ADAPTIVE NOISE CANCELLATION FOR A CLASS OF NONLINEAR IIR FILTERS

S.A. BILLINGS
M. EL HITMY
A.R.D. CURTIS

Department of Control Engineering
University of Sheffield
Mappin Street
Sheffield S1 3JD
U.K.

Research Report No. 351

February 1989

Abstract

A new filter structure known as the NARMAX model (Nonlinear AutoRegressive Moving Average model with eXogenous inputs) is introduced as a basis for the design of a class of nonlinear filters which can be used for noise cancellation. The new design is an extention of the classical IIR filter and it is shown that the filter parameters can be estimated using a suboptimal least squares algorithm originally developed for system identification. Conditions for the convergence of the algorithm are briefly discussed and simulation results are presented which show the effectiveness of the new structure.

introduction

One of the most important problems in signal processing is the estimation of signals from noisy measurements. Whilst the optimal filter designs of Wiener and Kalman can be used to provide a solution to this problem both these algorithms are computationally demanding especially if the noise characteristics are nonstationary and adaptive noise cancellation can be implemented as an alternative. Noise cancellation (Widrow et al, 1975; Harrison, Lim and Singer, 1986; Kang and Fransen, 1987; Widrow and Stearns, 1985) consists of using a reference noise source to predict the noise component and substract it from the noise corrupted signal. This latter approach is computationally cheap when compared with the Wiener and Kalman designs and has been successfully applied to many problems including echo cancellation, speech signal processing, featal heart monitoring etc (Kang and Fransen, 1987; Harrison, Lim and Singer, 1986; Hill, 1985).

Although noise cancellation is now an established branch of signal processing which has been widely studied almost all the designs are based on linear filters. Clearly linear filters are much easier to study than nonlinear design and provided the nonlinear distortion is mild linear designs which can adapt to changing characteristics will yield an acceptable performance. In general however, linear filters will be inadequate when applied to systems where the nonlinear characteristics dominate and a few authors have begun to consider nonlinear design based on the Volterra series and the LMS algorithm (Sicuranza, Bucconi and Mitri, 1984; Coker and Simkins, 1980). The disadvantage of this approach is the large number of parameters which are required to characterise even mildly nonlinear systems.

In the present study a new nonlinear filter design is presented. This is based on the NARMAX (Nonlinear AutoRegressive Moving Average model with eXogenous inputs) which can be interpreted as an extention of a linear IIR filter. The advantage of the NARMAX filter is that it can provide a concise description for a wide class of nonlinear systems using just a few nonlinear terms. Because the number of terms required in a NARMAX filter will typically be similar to the number of terms required for a linear FIR design the potential of real time adaptation is clearly much higher than for a Volterra design. It is shown that the estimation of the NARMAX filter parameters can be interpreted as a system identification problem and can be solved using a suboptimal least squares algorithm. The convergence properties of the algorithm have been studied in the control engineering literature and consequently only a brief summary will be provided in the present study. Simulated examples are presented to demonstrate the effectiveness of the new filter and to compare the performance with Volterra based designs.

2 The Concept Of Adaptive Noise Cancellation

The structure of an adaptive noise canceller is illustrated in Fig. 1 where the signal of interest $s(t)$ is corrupted by noise $n(t)$ to yield the measusured signal $d(t)=s(t)+n(t)$. It is assumed that $n(t)$ and $s(t)$ are uncorrelated and that the noise $n(t)$passes through a second sensor with unknown transfer function $T(.)$ to yield the reference input $x(t)$. The objective of the design is to estimate the parameters of the filter $F(.)$ which operate on $x(t)$ to produce $y(t)$ which cancells the noise in $d(t-t_d)$ by minimising the mean squared error $E[\varepsilon^2(t)]$. A delay block t_d is inserted in the primary channel to ensure that the filter $F(.)$ is causal. Mathematically the idea can be expressed very simply. Assume that all the signals are zero mean and wide sense stationary, then from the diagram

$$\varepsilon(t) = d(t-t_d) - y(t) \tag{1}$$

$$= s(t-t_d) + n(t-t_d) - y(t) \tag{2}$$

such that

$$E[\varepsilon^2(t)] = E[(s(t-t_d) + n(t-t_d) - y(t))^2] \tag{3}$$

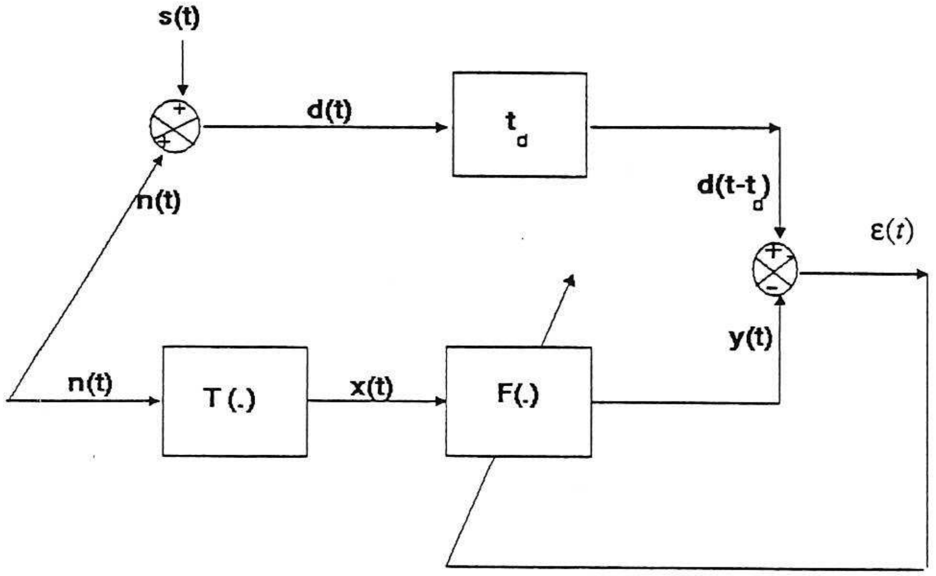

Fig. 1 Typical noise canceller

$$= E[s^2(t-t_d)] + E[(n(t-t_d) - y(t))^2] + 2E[s(t-t_d)(n(t-t_d) - y(t))] \tag{4}$$

Since $s(t)$ is uncorrelated with $n(t)$, it is also uncorrelated with $y(t)$ and therefore the last term on the RHS of eqn.(4) is equal to zero and

$$E[\varepsilon^2(t)] = E[s^2(t-t_d)] + E[(n(t-t_d) - y(t))^2] \tag{5}$$

Clearly

$$min(E[\varepsilon^2(t)]) = E[s^2(t-t_d)] + min(E[(n(t-t_d) - y(t))^2]) \tag{6}$$

which shows that $E[\varepsilon^2(t)]$ is minimised when $E[(n(t-t_d) - y(t))^2]$ is a minimum or when $y(t)$ the output of the filter cancells the noise $n(t-t_d)$ in the primary channel.

For the class of filter structures which are linear in the parameters it is possible to write $y(t) = \Phi^T(t)\theta$ where $\Phi^T(t)$ is a regression vector containing past data measurement and θ represents the vector of optimal filter parameters which are estimated to satisfy eqn.(6).

The type of filters for which $\Phi^T(t)$ can contain only lagged values of $x(t)$ include the well known FIR filters in the linear case, and the finite Volterra series in the nonlinear case. The set of equations which yield the optimal filter coefficients are obtained by expanding $E[\varepsilon^2(t)]$ as a function of the sequences $d(t-t_d)$ and $y(t)$

$$E[\varepsilon^2(t)] = E[d^2(t-t_d)] + E[y^2(t)] - 2E[d(t-t_d)y(t)] \tag{7}$$

Substituting for $y(t)$ by $\Phi^T(t)\theta$ and rearranging gives

$$E[\varepsilon^2(t)] = E[d^2(t-t_d)] + \theta^T E[\Phi(t)\Phi^T(t)]\theta - 2\theta^T E[d(t-t_d)\Phi(t)] \tag{8}$$

Defining $P=E[\Phi(t)\Phi^T(t)]$ and $R=E[d(t-t_d)\Phi(t)]$ and differentiating with respect to the filter parameter vector θ gives

$$\frac{\partial}{\partial\theta^T}(E[\varepsilon^2(t)]) = 2P\theta - 2R \tag{9}$$

The minimum is achieved by setting the gradient of the mean squared error function to zero

$$\frac{\partial}{\partial\theta}(E[\varepsilon^2(t)]) = 0$$

or assuming that P is invertible

$$\hat{\theta} = P^{-1}R \tag{10}$$

Various recursive algorithms have been derived to implement eqn.(10) and to estimate θ including stochastic gradient algorithms such as LMS. When the vector $\Phi(t)$ contains lagged $y(t)$ and $x(t)$ values it is convenient to reformulate the problem of estimating the filter parameters into a system identification formulation and this is considered in the next section.

3 Noise Cancellation And System Identification

Previous authors (Eweda and Macchi, 1987; Cioffi and Kailath, 1984; Cowan, 1987) have considered the design of FIR filters using the RLS algorithm as an alternative to the LMS method and have demonstrated the superior convergence time of the former approach. Other authors have interpreted the design of linear IIR filters in terms of an output error identification form using variants of the Steilgliz-McBride (Hong, 1986; Hong and Jenkins, 1987), recursive maxmimum likelihood and hyperstability algorithms (Friedlander, 1982; Larimore et al, 1980). The advantage of the IIR designs is the reduced number of filter parameters required to describe the system and hence the possibility of a rapid convergence rate. The problems of convergence of IIR designs which were noted in earlier studies based on the LMS algorithm have been partly mitigated by invoking results from system identification which contain extensive investigations into the convergence of the RLS algorithm and variants of it. An exellent discussion of these concepts was recently provided by Soderstrom and Stoica (1988) who proposed an instrumental variable algorithm for linear adaptive filtering.

If nonlinear noise cancellation is to be successful then the filter will need to be described by a small parameter set and this tends to exclude Volterra based FIR designs and points to nonlinear IIR filters. This in turn suggests that the results from system identification should be modified to apply to this problem and RLS type algorithms should be utilised. It is for these reasons that a suboptimal least squares (SOLS) estimator which is a variant of the RLS is introduced in the present study for a class of nonlinear IIR filters.

The interpretation of IIR filter design as a system identification problem and the introduction of the suboptimal least squares algorithm will be considered below for the general filter structure in Fig. 1 and then adopted for the NARMAX model design in section 4.

From Fig. 1 and the analysis of section 2 the optimal design of the filter $F(.)$ will be the inverse of $T(.)$ such that

$$y(t) = n(t-t_d) = T^{-1}(x(t)) = F(x(t)) \tag{11}$$

where $F(.)$ and $T(.)$ are to be interpreted as operators to be defined later and where it is assumed that T^{-1} exists and is stable. Since

$$d(t-t_d) = n(t-t_d) + s(t-t_d) \tag{12}$$

using eqn.(11) yields

$$d(t-t_d) = T^{-1}(x(t)) + s(t-t_d) \tag{13}$$

In system identification studies eqn.(13) would be interpreted as a system $T^{-1}(x(t)) = F(.)$ with input $x(t)$, measured output $d(t-t_d)$ and an unobservable coloured noise $s(t-t_d)$ as illustrated in Fig. 2

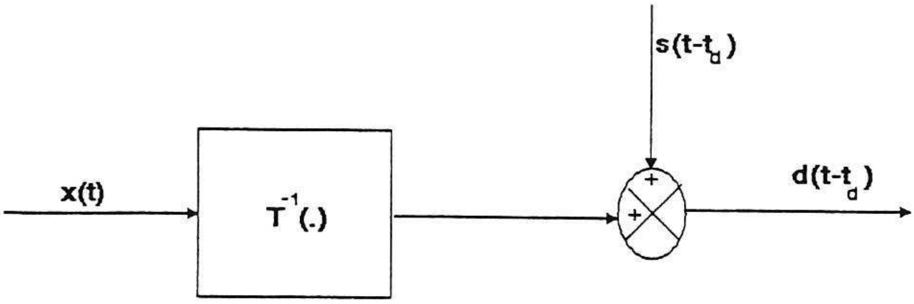

Fig.2 Interpretation of eqn.(13)

Estimation of the operator $T^{-1}(.)$ can be considered as a parallel identification scheme as illustrated in Fig. 3.

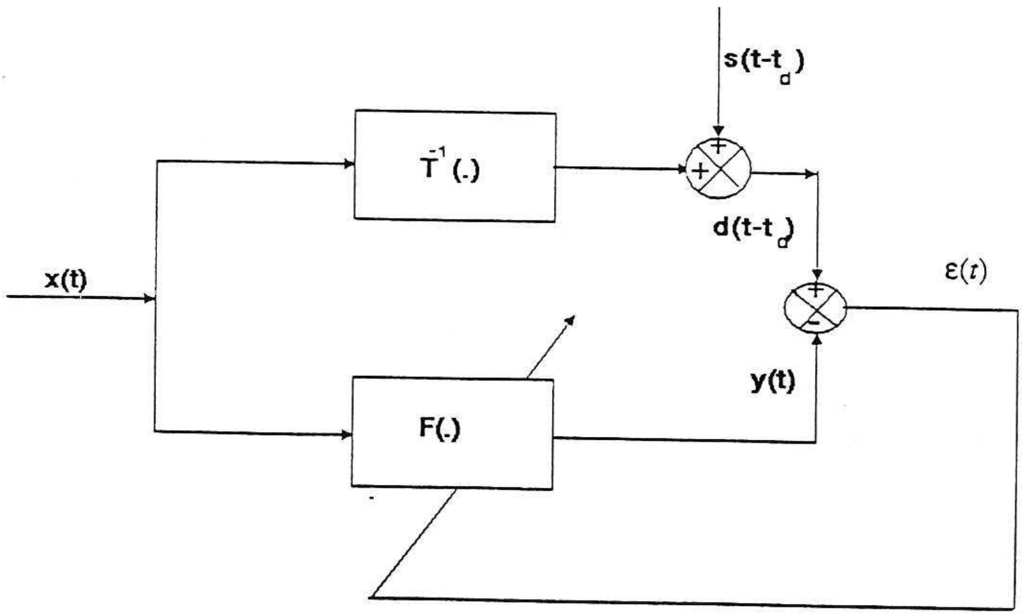

Fig. 3 Parallel system identification

Minimising the sum of the errors squared $\frac{1}{N}\sum_{i=1}^{N}\varepsilon^2(i)$ and assuming convergence would yield $\hat{F}(.) = T^{-1}(.)$ and $y(t)$ would be the optimal prediction of $n(t-t_d)$. Interpretation of the identification is dependent upon the form of $F(.)$. If a linear FIR filter is selected eqn.(13) can be expressed as

$$d(t-t_d) = F(x(t),x(t-1),....,x(t-n_x)) + s(t-t_d) \tag{14}$$

or more concisely

$$d(t-t_d) = \Phi_1^T(t)\theta_x + s(t-t_d) \tag{15}$$

where

$$\Phi_1^T(t) = [\ x(t),x(t-1),.....,x(t-n_x)]$$

$$\theta_x^T = [\ b_0,b_1,b_2,......,b_{n_x}]$$

Because $s(t)$ and $x(t+\tau)$ are zero mean uncorrelated for all τ an LMS or RLS algorithm can be applied to yield umbiased estimates of the parameter vector θ_x. The RLS algorithm for the model of eqn.(15) is defined by the choice

$$\Phi(t) = \Phi_1(t)$$

$$\theta(t) = \theta_x(t) \tag{16}$$

$$zz(t) = \Phi_1(t)$$

In the following unified algorithm (Soderstrom,Ljung and Gustavsson, 1978; Ljung and Soderstrom, 1983)

$$\hat{\theta}(t) = \hat{\theta}(t-1) + K(t)\varepsilon(t) \tag{17}$$

$$K(t) = \frac{P(t-1)zz(t)}{\lambda(t) + \Phi^T(t)P(t-1)zz(t)} \tag{18}$$

$$P(t) = \frac{1}{\lambda(t)}\left[P(t-1) - \frac{P(t-1)zz(t)\Phi^T(t)P(t-1)}{\lambda(t) + \Phi^T(t)P(t-1)zz(t)}\right] \tag{19}$$

$$\varepsilon(t) = d(t-t_d) - \Phi^T(t)\hat{\theta}(t-1) \tag{20}$$

$$\lambda(t) = \lambda_0\lambda(t-1) + 1 - \lambda_0 \tag{21}$$

where $\lambda(t)$ is a variable forgetting factor which usually takes values between 0.9 and 1.0.

If a linear IIR filter is selected eqn.(13) takes the form

$$d(t-t_d) = \frac{B(z^{-1})}{A(z^{-1})}x(t) + s(t-t_d) \tag{22}$$

where

$$T^{-1}(.) = \frac{B(z^{-1})}{A(z^{-1})}$$

$$A(z^{-1}) = 1 + a_1 z^{-1} + ... + a_{n_d} z^{-n_d}$$

$$B(z^{-1}) = b_0 + b_1 z^{-1} + ... + b_{n_x} z^{-n_x}$$

Expanding eqn.(22)

$$d(t-t_d) = [-d(t-t_d-1),.....,-d(t-t_d-n_d),x(t),....,x(t-n_x)]\ [a_1,a_2,\ldots,a_{n_d},b_0,\ldots,b_{n_x}]^T$$

$$+ A(z^{-1})s(t-t_d) \tag{23}$$

$$= \Phi_2^T(t)\theta_{dx} + A(z^{-1})s(t-t_d)$$

and the elements of $\Phi_2^T(t)$ are now correlated with $s(t-t_d)$. Application of the RLS algorithm would therefore provide biased estimates. This is analogous to the identification problem in the presence of coloured noise and whilst several solutions have been proposed the suboptimal least squares algorithm introduced by Moore (1982) for linear systems and Billings and Voon (1984) for nonlinear models matches the requirement of the IIR noise cancellation filter exactly.

If the output $y(t)$ could be monitored eqn.(23) could be expressed as

$$d(t-t_d) = [-y(t-1),.....,-y(t-n_d),x(t),....,x(t-n_x)]\ [a_1,a_2,\ldots,a_{n_d},b_0,\ldots,b_{n_x}]^T$$

$$+ s(t-t_d) \tag{24}$$

$$= \Phi_3^T(t)\theta_{dx} + s(t-t_d)$$

and the elements of Φ_3^T(t) are no longer correlated with the noise $s(t-t_d)$. Although the signal $y(t)$ is unavailable for measurement it can be estimated by recursively computing the predicted output using

$$\hat{y}(t) = \hat{\Phi}_3^T(t)\hat{\theta}_{dx}(t) \tag{25}$$

where $\hat{\theta}_{dx}(t)$ denotes the estimated parameter vector at the most recent iteration. Thus the estimation algorithm now becomes eqns. (17)-(21) followed by eqn. (25). $\hat{y}(t)$ is considered to be a better estimate of $y(t)$ than if it was estimated using the estimated parameter vector from the previous iteration $\hat{\theta}_{dx}(t-1)$ and

$$\hat{\Phi}_3^T(t) = [-\hat{y}(t-1),.....,-\hat{y}(t-n_d),x(t),....,x(t-n_x)]$$

The noise free output $y(t)$ in eqn.(24) is therefore replaced by the prediction $\hat{y}(t)$. The algorithm is computationally simple and can be implemented as a variant of RLS using the following definitions in (17)-(21)

$$\Phi(t) = \hat{\Phi}_3(t)$$

$$\theta(t) = \theta_{dx}(t) \qquad (26)$$

$$zz(t) = \Phi(t)$$

The advantage of SOLS compared with the alternative algorithms is that unbiased estimates of $F(.)$ can be obtained by operating on the predicted filter output $\hat{y}(t)$. This can only be achieved when the noise $s(t-t_d)$ in Fig. 3, is additive at the output but this corresponds exactly to the noise cancellation problem and means that it is unnessecary to consider the problem of estimating a noise model which would both increase the complexity of the algorithm and the size of the parameter vector. Whilst these are desirable properties for linear designs they become virtual necessities in the nonlinear case.

Convergence properties of the SOLS and related output error schemes have been extensively studied by Moore (1982) who showed that the algorithm will achieve global convergence in the presence of persistently exciting input signals and for $\lambda(t)=1$ in eqns. (17)-(20) providing $[A^{-1}(z^{-1})-\frac{1}{2}]$ is strictly positive real . Moore also showed that the convergence of the algorithm will not be critically sensitive to the colour of the noise as are related extended least squares type schemes which require simultaneous noise model estimation.

Whilst eqns. (17)-(20) represent the standard formulation given for RLS based estimators numerical ill conditioning of the algorithm may arise if the equations are coded directly. This problem occurs because the covariance matrix $P(t)$ may become nonpositive definite due to roundoff errors and can be avoided by using a numerically stable implementation such as the UD factorisation algorithm of Bierman (1977). (Appendix I)

4 Noise Cancellation for Nonlinear Systems

The design of noise cancellation methods for nonlinear systems depends criticaly on the model that is used to represent $F(.)$ in the reference path of Fig. 1. Three possible choices for $F(.)$ are considered below together with details of how the linear estimation algorithms of section 3 can be modified in each case.

4.1 The Volterra Model

The Volterra series

$$y(t) = h_0 + \sum_{i=0}^{n_1} h_1(i)x(t-i) + \sum_{i=0}^{n_{21}}\sum_{j=0}^{n_{22}} h_2(i,j)x(t-i)x(t-j) + ... \tag{27}$$

is a generalisation of the linear convolution integral and was introduced by Volterra early in the twentieth century. The functionals $h_i(k_1,....,k_i)$ which are referred to as Volterra kernels are bounded, symmetric functions of their arguments and for causal systems $h_i(k_1,....,k_i)=0$ for any $k_i<0$. Inspection of Fig. 2 shows that if the Volterra model were used to represent $T^{-1}(.)$ then

$$d(t-t_d) = b_0 + \sum_{i=0}^{n_1} b_1(i)x(t-i) + \sum_{i=0}^{n_{21}}\sum_{j=0}^{n_{22}} b_2(i,j)x(t-i)x(t-j) + ... + s(t-t_d) \tag{28}$$

where $b_j(.) = h_j(.)$

Eqn.(28) can be interpreted as a nonlinear extension of the FIR filter of eqn.(14) and could readily be implemented within an adaptive noise cancellation scheme by rewriting eqn.(28) as

$$d(t-t_d) = \Phi_4^T(t)\theta_V + s(t-t_d) \tag{29}$$

where

$$\Phi_4^T(t) = [1, x(t), x(t-1), ..., x(t-n_1), x^2(t), x(t)x(t-1), ...]$$

$$\theta_V^T = [b_0, b_1(0), b_1(1), ..., b_1(n_1), b_2(0,0), b_2(0,1), ...]$$

The filter parameters could then be estimated using an LMS algorithm or the RLS algorithm of eqns. (17)-(21) could be used by setting

$$\Phi(t) = \Phi_4(t)$$

$$\theta(t) = \theta_V(t) \tag{30}$$

$$zz(t) = \Phi_4(t)$$

Whilst Volterra models have been extensively studied in both system identification and nonlinear noise cancellation they require a very large parameter set to represent even simple nonlinear systems and this is often not compatible with the need for fast adaptation. This curse of dimensionality arises because in the Volterra model the output is expressed as a function of past inputs only. It is a nonlinear FIR filter. The advantage of using the Volterra model is that the linear estimation results whether based on LMS or RLS carry over directly with a simple redefinition of the relevant matrices. The disadvantage is that the enormous number of terms may prevent a real time implementation without very fast processors. On balance it would seem that the disadvantages will dominate and an alternative design must be saught.

4.2 The NARMAX Model

The nonlinear difference equation model

$$y(t) = f'(y(t-1),....,y(t-n_y),x(t),...,x(t-n_x)) \tag{31}$$

where $f'(.)$ is some nonlinear function which was originally introduced by Billings and Leontaritis (1981). A rigorous derivation of eqn.(31) together with conditions for its existence were given by Leontaritis and Billings (1985). An equivalent description for nonlinear stochastic systems can be derived by considering input-output maps based on conditional probability density functions to yield, in terms of the variables of Fig. 2,

$$d(t-t_d) = dc + f(d(t-t_d-1),....,d(t-t_d-n_d),x(t),...,x(t-n_x),s(t-t_d-1),.....,s(t-t_d-n_d)) + s(t-t_d) \tag{32}$$

where dc represents a dc level.

This model is referred to as the NARMAX model (Nonlinear AutoRegressive Moving Average model with eXogenous input). The Hammerstein, Wiener, bilinear, Volterra and other well known nonlinear models can be shown to be special cases of eqn.(32). In the original derivation of eqn.(32) $f(.)$ was defined as some general nonlinear function but it will suffice to expend $f(.)$ as a polynomial in the present analysis.

The NARMAX model is just a nonlinear IIR filter. The advantage of this description is that the NARMAX model will require a very small parameter set compared with the Volterra model because it is an expansion in terms of past inputs and outputs. For example a model of the form

$$d(t-t_d) = a_1 d(t-t_d-1) + a_2 d^3(t-t_d-1) + b_1 x(t) + b_2 x(t-1) + s(t-t_d) \qquad (33)$$

would require the estimation of four parameters a_1, a_2, b_1, b_2 to describe it as a NARMAX model. A Volterra series expansion of eqn.(33) would however contain many many more terms because lagged outputs cannot be modelled directly but have to be approximated by a possibly infinite expansion in terms of lagged inputs only. The NARMAX model thus provides a very concise description of nonlinear systems, often with fewer parameters than a linear FIR filter. The disadvantage of the NARMAX model, for noise cancellation, is that like the linear IIR designs the parameter estimation is more involved because the regression vector is composed of lagged inputs and outputs.

Notice that even though in the noise cancellation problem the noise is additive at the output, Fig. 2, the NARMAX model description will take the general form of eqn.(32) where the noise $s(t-t_d)$ appears in the model expansion as multiplicative with both $d(t-i)$ and $x(t-j)$. This is a direct consequence of using lagged outputs in the model. Expanding eqn.(32) as a polynomial NARMAX and grouping terms yields

$$\begin{aligned} d(t-t_d) = dc &+ G^{dx}[d(t-t_d-1),\ldots,d(t-t_d-n_d),x(t),\ldots,x(t-n_x)] \\ &+ G^{dsx}[d(t-t_d-1),\ldots,d(t-t_d-n_d),x(t),\ldots,x(t-n_x),s(t-t_d-1),\ldots,s(t-t_d-n_d)] \\ &+ G^{s}[s(t-t_d-1),\ldots,s(t-t_d-n_d)] + s(t-t_d) \end{aligned} \qquad (34)$$

where $G^{dx}[.]$ is a function of $d(t)$ and $x(t)$ only, $G^{dxs}[.]$ represents all the cross product terms involving $s(t)$ and $G^{s}[.]$ is a function of $s(t)$ only. All functions are polynomial. Separating out the unknown parameters gives

$$\begin{aligned} d(t-t_d) &= \Phi_S^T(t)\theta + s(t-t_d) \\ &= [\Phi_{dx}^T(t)\ \ \Phi_{dxs}^T(t)\ \ \Phi_s^T(t)] \begin{bmatrix} \theta_{dx} \\ \theta_{dxs} \\ \theta_s \end{bmatrix} + s(t-t_d) \end{aligned} \qquad (35)$$

where

$$G^{dx}[.] = \Phi_{dx}^T(t)\theta_{dx}$$

$$G^{dxs}[.] = \Phi_{dxs}^T(t)\theta_{dxs} \qquad (36)$$

$$G^{s}[.] = \Phi_{s}^T(t)\theta_{s}$$

Any design based on eqn.(35) directly would involve estimating θ_{dx} and the noise parameters θ_{dxs} ,θ_s to ensure unbiased estimates of the cancellation filter parameters θ_{dx}. Although parameter estimation algorithms which achieve this objective are available fortunately in the noise cancellation application this complexity can be avoided using the suboptimal least squares algorithm derived by Billings and Voon (1984).

Following the same arguments as in section 3 if the output $y(t)$ could be monitored eqn.(32) could be expressed as

$$d(t-t_d) = dc + f(y(t-1),....,y(t-n_d),x(t),...,x(t-n_x)) + s(t-t_d) \qquad (37)$$

and the cross product noise terms are eliminated. The signal $y(t)$ cannot be measured directly but it can be predicted using

$$\hat{y}(t) = \Phi_{\hat{y}x}^T(t)\hat{\theta}_{dx}(t) \qquad (38)$$

where

$$\Phi_{\hat{y}x}^T(t) = [-\hat{y}(t-1),.....,-\hat{y}(t-n_d),-\hat{y}^2(t-1),....,-\hat{y}(t-1)\hat{y}(t-2),....x(t),...,x(t-n_x),$$
$$x^2(t),x(t)x(t-1).....]] \qquad (39)$$

The estimation algorithm is given by eqns (17)-(21) with the definitions

$$\Phi(t) = \Phi_{\hat{y}x}(t)$$

$$\theta(t) = \theta_{dx}(t) \qquad (40)$$

$$zz(t) = \Phi(t)$$

so that unbiased estimates of the nonlinear canceller are obtained without the need to estimate a noise model. The maximum number of entries in the θ_{dx} vector is given by

$$n = M + 1$$

where

$$M = \sum_{i=1}^{l} n_i$$

$$n_i = \frac{n_{i-1}(n_d + n_x + i - 1)}{i} \quad , \quad n_0 = 1$$

and l is the degree of polynomial expansion for $f(.)$ in eqn.(37). To ensure a parsimonious model description and hence to minimize the computational load of the parameter estimator it is preferable to detect the correct structure or which terms to include in the NARMAX model prior to final estimation. Several parameter estimation, structure detection algorithms which achieve these objectives have been derived but in the present study it will be assumed that these have been applied prior to implementing a NARMAX noise canceller or that the structure is known a priori.

Convergence analysis of the linear SOLS algorithm derived in section 3 does not apply directly to the nonlinear case because there is no simple generalisation of the positive real condition. A discussion of the concepts involved has been given by Chen and Billings (1988) in a study of a recursive prediction error estimator for nonlinear models.

Notice that the SOLS algorithm is preferable to an extended least squares or recursive prediction error method because both these alternatives would introduce the additional complexity of noise model estimation. Whilst instrumental variable (IV) offers a viable alternative if the model is linear, subject to some restrictions noticed by Moore, IV will in general yield biased estimates when the system is nonlinear (Billings,Voon, 1984).

4.3 The ARNX Model

A subclass of the model eqn.(32) for which the convergence results of the linear SOLS algorithm are still applicable is the ARNX (AutoRegressive model with Nonlinear eXogenous inputs)

$$d(t-t_d) = \frac{f^x(x(t),x(t-1),....,x(t-n_x))}{A(z^{-1})} + s(t-t_d) \tag{41}$$

where $f^x(.)$ is a polynomial in lagged $x(t)$'s only. This model can be viewed as a linear IIR filter with nonlinear FIR terms.The ARNX filter is somewhat between the NARMAX model and the Volterra series with the advantage that the convergence of the SOLS will be dependent upon a positive real condition of the denominator as in linear IIR designs.

5 Simulation Results

The estimation of a saw tooth signal varying between -1 and 1 which was burried in a white noise signal was tested on the noise path models which were introduced in the preceding sections.

Fig. 4(a)-4(c) show a trace of 1000 samples of the noise n(t), the signal s(t) and the measured signal d(t). These signals remain the same for the three simulated cases considered in this analysis and it is the noise path model which takes a different form in each case.

5.1 Simulation S1

A linear system explicitly given by

$$x(t) = -0.5x(t-1) + n(t) - 1.722n(t-1) + 0.9n(t-2) \tag{42}$$

called simulation S1 was used to generate the pair $n(t),x(t)$ at each time sample t where $n(t)$ was a uniformly distributed signal varying between -2 and +2. The optimal noise cancelling filter $F(.)$ is the exact inverse of the noise model eqn.(42)

$$y(t) = x(t) + 0.5x(t-1) + 1.722y(t-1) - 0.9y(t-2) \tag{43}$$

or

$$F(z^{-1}) = \frac{1 + 0.5z^{-1}}{1 - 1.722z^{-1} + 0.9z^{-2}} \tag{44}$$

So in this case t_d in Fig.1 is zero. The poles of this optimal transfer function are close to the unit circle and in the time domain the impulse response is given by

$$h(t) = 0.95^t \{ cos(0.436t) + 3.4sin(0.436t) \} \tag{45}$$

whose magnitude decreases to 10% only after 44 terms. $F(z^{-1})$ was modelled by an IIR filter of the form

$$\hat{y}(t) = \hat{dc}(t) + \hat{a}_1(t)x(t) + \hat{a}_2(t)x(t-1) + \hat{b}_1(t)\hat{y}(t-1) + \hat{b}_2(t)\hat{y}(t-2) \tag{46}$$

where the parameters were updated using the SOLS algorithm with $P(0) = 10^6 I$, I is the unit matrix and $\lambda(0) = 0.95$, $\lambda_0 = 0.99$. A high degree of noise suppression, around 18 dB improvement, is acheived in about 2000 adaptations. The parameter estimates and the trace of 1000 samples of the estimated signal $\hat{s}(t-t_d)$ are shown in *Table I(a)*. and Fig. 5(a) respectively.

For comparison, a conventional FIR noise cancelling filter of the form

$$\hat{y}(t) = \hat{dc}(t) + \sum_{i=0}^{n_x} \hat{a}_i(t)x(t-i) \tag{47}$$

was applied to the same data. *Table I(b)* illustrates the improved performance as the number of adaptive weights is increased but only at a severe computational cost. Fig. 5(b)-(c) illustrates the traces of the estimated signal for 1000 adaptations, which were taken between iteration N^0 1000 and 2000. The filter length considered was 40 in Fig. 5(b) and 30 in Fig. 5(c) respectively.

	dc	$x(t)$	$x(t-1)$	$y(t-1)$	$y(t-2)$
True Values	0.0	1.0	0.5	1.722	-0.9
Estimated Values	4.34e-03	0.9793	0.4875	1.709	-0.87
Standard Deviation	0.397e-02	0.5e-02	0.3e-02	0.4e-02	0.14e-01

Table I(a). IIR filter parameter estimates at iteration N^0 2000 for *simulation S1*

N^0*adaptive weights*	$n_x=8$ $n=10$	$n_x=18$ $n=20$	$n_x=28$ $n=30$	$n_x=38$ $n=40$
Noise Reduction	9.6	14.2	19.8	22.7

Table I(b). Noise Reduction in dB obtained in 2000 *adaptations* for *simulation S1*

5.2 Simulation S2

A Wiener model with third order dynamics and third degree nonlinearity described as

$$v(t) = -0.5v(t-1) + n(t-1) - 1.722n(t-2) + 0.9n(t-3) \tag{48}$$

$$x(t) = 0.1v(t) + 0.02v^2(t) + 0.005v^3(t)$$

defined simulation S2 which used the same input data as S1 but with t_d in Fig.1 equal to one. The adaptive filter was modelled by three different structures namely, the nonlinear IIR filter, the ARNX model and the nonlinear FIR filter. Fig. 6 shows the traces of the estimated signal $\hat{s}(t-t_d)$ for each choice of filter and *Table II* illustrates the noise power reduction obtained in each case. The structure for these filters used in the simulations is specified in *Table II* where n is the total number of terms in the parameter vector, l represents the degree of polynomial expansion and n_x and n_d represent the maximum lag in the input and output respectively. As expected, improved performance is achieved when the filter model incorporates past output terms. The best performance obtained in this case is when the filter is of an ARNX type because this structure matches closely the inverse model of the Wiener system.

N^0 adaptive weights	n_x=3 l=3,n=35	n_x=1, n_d=5 l=3,n=15	n_x=2, n_d=1 l=2,n=15
Filter Type	NFIR	ARNX	NIIR
Noise Reduction	14.8	16.2	15.8

Table II. Noise Reduction in dB obtained in 2000 *adaptations* for *simulation S2*

5.3 Simulation S3

An implicit NARX model with second order dynamics and second degree non-linearity described by

$$x(t) = 0.25x(t-1) + 0.1x(t-2) + 0.5n(t-1) + 0.1n(t-2) - 0.2n(t-3) + 0.1n^2(t-2) + 0.08n(t-2)x(t-1) \quad (49)$$

defined simulation S3 which used the same input $n(t)$ as S1 but with t_d=1. The optimal noise cancelling filter $F(.)$ is given by the exact inverse of the noise channel model eqn.(49)

$$y(t) = 2x(t) - 0.5x(t-1) - 0.2x(t-2) - 0.2y(t-1) + 0.4y(t-2) - 0.2y^2(t-1) - 0.16y(t-1)x(t-1) \quad (50)$$

if the filter structure matches the noise channel inverse model then the filter will be of the form

$$\hat{y}(t) = \hat{a}_1(t)x(t) + \hat{a}_2(t)x(t-1) + \hat{a}_3(t)x(t-2) + \hat{b}_1(t)\hat{y}(t-1) + \hat{b}_2(t)\hat{y}(t-2) + \hat{c}_1(t)\hat{y}^2(t-1) + \hat{c}_2(t)\hat{y}(t-1)x(t-1) \quad (51)$$

The parameter estimates associated with the filter structure of eqn. (51) and the trace of the estimated signal $\hat{s}(t-t_d)$ are shown in *TableIII(a)*. and Fig. 7(a) respectively. A high degree of noise suppression giving around 25 dB reduction was achieved.

	$x(t)$	$x(t-1)$	$x(t-2)$	$y(t-1)$	$y(t-2)$	$y^2(t-1)$	$y(t-1)x(t-1)$
True Values	2.0	-0.5	-0.2	-0.2	0.4	-0.2	-0.16
Estimated Values	1.97	-0.5477	-0.2457	-0.1795	0.4347	-0.1862	-0.1652
Standard Deviation	0.016	0.02	0.0156	0.014	0.01	0.019	0.039

Table III (a), NIIR filter parameter estimates at iteration N^0 2000 for *simulation S3*

For comparison, an ARNX noise cancelling filter and a nonlinear FIR filter were successively applied to the same data. The results obtained are illustrated in *TableIII (b)*.

N^0 *adaptive weights*	$n_x=3$ $l=3, n=35$	$n_x=2,\ n_d=5$ $l=2, n=15$	*Filt. eqn.*(51) $n=7$
Filter Type	NFIR	ARNX	NIIR
Noise Reduction	12.5	12.7	25.0

Table III (b) Noise Reduction in dB obtained in 2000 *adaptations* for *simulation S3*

Inspection of the results in *Table III(b)* shows that whilst the Volterra (NFIR) and ARNX filters have produced a noise reduction of just over 12 dB the NARMAX (NIIR) design has performed significantly better with a 25 dB noise reduction. This was achieved using just seven estimated parameters for the NARMAX filter compared with thirty five or five times as many parameters for the NFIR design and twice as many parameters, fifteen, for the ARNX filter.

In simulations S1 and S2 the cancellation filter could be realised by all the designs NFIR, ARNX and NIIR. With different parameter sets they all provided comparable noise reductions. In simulation S3 however the NFIR and ARNX filters, even with a considerable number of filter parameters, cannot realise exactly the required cancellation filter and this is why they perform so badly compared with the NARMAX which can represent a much wider class of noise cancellers.

Conclusion

A new design for nonlinear IIR filters has been proposed based on the NARMAX representation. It has been shown that the parameters in both linear and nonlinear IIR designs can be estimated using a suboptimal least squares algorithm and the NARMAX filter has been shown to perform significantly better than alternative designs for a wide class of nonlinear systems. Extensions of this work to include alternative designs for nonlinear systems and the formulation of a subsumption architecture for detecting the filter structure and implementing the filters will be considered in forthcoming publications.

References

BIERMAN, G. J., (1977). "Factorisation Methods For Discrete Seqential Estimation," *New York Academic Press.*

BILLINGS, S. A. AND LEONTARITIS, I. J., (1981). "Identification of Nonlinear Systems Using Parameter Estimation Techniques," *Proc. IEE, Conference on Control and its Applications, Warwick Univ.*, pp. 183-187.

BILLINGS, S. A. AND VOON, W. S. F., (1984). "Least Squares Parameter Estimation Algorithms for Nonlinear Systems," *Int. J. Systems Sci*, vol. 15 No. 6, pp. 601-615.

CHEN, S. AND BILLINGS, S. A., (1988). "A Recursive Prediction Error Parameter Estimator For Nonlinear Models," *Tnt. J. Control (To appear).*

CHEN, S. AND BILLINGS, S. A., (1988). "Representations of Nonlinear Systems: The Narmax Model," *Research Report No.330*, Department of Control Engineering, University of Sheffield, Mappin Street, Sheffield S13JD.

CIOFFI, J. M. AND KAILATH, T., (1984). "Fast, Recursive-Least-Squares Transversal Filters for Adaptive Filtering," *IEEE Trans. Accoustics, Speech and Signal Processing*, vol. ASSP-32 No.2, pp. 304-337.

COKER, M. J. AND SIMKINS, D. N., (1980). "A Nonlinear Adaptive Noise Canceller," *proc IEEE*, pp. 470-473.

COWAN, C. F. N., (1987). "Performance Comparison of Finite Linear Adaptive Filters," *IEE Proceedings*, vol. 134, Pt. F, pp. 211-215.

EWEDA, E. AND MACCHI, O., (1987). "Convergence of the RLS and LMS Adaptive Filters," *Trans. Circuits and Systems*, vol. 7, pp. 799-803.

FRIEDLANDER, B., (1982). "System Identification Techniques for Adaptive Signal Processing," *Trans. Circuits and Systems*, vol. 1, pp. 3-41.

GIORDANO, A. A. AND HSU, F. M., (1985). "Least Square Estimation With Application To Digital Signal Processing," *John Wiley & Sons,Inc.*.

HARRISON, W. A., LIM, J. S., AND SINGER, E., (1986). "A New Application of Adaptive Noise Cancellation," *Trans. Accoustics, Speech and Signal Processing*, vol. ASSP-3 No.1, pp. 21-2.

HILL, P. D., (1985). "Real Time Real Life Adaptive Noise Cancellation Using The Homogenious Algorithm," *MSEE Thesis*, West Virginia University.

HONG, F., (1986). "New Adaptive IIR Filtering Algorithms," *Ph.D Thesis*.

HONG, F. AND JENKINS, W. K., (1987). "A New Adaptive IIR Filter," *Trans. Circuits and Systems*, vol. CAS-33 No. 10, pp. 939-946.

KANG, G. S. AND FRANSEN, L. J., (1987). "Experimentation with an Adaptive Noise Cancellation Filter," *Trans. Circuits and Systems*, vol. CAS-34 No.7, pp. 753-758.

LARIMORE, M. J., TREICHLER, J. R., AND JOHNSON, R., (1980). "SHARF: An Algorithm for Adapting IIR Digital Filters," *Trans. Accoustics, Speech and Signal Processing*, pp. 428-440.

LEONTARITIS, I. J. AND BILLINGS, S. A., (1985). "Input Output Parametric Models for Nonlinear Systems Part I Deterministic Nonlinear Systems. Part II Stochastic Nonlinear Systems," *Int. J. Control*, vol. 41, pp. 303-344.

LJUNG, L. AND SODERSTROM, T., (1983). "Theory and Practice of Recursive Identification," *MIT Press, Cambridge*.

MOORE, J. B., (1982). "Global Convergence of Output Error Recursions In Colored Noise," *Proc. IEEE Trans. on Automatic Control*, vol. vol. AC-27 No.6, pp. 1189-1199.

SICURANZA, G. L., BUCCONI, A., AND MITRI, P., (1984). "Adaptive Echo Cancellation with Nonlinear Digital Filters," *proc IEEE*, pp. 3101-3104.

SODERSTROM, T., LJUNG, L., AND GUSTAVSSON, I., (1978). "A Theoritical Analysis of recursive identification methods," *Automatica*, vol. 41, pp. 177-191.

SODERSTROM, T. AND STOICA, P., (1988). "On Some System Identification Techniques for Adaptive Filtering," *Trans. Circuits and Systems*, vol. vol. 35 No.4, pp. 457-461.

WIDROW, B., GLOVER, J. R., MCCOOL, J. M., KAUNITZ, J., WILLIAMS, C. S., HEARNS, R. H., ZEIDLER, J. R., DONG, E., AND GOODLIN, R. C., (1975). "Adaptive Noise Cancelling Principles and Applications," *Proc IEEE*, vol. 63, pp. 1692-1716.

WIDROW, B. AND STEARNS, S. D., (1985). "Adaptive Signal Processing," *Prentice Hall Inc.*

Appendix I
UD updating algorithm

The matrix P(t) is factorised as $U(t)D(t)U^T(t)$ where D is a diagonal matrix $[diag(d_i)]$ and U is a unit upper-triangular matrix, i.e $u_{ii}=1$, i=1,2,...,n and $u_{ij}=0$, i>j. From the updating eqn. (19) for P(t):

$$P(t) = U(t)D(t)U^T(t) = \frac{1}{\lambda(t)}\left[P(t-1) - \frac{P(t-1)\Phi(t)\Phi^T(t)P(t-1)}{\lambda(t) + \Phi^T(t)P(t-1)\Phi(t)}\right] \tag{A.1}$$

$$= \frac{1}{\lambda(t)}U(t-1)\left[D(t-1) - \frac{D(t-1)U^T(t-1)\Phi(t)(D(t-1)U^T(t-1)\Phi(t))^T}{\lambda(t) + \Phi^T(t)P(t-1)\Phi(t)}\right]U^T(t-1)$$

Denoting $U^T(t-1)\Phi(t)$ by **f** and $D(t-1)$**f** by **g** and $\alpha(t) \overset{\Delta}{=} \lambda(t) + \Phi^T(t)P(t-1)\Phi(t) = \lambda(t) + \mathbf{f}^T\mathbf{g}$ then eqn.(A.1) becomes:

$$U(t)D(t)U^T(t) = \frac{1}{\lambda(t)}U(t-1)\left[D(t-1) - \frac{\mathbf{g}\mathbf{g}^T}{\alpha(t)}\right]U^T(t-1) \tag{A.2}$$

If U^* and D^* are found such that

$$D(t-1) - \frac{\mathbf{g}\mathbf{g}^T}{\alpha(t)} = U^*D^*U^* \tag{A.3}$$

then from (A.2)

$$D(t) = \frac{D^*}{\lambda(t)}\ ,\ U(t) = U(t-1)U^* \tag{A.4}$$

The factorisation (A.3) is performed by the Agee-Turner algorithm, giving D^* explicitly and U^* implicitly (Giordano and Hsu, 1985)

$k_f = 0;\ k_u = 0;\ \beta_0 = \lambda(t)$

for j=1,2,...,n
$f_j = \Phi(j)$

for i=1,...,j-1
j>1; $k_f = k_f + 1$
$f_j = f_j + \Phi(i)u_{k_f}$
end

$v_j = f_j d_j$
$\beta_j = \beta_{j-1} + v_j f_j$
$p_j = \frac{-f_j}{\beta_{j-1}}$
$d_j = \frac{d_j}{\lambda(t)}\frac{\beta_{j-1}}{\beta_j}$

for i=1,...,j-1
j>1; $k_u = k_u + 1$
$w_d = u_{k_u} + K_g(i)p_j$
$K_g(i) = K_g(i) + u_{k_u}v_j$
$u_{k_u} = w_d$
end

end

Notice that, β_n coincides with $\alpha(t)$, the updating of the algorithm gain $K(t)$ is $\frac{K_g}{\beta_n}$.

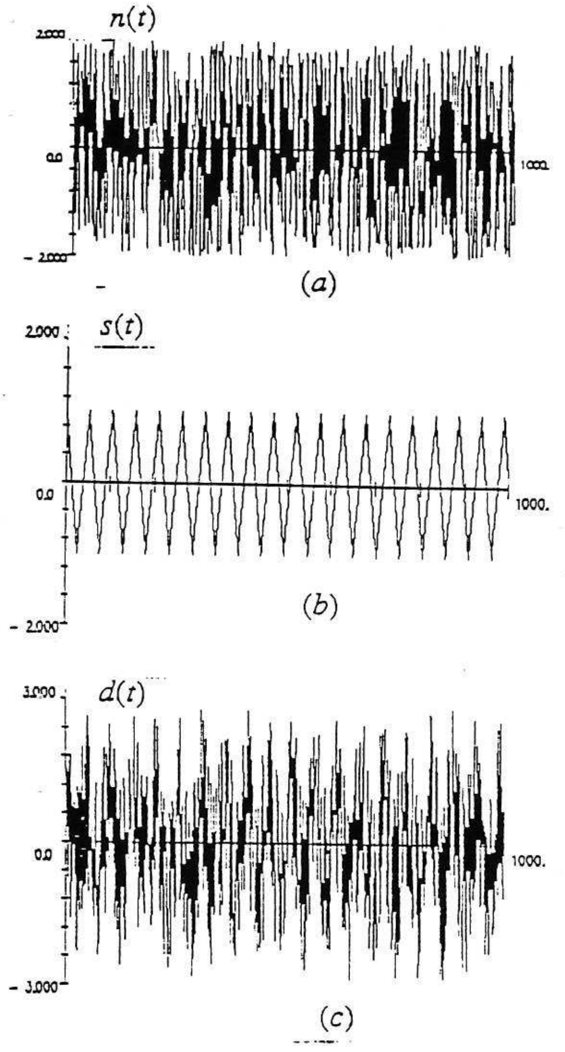

Fig. 4 *Form of the signals*
4(*a*) $n(t)$, 4(*b*) $s(t)$, 4(*c*) $d(t)$

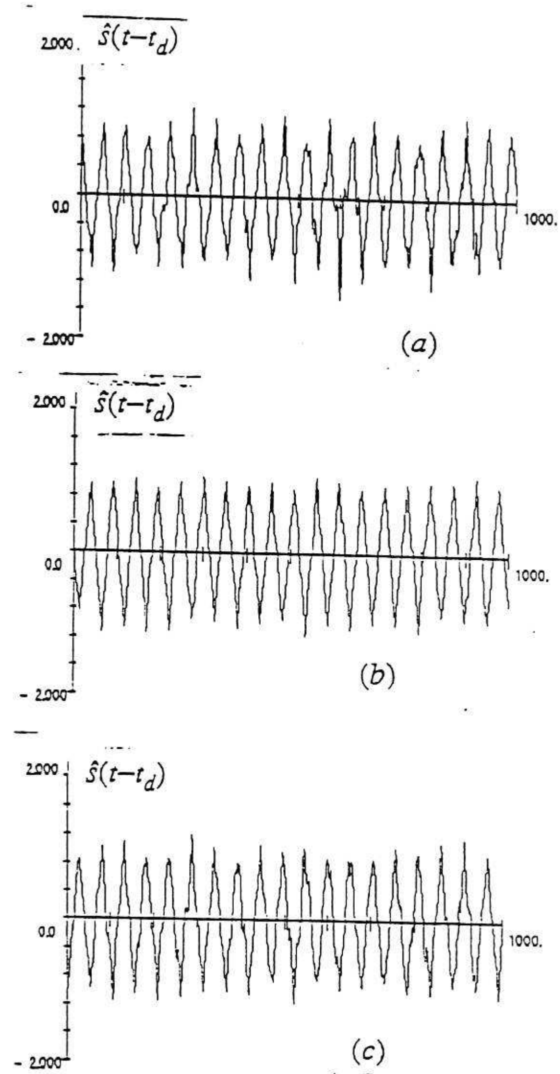

Fig. 5 *Estimated signal taken between iterations* 1000 and 2000 for *simulation S*1
5(*a*) *IIR filter described in eqn.*(46)
5(*b*) *FIR filter with* 40 *weights*
5(*c*) *FIR filter with* 30 *weights*

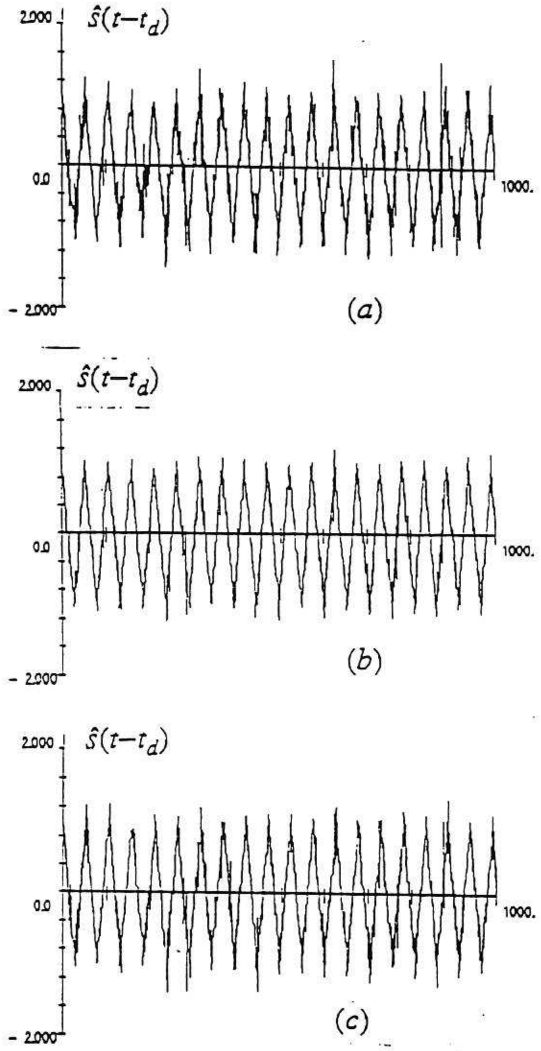

Fig. 6 *Estimated signal taken between iterations* 1000 and 2000 for *simulation S2*

6(*a*) *NFIR filter described in Table II.*

6(*b*) *ARNX filter described in Table II.*

6(*c*) *NIIR filter described in Table II.*

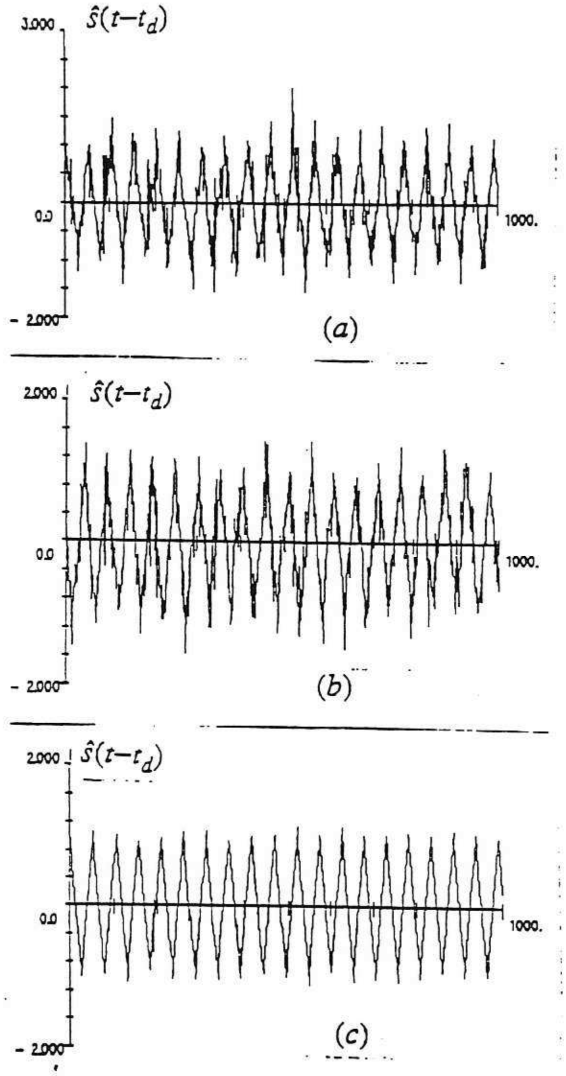

Fig. 7 Estimated signal taken between iterations 1000 and 2000 for *simulation S3*
7(*a*) *NFIR filter described in Table III*(*b*).
7(*b*) *ARNX filter described in Table III*(*b*).
7(*c*) *NIIR filter described in Table III*(*a*),(*b*).