

ADAPTIVE RESTORATION OF UNKNOWN SAMPLES IN CERTAIN TIME-DISCRETE  
SIGNALS

R.N.J. Veldhuis\*, A.J.E.M. Janssen\* and L.B. Vries\*\*

\*Philips Research Laboratories, Eindhoven, The Netherlands

\*\*Philips Audio Department, Eindhoven, The Netherlands

ABSTRACT

In this paper algorithms for the restoration of unknown samples at known positions embedded in a neighbourhood of known samples are discussed. First this restoration problem is treated as a (non-adaptive) linear minimum variance estimation problem. It is shown that the optimal linear minimum variance interpolator for unknown samples from an autoregressive process uses only a finite neighbourhood of known samples, whereas in general this neighbourhood is infinite. Secondly, for signals that can be modelled as autoregressive processes an adaptive solution to the restoration problem is given.

INTRODUCTION

This paper discusses a method for the restoration, or interpolation, of lost samples that occur at known positions in a time-discrete signal. It can, for instance, be applied in the field of digital audio (e.g. Compact Disc<sup>®</sup>) where the samples are sometimes provided with flags indicating their reliability. Lost samples may occur consecutively (bursts), as well as non-consecutively. The only requirement is that they are embedded in a sufficiently large neighbourhood of known samples. The method discussed in this paper can restore, without audible distortion, up to 32 unknown samples at a time in a music signal, sampled at 44100 Hz. When applied to speech, sampled at 8000 Hz, it can restore up to 100 unknown samples. The organization of this paper is as follows. First the restoration problem is discussed as a minimum variance estimation problem. The

estimates are weighted sums of the known samples, the weight coefficients depending on the autocorrelation function of the signal. The results are derived for finite sequences of samples and generalized to infinite sequences. It follows that for autoregressive processes the best linear minimum variance estimates use only a finite number of samples. In this approach, however, it is required that the process parameters (order, prediction coefficients) be known. This is not always the case. Therefore, an adaptive version of the interpolation method is discussed in which both the parameters and the unknown samples are estimated. This is done by iteratively minimizing a function that involves quadratically the unknown samples as well as the parameters, with respect to the parameters and subsequently with respect to the unknown samples. It often turns out that only one iteration step already produces good interpolation results. The paper is concluded with the presentation of some results.

A LINEAR MINIMUM VARIANCE ESTIMATE

Let  $\underline{s} = [s_1, \dots, s_N]^T$  be a finite vector of samples taken from a stationary time-discrete stochastic signal that has zero mean. The superscript  $T$  denotes transposition. Assume that  $s_{t(1)}, \dots, s_{t(m)}$  are unknown. Here  $1 < t(1) < \dots < t(m) < N$  are known. The unknown samples are estimated by

$$(1) \quad \hat{s}_{t(i)} = \sum_{\substack{j=1 \\ j \neq t(1), \dots, t(m)}}^N H_{ij} s_j, \quad i=1, \dots, m.$$

The coefficients  $H_{ij}$  from (1) are chosen such that they minimize the total expected square interpolation error

$$(2) \quad W(t(1), \dots, t(m)) = E \left[ \sum_{i=1}^m (s_{t(i)} - \hat{s}_{t(i)})^2 \right].$$

The  $H_{ij}$  that minimize  $W(t(1), \dots, t(m))$  are solutions of

$$(3) \quad \sum_{j=1, j \neq t(1), \dots, t(m)}^N H_{ij} r(k-j) - r(t(i)-k) = 0,$$

where  $i=1, \dots, m$ ,  $k=1, \dots, N$ ,  $k \neq t(1), \dots, t(m)$  and  $r(j) = E[s_k s_{k+j}]$  is the  $j^{\text{th}}$  autocorrelation coefficient of  $\underline{s}$ .

Expression (3) can be brought into a more convenient form as follows. Let  $H = (H_{ij})_{i=1, \dots, m, j=1, \dots, N}$ , where  $(H_{it(j)})_{i,j=1, \dots, m} = -I$ . Furthermore, let  $R = (R_{ij})_{i,j=1, \dots, N} = (r(i-j))_{i,j=1, \dots, N}$  and let  $R'$  be the  $(N-m) \times N$ -submatrix of  $R$  obtained by deleting the rows of  $R$  with indices  $t(1), \dots, t(m)$ . Then (3) takes the form

$$(4) \quad (a) \quad R'H^T = [0], \quad (b) \quad (H_{it(j)})_{i,j=1, \dots, m} = -I,$$

[0] being the all-zero matrix.

The  $m \times N$ -matrix  $G$  that satisfies

$$(5) \quad RG^T = [\underline{u}_{t(1)}, \dots, \underline{u}_{t(m)}],$$

where  $\underline{u}_{t(i)}$  is the  $t(i)^{\text{th}}$  unit vector of length  $N$ , can be used to find the solution  $H$  of (4). Indeed, let

$$(6) \quad G' = (G_{it(j)})_{i,j=1, \dots, m},$$

Then it can be shown that  $H$ , defined by

$$(7) \quad H = -G'^{-1}G$$

solves (4) if  $G'$  is non-singular. Now  $G'$  is non-singular if  $R$  is non-singular and one sees that consequently

$$(8) \quad G' = ((R^{-1})_{t(i)t(j)})_{i,j=1, \dots, m},$$

and that  $G'$  is positive definite.

If the  $N$ -vector  $\underline{v}$  is obtained from  $\underline{s}$  by substituting zeros for the unknown samples and if

$$\hat{\underline{x}} = [\hat{s}_{t(1)}, \dots, \hat{s}_{t(m)}]^T, \text{ then}$$

$$(9) \quad \hat{\underline{x}} = H\underline{v} = -G'^{-1}G\underline{v}.$$

Sometimes it is more convenient to solve  $\hat{\underline{x}}$  from

$$(10) \quad G'\hat{\underline{x}} = -G\underline{v} =: \underline{z}.$$

The vector  $\underline{z}$  is referred to as the syndrome.

A further observation is that one can derive an expression for the error covariance matrix

$C := E[(\hat{\underline{x}} - \underline{x})(\hat{\underline{x}} - \underline{x})^T]$ , for it follows from (9) and (4) that

$$(11) \quad \hat{\underline{x}} = H\underline{s} + \underline{x},$$

with  $\underline{x} = [s_{t(1)}, \dots, s_{t(m)}]^T$ . Therefore,

$$(12) \quad C = E[H\underline{s}\underline{s}^T H^T] = HRH^T = G'^{-T} = G'^{-1}.$$

Also,

$$(13) \quad W(t(1), \dots, t(m)) = \text{trace}(C) = \text{trace}(G'^{-1}).$$

Until now the number of samples used to estimate the unknown samples was finite. If  $R, H$  and  $G$  are allowed to be (semi-)infinite matrices the previously obtained results also apply for the estimation of unknown samples in an infinite sequence. In that case there are some additional results. If  $G^T = [g_1, \dots, g_m]$  then

$$(14) \quad (g_i)_j = g_{j-t(i)}, \quad i=1, \dots, m, j=-\infty, \dots, \infty,$$

with the rows of  $G$  being shifted versions of an infinite vector  $\underline{g}$ , defined by

$$(15) \quad g_k = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{1}{S(\theta)} \exp(j\theta k) d\theta, \quad k=-\infty, \dots, \infty,$$

where  $S(\theta) = \sum_{k=-\infty}^{\infty} r(k) \exp(-j\theta k)$  is the spectrum of the signal  $(s_k)_{k=-\infty, \dots, \infty}$ . Of course it is required that  $S(\theta) > 0$  for  $|\theta| \leq \pi$ . A possible solution to the interpolation problem for a case where this is not true is given in [1].

For  $G'$  one has

$$(16) \quad G'_{ij} = g_{t(i)-t(j)}, \quad i, j=1, \dots, m,$$

and for the  $\underline{z}$  of (10)

$$(17) \quad z_i = \sum_{k=-\infty}^{\infty} g_{t(i)-k} v_k, \quad i, j=1, \dots, m.$$

Observe that in the case of consecutive errors  $G'$  is Toeplitz and (10) can efficiently be solved in  $O(m^2)$  operations by using the Levinson algorithm.

In general  $\underline{g}$  has infinite length. Therefore, in practical applications a finite approximation must be used to calculate  $\underline{z}$  in (17). However, if

$(s_k)_{k=-\infty, \dots, \infty}$  is an autoregressive process of order  $p$ , then  $\underline{g}$  has length  $2p+1$ . This can easily be shown as follows. For an autoregressive process  $(s_k)_{k=-\infty, \dots, \infty}$  of order  $p$  and with prediction coefficients  $a_0, a_1, \dots, a_p$ ,  $a_0=1$ , one has

## 27.1.2

$$(18) \quad \sum_{j=0}^p a_j s_{k-j} = e_k, \quad k=-\infty, \dots, \infty,$$

where  $(e_k)_{k=-\infty, \dots, \infty}$  is a white noise process with zero mean and variance  $\sigma_e^2$ . The signal spectrum  $S(\theta)$  is given by

$$(19) \quad S(\theta) = \sigma_e^2 \left( \sum_{k=-p}^p b_k \exp(-j\theta k) \right)^{-1},$$

where

$$(20) \quad b_k = \sum_{j=0}^{p-|k|} a_j a_{j+|k|}, \quad k=-p, \dots, p.$$

On substituting (19) into (15) one has for  $\underline{g}$

$$(21) \quad g_k = \begin{cases} \sigma_e^{-2} b_k, & |k| \leq p, \\ 0, & \text{otherwise.} \end{cases}$$

Also,  $G = \sigma_e^{-2} B$ , with

$$(22) \quad B_{ij} = \begin{cases} b_{t(i)-t(j)}, & |t(i)-t(j)| \leq p, \\ 0, & \text{otherwise.} \end{cases}$$

If, in this case, one defines the syndrome  $\underline{z}$  by

$$(23) \quad z_i = - \sum_{k=-p}^p b_k v_{t(i)-k},$$

then  $\hat{\underline{x}}$  can be obtained as the solution of

$$(24) \quad B\hat{\underline{x}} = \underline{z}.$$

Part of the results of this section can be found in [2,3].

#### ADAPTIVE INTERPOLATION OF AUTOREGRESSIVE PROCESSES

In this section the interpolation of autoregressive processes is discussed in the case that the parameters  $p, a_1, \dots, a_p$  as well as the unknown samples  $s_{t(1)}, \dots, s_{t(m)}$  have to be estimated from the data  $s_1, \dots, s_N$ . It is assumed that  $t(1) \geq p+1$  and that  $t(m) \leq N-p$ . This problem is discussed in great detail in [4].

Although several algorithms exist for the estimation of  $p$  [5], the rather arbitrary choice  $p=3m+2$  is used instead. It produces satisfactory interpolation results for the experiments done with digitized music and speech. If  $\underline{a} = [a_1, \dots, a_p]^T$  and  $\underline{x} = [s_{t(1)}, \dots, s_{t(m)}]^T$ , then

the estimates  $\hat{\underline{a}}$  and  $\hat{\underline{x}}$  are those values of  $\underline{a}$  and  $\underline{x}$  that minimize

$$(25) \quad Q(\underline{a}, \underline{x}) = \sum_{k=p+1}^N \left| \sum_{j=0}^p a_j s_{k-j} \right|^2 = \sum_{k=p+1}^N |e_k|^2,$$

with  $s_{t(i)} = x_i$ , for  $i=1, \dots, m$ . This choice can be motivated by the facts that a) minimizing  $Q(\underline{a}, \underline{x})$  as a function of  $\underline{x}$  under the assumption that  $\underline{a}$  is known leads to an estimate  $\hat{\underline{x}}$  solving (24) and b) minimizing  $Q(\underline{a}, \underline{x})$  as a function of  $\underline{a}$  under the assumption that  $\underline{x}$  is known and that  $(s_k)_{k=-\infty, \dots, \infty}$  has a Gaussian distribution is the same as maximizing the likelihood function

$$(26) \quad L(\sigma_e^2, \underline{a}) = \prod_{s_{p+1}, \dots, s_N} L(s_{p+1}, \dots, s_N | s_1, \dots, s_p, \sigma_e^2, \underline{a}) \\ = \left( \frac{1}{2\pi\sigma_e^2} \right)^{(N-p)/2} \exp\left( - \frac{Q(\underline{a}, \underline{x})}{2\sigma_e^2} \right)$$

as a function of  $\underline{a}$ . Note that in  $L(\sigma_e^2, \underline{a})$  a conditional probability density function is used instead of the commonly used unconditional probability density function. Maximizing  $L(\sigma_e^2, \underline{a})$  as a function of  $\underline{a}$  is the same as minimizing  $Q(\underline{a}, \underline{x})$  as a function of  $\underline{a}$ .

Since  $Q(\underline{a}, \underline{x})$  contains 4<sup>th</sup> order terms, this minimization is a non-trivial problem. One can try the following iterative procedure. Starting with a zeroth estimate  $\hat{\underline{x}}^{(0)}$ , ( $\hat{\underline{x}}^{(0)} = \underline{0}$ , for instance), one produces a first estimate  $\hat{\underline{a}}^{(1)}$  for  $\underline{a}$  by minimizing  $Q(\hat{\underline{a}}^{(1)}, \hat{\underline{x}}^{(0)})$  as a function of  $\underline{a}$ . Then, by minimizing  $Q(\hat{\underline{a}}^{(1)}, \underline{x})$  as a function of  $\underline{x}$ , one produces a first estimate  $\hat{\underline{x}}^{(1)}$  for  $\underline{x}$ . This can be repeated to obtain second estimates  $\hat{\underline{a}}^{(2)}$  and  $\hat{\underline{x}}^{(2)}$  and so on. It is clear that  $Q(\hat{\underline{a}}, \hat{\underline{x}})$  decreases to some non-negative number but it seems hard to determine whether this number is a global minimum or not.

The result of the minimization of  $Q(\underline{a}, \underline{x})$  as a function of  $\underline{x}$  with known  $\underline{a}$  was given in the previous section. An additional remark is that solving (24) by means of a Cholesky decomposition, in which  $B$  is decomposed as a product  $B=LL^T$ , with  $L$  a lower triangular matrix, is very stable. It is shown in [4] that the magnitudes of the nonzero elements are all bounded by  $b_0^{1/2}$ ,  $b_0$  being defined in (20), and that in particular the  $L_{ii}$ ,  $i=1, \dots, m$ , which act as divisors, are bounded by  $L_{ii} \leq b_0^{1/2}$ . It has not been observed for music that  $b_0 \ll 1$ .

Minimization of  $Q(\underline{a}, \underline{x})$  with respect to  $\underline{a}$  with known  $\underline{x}$  leads to the well-known covariance method [6] for estimating the prediction coefficients  $a_1, \dots, a_p$  from a given sequence of samples  $s_1, \dots, s_N$ . Instead of the covariance method the more popular autocorrelation method [6] can also be used to estimate  $a_1, \dots, a_p$ . Experiments have revealed that this modification hardly affects the results.

The iterative procedure just described already produces good results, without audible distortion, for digitized music if  $m \leq 16$  and for digitized speech if  $m \leq 100$  after only one iteration, although more iterations can improve the results.

The interpolation algorithm described in this section has been tested extensively on digitized music and speech, both represented in 16 bits. Some typical results are presented below. The original signals, with the correct values of  $s_{t(i)}$ ,  $i=1, \dots, m$ , substituted, are marked with A, the restored signals obtained after one iteration are marked with B and the restored signals obtained after 3 iterations are marked with C. For a further account of the performance of this algorithm the reader is referred to [4].

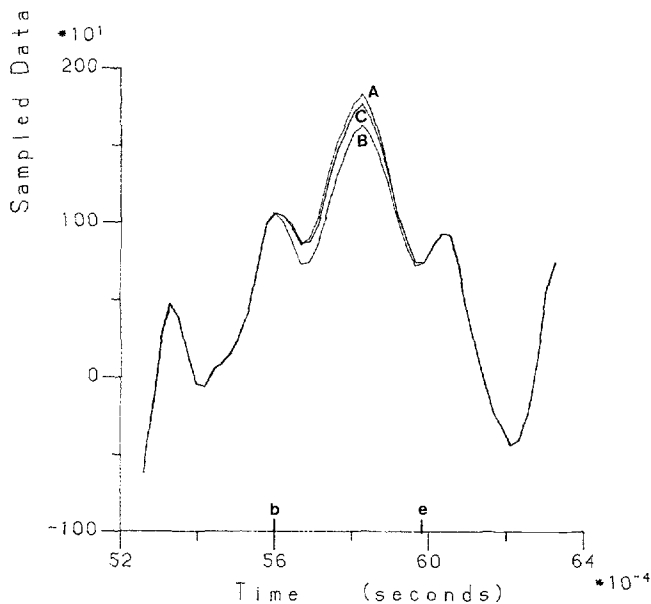


Figure 1. Interpolation results for a fragment of Beethoven's Violin Concert. Sampling frequency 44100 Hz. Segment length  $N = 512$ , burst length  $m = 16$ , process order  $p = 50$ . The position of the burst is marked on the time axis. A: original error free signal, B: result of 1 iteration, C: result of 3 iterations.

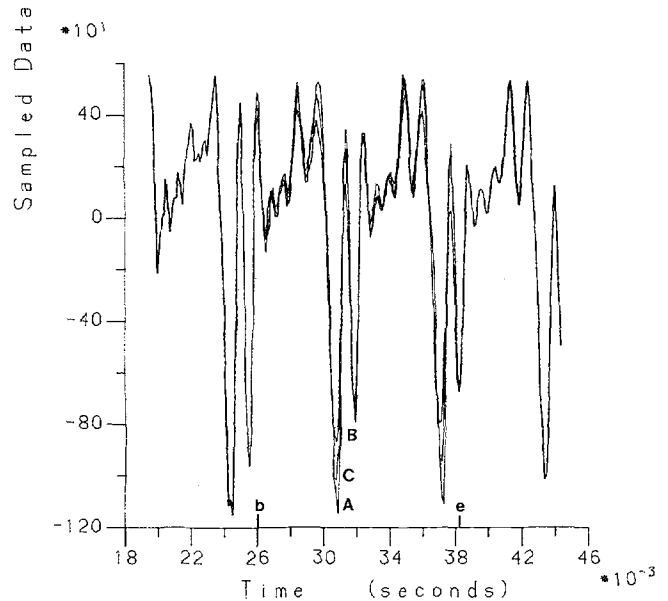


Figure 2. Interpolation results for a fragment of male speech. Sampling frequency 8000 Hz. Segment length  $N = 512$ , burst length  $m = 100$ , process order  $p = 50$ . Note that  $p=3m+2$  does not hold here. The position of the burst is marked on the time axis. A: original error free signal, B: result of 1 iteration, C: result of 3 iterations.

#### References

- [1] A.J.E.M. Janssen, L.B. Vries, "Interpolation of Band Limited Discrete Time Signals by minimizing out of Band Energy", Proc. ICASSP 84 (1984), San Diego.
- [2] R. Steel, F. Benjamin, "Sample Reduction and Subsequent Adaptive Interpolation of Speech Signals", Bell Syst. Tech. J. 62, (1983) 1365-1398.
- [3] S. Kay, "Some Results in Linear Interpolation Theory", IEEE Trans. ASSP 31 (1983) 746-749.
- [4] A.J.E.M. Janssen, R.N.J. Veldhuis, L.B. Vries, "Adaptive Interpolation of Time Discrete Signals that can be modelled as Autoregressive Processes", submitted to IEEE Trans. ASSP.
- [5] H. Akaike, "A New Look at the Statistical Model Identification", IEEE Trans. Aut. Cont. 19 (1974) 716-728.
- [6] S.M. Kay, S.L. Marple, "Spectrum Analysis-A Modern Perspective", Proc. IEEE 69, (1983) 1380-1419.