Abstract—In this paper the problem of estimating the time delay between two spatially separated noisy sinusoidal signals by system identification modeling is addressed. The system is assumed to be perturbed by both input and output additive white Gaussian noise. The presence of input noise introduces bias in the time delay estimates. Normally the solution requires a priori knowledge of the input-output noise variance ratio. We utilize the cascade of a self-tuned filter with the time delay estimator, thus making the delay estimates robust to input noise. Simulation results are presented to confirm the superiority of the proposed approach at low input signal-to-noise ratios.

Keywords—LMS algorithm, Self-tuned filter, System identification, Time delay estimation.

I. INTRODUCTION

Time delay estimation (TDE) between noise-corrupted signals incident on two spatially separated sensors is important in various fields such as radar, sonar and geophysics [1]. The LMS algorithm is popularly used for TDE due to its simplicity [2], [3]. However, the LMS algorithm gives biased time delay estimates. Alternatively, the estimation of time delay can be modeled as a system identification problem but the presence of additive input noise brings further complications. In many practical situations the additive noise is unavoidable because of imperfections in the signal acquisition process [4]. Under the abovementioned circumstances, one way of making it possible to track time delays is by assuming a priori knowledge of the input to output noise variance ratio and then proceed to use this information to modify the error function in the LMS algorithm[5]. However, the a priori knowledge of noise variances could be unavailable.

In this paper, we propose an alternative LMS-based TDE approach that relaxes the assumptions about input and output noise for sinusoidal input signals. The proposed method uses a cascade of a self-tuned filter (STF) with the time delay estimator and this structure is hereinafter referred to as the STFTDE for brevity. It is assumed that the increased robustness of the resulting system outweighs the moderate increase in computational complexity associated with the introduction of the STF. Some exploratory results related to this work have been presented in [6].

The paper is organized as follows. Section II describes the proposed method beginning with the preliminaries, followed by the effect of input noise, then MSE and bias analysis, ending the section with delay convergence properties of the STFTE. A simulation example is given in Section III. Finally, Section IV concludes this paper.

II. PROPOSED METHOD

A. Preliminaries

The TDE considers two discrete-time signals incident on two sensors, which are sampled at time \( t = kT_s \), where \( T_s \) is the sampling period (assumed to be unity in this paper for simplicity) and expressed as

\[
x(k) = As(k) + v(k),
\]

\[
y(k) = As(k - D) + n(k),
\]

where \( s(k) \) corresponds to the noise-free source signal, \( s(k - D) \) is its delayed one, and \( A \) is their constant amplitudes. Where not explicitly stated, and without of loss of generality, the value of \( A \) will be assumed to be unity in the sequel. The \( n(k) \) and \( v(k) \) are the uncorrelated zero-mean white Gaussian noise of variance \( \sigma_n^2 \) and \( \sigma_v^2 \), respectively. The variable \( D \) represents the unknown time delay to be estimated, which is approximated to an integer closest to the true delay in the discrete-time model. Subsample delays can be obtained by interpolation, where necessary [8].

![Fig. 1 Modeling time delay estimation as system identification.](image)

Figure 1 depicts the configuration of the proposed method for TDE. The sequences \( \{h_i\} \) and \( \{g_i\} \) are the system...
impulse response and the system identifier (time delay estimator filter) coefficients, respectively. The LMS algorithm is used to update all the filter coefficients. The output $z(k)$ of the STF when excited by input $x(k)$ is given by

$$z(k) = \sum_{i=0}^{M-1} w_i(k)x(k-i - D_i),$$

where $M$ is the length of the STF, $D_i$ is a delay appropriately chosen to decorrelate the input noise, and the $w_i(k)$'s are STF coefficients at time $k$. The error term for adjusting the coefficients of the STF becomes

$$\xi(k) = x(k) - z(k).$$

The desired sequence is given by

$$y(k) = \sum_{i=0}^{M-1} h_i(k)s(k-i) + n(k),$$

where $[h_0(k), \ldots, h_{M-1}(k)]$ is the system impulse response vector. The output of the time delay estimator, $\hat{y}(k)$, can be expressed as

$$\hat{y}(k) = \sum_{i=0}^{M-1} g_i(k)z(k-i),$$

where $[g_0(k), \ldots, g_{M-1}(k)]$ is an estimate of the system impulse response vector and is denoted by $g(k)$. The time delay estimator output error is given by

$$e(k) = y(k) - \hat{y}(k).$$

The time delay estimate at the $k$ th iteration, $\hat{D}(k)$, is obtained as

$$\hat{D}(k) = \arg \max_i \left[g(k)\right],$$

where \([\cdot]_i\) denotes the $i$th element.

**B. Effect of Input Noise**

The bias in the time delay estimates when the time delay estimator is directly excited by $x(k)$ can be obtained as follows. Both sides of ((2)) are first squared, expanded, the expected value taken, differentiated with respect to the $g_i(k)$'s, and finally the result is equated to zero. The process yields

$$\sum_{i=0}^{M-1} g_i(k) = \beta(k) \sum_{i=0}^{M-1} h_i(k),$$

where

$$\beta(k) = \frac{\sigma_i^2}{\sigma_i^2 + \sigma_e^2(k)} = \frac{SNR_i(k)}{SNR_i(k) + 1}$$

and $SNR_i(k) = \frac{\sigma_i^2}{\sigma_i^2 + \sigma_e^2(k)}$. A similar result has also been obtained in [5]. It is clearly seen that the input noise variance introduces the bias in the estimated coefficients through the parameter $\beta(k)$. Since there is a one-to-one correspondence between the system response and the time delay estimator response, we get the relationship

$$g_i(k) = \beta(k)h_i(k).$$

By defining the coefficient bias $B_i(k)$ as $B_i(k) = E[g_i(k)] - h_i(k)$, we can write

$$B_i(k) = [\beta(k) - 1]h_i(k).$$

For high input SNR, that is $SNR_i(k) >> 1$, the $\beta(k) \approx 1$ and $B_i(k) \approx 0$. This means that for all $k$, $g_i(k) = h_i(k)$. In that case, the delay can ideally be obtained in one iteration since only the index corresponding to the maximum of $g(k)$ is required. However, this is not practically possible since the time delay estimator coefficients are determined by the transient behavior of the filter. For very low input SNR, that is $SNR_i(k) << 1$, we can deduce that $\beta(k) \approx SNR_i(k)$, which gives $B_i(k) \approx -h_i(k)$. This means that during the initial adaptation period, the delay estimator coefficients could be very small and indistinguishable. It is, therefore, possible to get a delay estimate that is close to the maximum length of the filter since there could be fluctuations in the estimated coefficient values due to round-off errors and the $\max(\cdot)$ operator simply picks the biggest coefficient.

In the LMSTDE method, the desired output is $d(k) = As(k - D)$, so that we can write $y(k)$ as

$$y(k) = d(k) + n(k).$$

The input to the delay estimator is $x(k)$ and the output error at time $k$ becomes

$$e(k) = y(k) - \hat{y}(k).$$

In ((4)), $\hat{y}(k)$ can be written as
\[ \hat{y}(k) = \sum_{i=0}^{M-1} g_i(k)x(k - i) \]
\[ = \sum_{i=0}^{M-1} g_i(k)s(k - i) + \sum_{i=0}^{M-1} g_i(k)v(k - i) \]
\[ = \hat{d}(k) + v_o(k), \quad (5) \]

where \( \hat{d}(k) = \sum_{i=0}^{M-1} g_i(k)s(k - i) \) denotes the estimate of the desired response, \( d(k) \), to the input signal, and \( v_o(k) = \sum_{i=0}^{M-1} g_i(k)v(k - i) \) denotes the filtered input noise. From ((4)), we can write \( e(k) \) as
\[ e(k) = d(k) + n(k) - \hat{d}(k) - v_o(k) \]
\[ = [d(k) - \hat{d}(k)] + [n(k) - v_o(k)] \]
\[ = e_d(k) + e_{nv}(k), \quad (6) \]

where \( e_d(k) = d(k) - \hat{d}(k) \) is the component of the output error \( e(k) \) due to the difference between the delay estimate \( \hat{D}(k) \) and the true delay \( D \), while \( e_{nv} = n(k) - v_o(k) \) the component due to the presence of noise. Although the steady-state values of the filter coefficients are only affected by the input noise as previously shown, the transient values of the coefficients are affected by both the input and the output noise. We also note that the error component \( e_d(k) \) cannot be controlled by the adaptive filter. This is because \( e_d(k) \) is a determined by filter length. If the filter is of infinite length, then the maximum coefficient can exactly coincide with the true delay, making \( e_d(k) = 0 \). However, infinite-length filters are practically unrealizable. It is, therefore, most likely that \( e_d(k) \) would be nonzero for all \( k \). One way to get around the problem would be to incorporate a separate interpolation algorithm into the adaptation process. This obviously increases the overall computational complexity of the delay estimation process. On the other hand, \( e_{nv}(k) \) can easily be controlled by some form of noise compensation. This can be done either explicitly or implicitly. Explicitly the input noise can be controlled from the input side, while implicitly the input noise can be controlled by modifying the error \( e(k) \) at the output of the delay estimator. In the next subsection, we discuss the MSE and bias of the STFTDE with respect to the input noise.

C. MSE and Bias of the STFTDE

The output \( z(k) \) of the STF when excited by the delayed input \( x(k) \) is given by
\[ z(k) = \sum_{i=0}^{M-1} w_i(k)x(k - i - D_k). \]

The desired output sequence is
\[ y(k) = \sum_{i=0}^{M-1} h_i(k)s(k - i) + n(k). \]

The output of the time delay estimator, \( \hat{y}(k) \), is
\[ \hat{y}(k) = \sum_{j=0}^{M-1} g_j(k)z(k - j) \]
\[ = \sum_{j=0}^{M-1} g_j(k)\sum_{i=0}^{M-1} w_i(k - j)x(k - j - i - D_k) \]
\[ = \sum_{j=0}^{M-1} g_j(k)w_j(k - j)x(k - j - i - D_k) \]

where we substituted for \( z(k - j) = \sum_{i=0}^{M-1} w_i(k - j)x(k - j - i - D_k) \).

By substituting for \( x(k - j - i - D_k) \) in \( \hat{y}(k) \), we have
\[ \hat{y}(k) = \sum_{j=0}^{M-1} \sum_{i=0}^{M-1} g_j(k)w_i(k - j)s(k - j - i - D_k) + \sum_{j=0}^{M-1} g_j(k)w_i(k - j)v(k - j - i - D_k) \]
\[ + \sum_{j=0}^{M-1} g_j(k)w_i(k - j)n(k). \]

The output error becomes
\[ e(k) = y(k) - \hat{y}(k) \]
\[ = \sum_{m=0}^{M-1} h_m(k)s(k - m) + n(k) \]
\[ - \sum_{j=0}^{M-1} g_j(k)w_j(k - j)s(k - j - i - D_k) \]
\[ - \sum_{j=0}^{M-1} g_j(k)w_j(k - j)v(k - j - i - D_k) \]
\[ = \sum_{m=0}^{M-1} h_m(k)s(k - m) \]
\[ - \sum_{j=0}^{M-1} g_j(k)w_j(k - j)s(k - j - i - D_k) \]
\[ - \sum_{j=0}^{M-1} g_j(k)w_j(k - j)v(k - j - i - D_k) + n(k). \]

Squaring ((11)) above results in
\[ e^2(k) = \left[ \sum_{m=0}^{M-1} h_m(k)s(k-m) \right]^2 + n^2(k) \]

\[ + \left[ \sum_{j=0}^{M-1} \sum_{l=0}^{M-1} g_j(k)w_j(k-l)s(k-j-i-D_i) \right]^2 \]

\[ - 2 \sum_{m=0}^{M-1} \sum_{j=0}^{M-1} h_m(k)g_j(k)w_j(k-l)s(k-j-i-D_i) \]

\[ + \text{crossterms}. \]  

By taking expectation of the squared error, we can approximate the MSE to

\[ E[e^2(k)] \approx \sigma_s^2 \sum_{m=0}^{M-1} h_m^2(k) + \sigma_s^2 \sum_{j=0}^{M-1} \sum_{l=0}^{M-1} g_j^2(k)w_j^2(k-l) \]

\[ + \sigma_s^2(k) + \sigma_s^2(k) \sum_{j=0}^{M-1} \sum_{l=0}^{M-1} g_j(k)w_j(k-l) \]

\[ - 2\sigma_s \sum_{m=0}^{M-1} \sum_{j=0}^{M-1} h_m(k)g_j(k)w_j(k-l). \]  

The minimum MSE can be obtained by differentiating \( E[e^2(k)] \) with respect to the \( g_j(k) \)'s and equating to zero to give

\[ \left( \sigma_s^2 + \sigma_s^2(k) \right) \sum_{j=0}^{M-1} \sum_{l=0}^{M-1} g_j(k)w_j(k-l) \]

\[ = \sigma_s^2 \sum_{m=0}^{M-1} h_m(k)w_i(k-l) \]  

or

\[ \sum_{j=0}^{M-1} g_j(k) = \frac{\sigma_s^2}{\sigma_s^2 + \sigma_s^2(k)} \sum_{m=0}^{M-1} h_m(k) \]  

which is similar to ((4)). However, by using the Wiener filter solution of the STF for sinusoidal input signals [9], we can approximate the expected value of \( z(k) \) as

\[ E[z(k)] \approx \frac{\text{SNR}_c(k)(M-1)/2}{1 + \text{SNR}_c(k)(M-1)/2} s(k) + v_f(k) \]

\[ = \alpha(k)s(k) + v_f(k) \]  

where \( \alpha(k) = \frac{\text{SNR}_c(k)(M-1)/2}{1 + \text{SNR}_c(k)(M-1)/2} \) and \( v_f(k) \) is the STF output noise with variance \( \mu \sigma_s^2(k)(M-1) \). It is evident from ((16)) that the STF can greatly reduce input noise effect for \( \mu \). Substituting for \( z(k) \) by its expected value in \( e(k) \) we can simplify the output error function to

\[ e(k) = \sum_{i=0}^{M-1} h_i(k)s(k-i) + n(k) \]

\[ = \sum_{i=0}^{M-1} g_i(k)[\alpha(k-i)s(k-i) - v_f(k-i)] \]

\[ = \sum_{i=0}^{M-1} h_i(k)s(k-i) - \sum_{i=0}^{M-1} \alpha(k-i)g_i(k)s(k-i) \]

\[ + n(k) - \sum_{i=0}^{M-1} g_i(k)v_f(k-i). \]  

Squaring both sides of ((17)) and taking expectations gives the MSE as

\[ E[e^2(k)] = \sigma_s^2 \sum_{i=0}^{M-1} h_i^2(k) + \sigma_s^2 \sum_{i=0}^{M-1} \alpha^2(k-i)g_i^2(k) \]

\[ - 2\sigma_s \sum_{i=0}^{M-1} \alpha(k-i)h_i(k) + \sigma_s^2(k) \]

\[ + \sigma_v^2 \sum_{i=0}^{M-1} g_i^2(k). \]  

Differentiating \( E[e^2(k)] \) wrt the \( g_i(k) \)'s and equating to zero to obtain the minimum MSE reduces the output error function to

\[ - 2\sigma_s \sum_{i=0}^{M-1} \alpha(k-i)h_i(k) + 2\sigma_s^2 \sum_{i=0}^{M-1} \alpha^2(k-i)g_i(k) \]

\[ + 2\sigma_v^2 \sum_{i=0}^{M-1} g_i^2(k) = 0. \]  

If \( \alpha(k) \) is taken to be constant and equal to \( \alpha_c \) we can write ((19)) as

\[ - 2\sigma_s^2 \alpha_c \sum_{i=0}^{M-1} h_i(k) + 2[\sigma_s^2 \alpha_c^2 + \sigma_v^2] \sum_{i=0}^{M-1} g_i(k) = 0 \]  

or

\[ - 2\sigma_s^2 \alpha_c \sum_{i=0}^{M-1} h_i(k) + 2[\sigma_s^2 \alpha_c^2 + \sigma_v^2] \sum_{i=0}^{M-1} g_i(k) = 0 \]  

(20)
When $\alpha_c$ is approaches 1, the effect of input noise is less severe and could give less biased coefficient estimates when the step-size is appropriately chosen to be small.

D. Delay Convergence Properties

The time delay estimates converge to their optimum values much earlier in the adaptation process compared to adaptive time constant. This can be explained as follows. From the transient analysis of the LMS algorithm, it is known [10] that the adaptive time constant $\tau_a$ lies in the range

$$-\frac{1}{\ln(1-\mu\lambda_{\text{max}})} \leq \tau_a \leq -\frac{1}{\ln(1-\mu\lambda_{\text{min}})},$$

where $\mu$ is the step-size parameter, and $\lambda_{\text{max}}$ and $\lambda_{\text{min}}$ are the maximum and minimum eigenvalues of input autocorrelation matrix $R_{xx}(m) = E[x(k)x(k+m)]$, respectively. However, for the TDE problem, we are mainly interested in the maximum filter coefficient value among $g_0(k) \ldots g_{M-1}(k)$. Therefore, only the lower bound of adaptive time constant is of interest. The time delay adaptive time constant, denoted by $\tau_d$, becomes

$$\tau_d = -\frac{1}{\ln(1-\mu\lambda_{\text{max}})}.$$  \(22\)

For a real sinusoidal input signal corrupted by white noise as given by ((1)), the maximum eigenvalue $\lambda_{\text{max}}$ is $A_0^2/2 + \sigma_v^2$, and substituting it in ((22)) gives

$$\tau_d = -\frac{1}{\ln(1-\mu(A_0^2/2 + \sigma_v^2))},$$

which could be much less than the adaptive constant $\tau_a$ as observed in [3]. Furthermore, we can see that the effect of input noise is to slow the convergence to the true delay by a factor $\eta$ equal to

$$\eta = \frac{\ln(1-\mu(A_0^2/2)}{\ln[1-\mu(A_0^2/2 + \sigma_v^2)]}.$$  \(21\)

Thus, the STF could improve convergence and also reduce bias in the time delay estimates. Moreover, a small residual error will always exist between $\epsilon(k)$ and the input noise. Thus, the stability of the time delay estimator is guaranteed.

One desirable characteristic for the STF is that only the Wiener filter solution is required for good performance [9]. The optimal Wiener filter coefficients for a narrowband sinusoidal signal of amplitude $A_0$, center radian-frequency $\omega_0$, and phase angle $\theta$, corrupted by broadband white Gaussian noise, have the form

$$w_{wi} = \frac{A_0^2 \cos(\omega_0(i + D_1))}{2\sigma_v^2(1 + \frac{M^2A_0^2}{4\sigma_v^2})}. $$

The expected value of $z(k)$ is given by

$$E[z(k)] = \sum_{i=0}^{M-1} w_{wi} A_0 \sin(\omega_0(k - D_1 - i) + \theta).$$

By substituting $w_{wi}$ in $E[z(k)]$ and assuming that $\omega_0 M$ is not close to zero or $\pi$, we get

$$E[z(k)] \approx \frac{\text{SNR}M/2}{1 + \text{SNR}M/2} A_0 \sin(\omega_0 k + \theta).$$

Thus, it is verified that $E[z(k)]$ becomes independent of $D_1$.

III. SIMULATION EXAMPLE

Computer simulation was carried out to evaluate the performance of the proposed method. The source signal $s(k)$ consisted of a single sinusoid and the input signal $x(k)$ was corrupted by white Gaussian noise $\nu(k)$ as given by

$$x(k) = A_0 \cos(2\pi f_0 k + \nu(k),$$

where the frequency of the input signal $f_0$ was set to 120 Hz per sample and the amplitude $A_0$ was set to unity. The signal was sampled at 1 kHz. The corresponding output signal $y(k)$ was generated according to the relationship ((2)). In the simulation, the ratio of the input to output noise variance was assumed to be 1 for the So method [5] while no assumptions were made for both the proposed method and the LMSTDE method [3]. For all methods, a fixed step-size of 0.001 was used and 100 independent runs were averaged. The filter
lengths for the proposed method were set to $M = 16$, and $D_1$ was set to 1.

Figure 2 shows the delay convergence properties of the simulated methods. The effectiveness of the proposed method can clearly be seen where the proposed method converges to the true delay of $3 \times T_s$ much faster than the other two methods. To see the effect of sudden changes in delay, the time delay was changed at the 200th iteration from $3 \times T_s$ to $4 \times T_s$. Figure 3 shows the performance of the proposed method compared with the two other methods. The results show that the proposed method performs much better for sudden changes in delay by adapting to the new delay faster than the other methods. We also conducted an investigation into the tracking ability of the algorithm when the noise variance suddenly changed from 10dB to 0dB at the 500th iteration. The results show superior tracking ability of the proposed method as depicted in Fig. 4.

IV. CONCLUSION

We have proposed a novel TDE method based on noisy input-output system identification modeling. The attractiveness is rapid convergence and noise robustness when compared to the other competing methods. An analysis of MSE and bias properties of the proposed method has been given. We have also shown that only the maximum eigenvalue of the input correlation matrix is important in speeding-up the convergence to true time delays. Simulation results show that the proposed method performs better than existing methods in its tracking ability for both step time-variant delays and input noise. Nevertheless, further improvements with respect to convergence speed and fractional delay adaptation are possible. These have been deferred to our future work.

ACKNOWLEDGMENT

The author thanks the reviewers for their valuable comments and suggestion.

REFERENCES


J. Gamba received the BSc.(Hons) degree in electrical and electronics engineering from the University of Zimbabwe, Harare, Zimbabwe, in 1994, the MSc. degree in computer science and engineering from Zhejiang University, Hangzhou, China, in 2000 and the PhD in Mathematical Information Science from Saitama University, Japan, in 2005. He is currently with Tsukuba University as post-doctoral researcher of multimedia information systems applications. His present research interests include time delay estimation, adaptive channel equalization and image restoration. He is a member of IEEE and IEICE.