  $R$  is an upper triangular  $m \times m$ -matrix. On substituting  $A = QR$  into (C.1), one obtains

$$B(a) = R^T Q^T Q R = \tilde{L} D \tilde{L}^T \quad (\text{C.3})$$

where  $\tilde{L}$  and  $D$  are as in (III.3). Clearly,  $D_{jj} = |q_j|^2$ . The  $QR$  decomposition of  $A$  can be done iteratively. In every iteration step,  $q_j$  is found by subtracting from  $a_j$  the projection of  $a_j$  onto the space spanned by  $q_1, \dots, q_{j-1}$

$$q_j = a_j - \sum_{k=1}^{j-1} \frac{a_j^T q_k}{|q_k|^2} q_k \quad (\text{C.4})$$

The space  $\text{sp}\{q_1, \dots, q_{j-1}\}$  spanned by  $q_1, \dots, q_{j-1}$  is the same as the space  $\text{sp}\{a_1, \dots, a_{j-1}\}$  spanned by  $a_1, \dots, a_{j-1}$ . Therefore,

$$\begin{aligned} |q_j|^2 &= \min_{v \in \text{sp}\{q_1, \dots, q_{j-1}\}} |a_j - v|^2 \\ &= \min_{v \in \text{sp}\{a_1, \dots, a_{j-1}\}} |a_j - v|^2 \\ &= \min_{w \in \mathbb{R}^{j-1}} \left| a_j + \sum_{k=1}^{j-1} w_k a_k \right|^2. \end{aligned} \quad (\text{C.5})$$

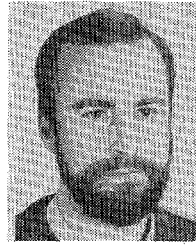
Since  $(a_j)_{t(j)-t(1)+p+1} = a_0 = 1$  and  $(a_k)_{t(j)-t(1)+p+1} = 0$  for  $k = 1, \dots, j$ , by (C.2), it follows easily that  $|q_j|^2 \geq 1$ . This proves the left-hand inequality of (III.4).

### ACKNOWLEDGMENT

The authors wish to thank their colleague H. J. Prins for his help in finding a convenient approach to the solution of the interpolation problem, and for drawing their attention to the EM algorithm.

## REFERENCES

- [1] R. Steele and F. Benjamin, "Sample reduction and subsequent adaptive interpolation of speech signals," *Bell. Syst. Tech. J.*, vol. 62, no. 6, pp. 1365-1398, 1983.
- [2] S. M. Kay, "Some results in linear interpolation theory," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-31, pp. 746-749, June 1983.
- [3] A. J. E. M. Janssen and L. B. Vries, "Interpolation of band-limited discrete-time signals by minimizing out-of-band energy," in *Proc. ICASSP '84*, San Diego, CA, 1984.
- [4] H. Akaike, "A new look at the statistical model identification," *IEEE Trans. Automat. Contr.*, vol. AC-19, pp. 716-728, June 1974.
- [5] S. M. Kay and S. L. Marple, Jr., "Spectrum analysis—A modern perspective," *Proc. IEEE*, vol. 69, no. 11, pp. 1380-1419, 1981.
- [6] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. Roy. Stat. Soc., Series B*, vol. 39, pp. 1-38, 1977.
- [7] C. F. J. Wu, "On the convergence properties of the EM algorithm," *Ann. Stat.*, vol. 11, no. 1, pp. 95-103, 1983.
- [8] R. A. Bayles, "On the convergence of the EM algorithm," *J. Roy. Stat. Soc., Series B*, vol. 45, no. 1, pp. 47-50, 1983.
- [9] F. Itakura and S. Saito, "A statistical method for estimation of speech spectral density and formant frequencies," *Electron. Commun. Japan*, vol. 53-A, pp. 36-43, 1970.
- [10] I. I. Hirschmann, Jr., "Recent developments in the theory of finite Toeplitz operators," in *Advances in Probability and Related Topics*, vol. 1. New York: Marcel Dekker, 1971.
- [11] M. Marcus and H. Minc, *Introduction to Linear Algebra*. New York: MacMillan, 1965.
- [12] J. Durbin, "The fitting of time-series models," *Rev. Inst. Int. Stat.*, vol. 28, pp. 233-243, 1960.
- [13] G. Cybenko, "The numerical stability of the Levinson-Durbin algorithm for Toeplitz systems of equations," *SIAM J. Sci. Stat. Comput.*, vol. 1, pp. 303-320, 1980.
- [14] J. H. Wilkinson, *The Algebraic Eigenvalue Problem*. Oxford, England: Clarendon, 1965.
- [15] N. Levinson, "The Wiener rms (root mean square) error criterion in filter design and prediction," *J. Math Phys.*, vol. 25, pp. 261-278, 1947.
- [16] P. Delsarte, Y. Genin, and Y. Kamp, "A polynomial approach to the generalized Levinson algorithm, based on the Toeplitz distance," *IEEE Trans. Inform. Theory*, vol. IT-29, pp. 268-278, 1983.
- [17] J. H. Wilkinson, "Error analysis of direct methods of matrix inversion," *J. Assoc. Comput. Mach.*, vol. 8, pp. 281-330, 1961.
- [18] L. R. Rabiner and R. W. Schafer, *Digital Processing of Speech Signals*. Englewood Cliffs, NJ: Prentice-Hall, 1978.
- [19] S. M. Kay, "Recursive maximum likelihood estimation of autoregressive processes," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-31, pp. 56-65, Feb. 1983.



**A. J. E. M. Janssen** was born in Breda, The Netherlands, in 1953. He received the Engineer degree and the Ph.D. degree in mathematics in 1976 and 1979, respectively, both from the Eindhoven University of Technology, Eindhoven, The Netherlands.

From 1979 to 1981 he was with the California Institute of Technology, Pasadena, as a Bateman Research Instructor in Mathematics. He is presently with the Philips Research Laboratories, Eindhoven, The Netherlands. His interests are time-frequency descriptions of signals with emphasis on fundamental aspects, spectral estimation, and applications of mathematical analysis in signal theory.



**Raymond N. J. Veldhuis** was born in The Hague, The Netherlands, on April 8, 1955. He received the Ingenieur degree in electronics from the Twente University of Technology in Enschede, The Netherlands, in 1981.

In 1982 he joined Philips Laboratories, Eindhoven, The Netherlands. He has been engaged in various fields of digital signal processing, including the processing of digital audio and video signals.



**Lodewijk B. Vries** was born in The Hague, The Netherlands, on November 18, 1947.

He joined Philips Research Laboratories in Eindhoven, The Netherlands, in 1974, after his graduation in electrical engineering from the Delft University of Technology, Delft, The Netherlands. Since 1976 he has been involved with development of an error-correcting system for the Compact Disc.<sup>®</sup> After the standardization of the CD in 1980, his field of interest changed to digital signal processing with applications to high-fidelity audio. Since March 1983 he has been working at the Advanced Projects Laboratory of the Consumer Electronics Department, Nederlandse Philips Bedrijven B.V.