Experimental Evaluation of Iterative Learning Control Algorithms for Non-Minimum Phase Plants

Chris Freeman, Paul Lewin and Eric Rogers

School of Electronics and Computer Science,

University of Southampton, UK, SO17 1BJ

cf@ecs.soton.ac.uk

Abstract

The purpose of this paper is two-fold, firstly it describes the development and modelling of an experimental test facility as a platform on which to assess the performance of Iterative Learning Control (ILC) schemes. This facility includes a non-minimum phase component. Secondly, P-Type, D-Type and phase-lead types of the algorithm have been implemented on the test-bed, results are presented for each method and their performance is compared. Although all the ILC strategies tested experience eventual divergence when applied to a non-minimum phase system, it is found that there is an optimum phase-lead ILC design that maximizes convergence and minimizes error. A general method of arriving at this phase-lead from knowledge of the plant model is described. A

variety of filters have been applied and assessed in order to improve the overall performance of the algorithm.

1 Introduction

Iterative Learning Control (ILC) is a control method that is applicable to systems which perform the same action repeatedly. Operating in this way it is able to use past control information such as input signals and tracking errors in the construction of the present control action. This sets ILC apart from most other control techniques and has allowed it to provide improved performance with reduced knowledge of the plant when compared with other control approaches. Practical testing of ILC algorithms has generally been performed in order to validate a single algorithm as part of a theoretical development. Exceptions to this occur in publications reporting experimental work, but these often do not justify the choice of algorithm used, or are in such specific areas as to preclude the use of more general ILC methods. It is the aim of this research to investigate and critically compare using experimental data a variety of ILC algorithms. This paper deals with the simplest methods that, as well as providing a reference point for more advanced algorithms, play an important role in ensuring the widespread acceptance of this approach to control systems design.

One of the first ILC algorithms was proposed by Arimoto, Miyazaki and Kawamura [1], and consisted of a correction term comprising of some measure of the error from the same instant in the previous trial.

$$u_{k+1}(t) = u_k(t) + \Gamma e_k(t) \tag{1}$$

Where Γ is the learning gain, $u_k(t)$ is the control signal at the k^{th} iteration while $e_k(t) = y_d(t) - y_k(t)$ is the tracking error. Here, time $t \in [0, T]$, where T is known and finite. Another of the original continuous-time algorithms considered was the so-called D-type Algorithm proposed by the same authors [2], in which the error derivative is used.

$$u_{k+1}(t) = u_k(t) + \Gamma \dot{e}_k(t) \tag{2}$$

Much work has been done extending the proof to various types of system, producing a discrete version, and establishing bounds on the gain for convergence. There has followed a process of extending and augmenting the structure of these early algorithms. Examples of such work include the use of more than one previous cycle and also higher derivatives as seen in [3, 4, 5]. The former increases the robustness, as defined by these authors, at the price of convergence speed. The latter is equivalent to a higher-order approximation of the plant inverse, and this naturally creates great difficulty in selecting the gains and reduces robustness. It should, however, reduce the final error bound.

Feedback controllers have been found to be extremely useful in stabilizing the plant during the process of ILC, and have been included in a variety of positions within the control structure. Current cycle information achieves the same effect, but is integral to the update algorithm, as discussed in, for example [6, 7, 8]. Additions to the basic ILC setup have been parameter estimators and filters, as well as many modifications to deal with time-delays, initialization error, and uncertainty. Work has also been done to summarise some of the material on ILC and produce practical guidelines intended to make the simpler laws easy to use on a wide range of plants [9].

The following section describes the non-minimum phase experimental test-bed

and approach to system modelling. Section 3 focuses on the feedback controller. Sections 4 and 5 evaluate the performance of P-type and D-type algorithms respectively, while section 6 proposes and assesses phase-lead ILC. Sections 7 and 8 use both causal and non-causal filters to improve the algorithm's performance. Conclusions and further work are given in section 9.

2 Experimental design

The experimental apparatus was chosen to be non-minimum phase since this characteristic has presented difficulties throughout the history of ILC. Advanced techniques that are proven to handle this type of system, or are formulated especially for it, exist but are generally complex. To fully justify their use, the practical failure of simple algorithms must be established and this is an area which has received little attention.

The plant has been designed to be linear time invariant (LTI) so as to increase the range of applicable algorithms; it is expected that a certain amount of non-linear behaviour will arise in the real world - enough for appreciable use by non-linear algorithms. There will inevitably be measurement noise and disturbances which will benefit from robust approaches, and certain other control challenges will be afforded by the motor's and inverter's characteristics. The non-minimum phase characteristic was obtained by means of an electrical analogue which could be realized mechanically with just an inertia, a damper, a torsional spring, a timing belt, pulleys and gears. This is shown schematically in Figure 1, in which the gearing on the right reverses the direction of rotation whilst the timing belt on the left maintains it. Two spring-mass-damper sys-

tems have also been constructed which can be inserted before the non-minimum phase component in order to increase the relative degree and system complexity for future ILC research, but are not used in the work described here.

Component values for the inertias J and J_g, the gearing G, the spring K and damper B were chosen using simulations to produce a relatively stable system with a large amount of error when following a demand. The entire system is shown in Figure 2 with the non-minimum phase section located in the upper left corner of the test-bed.

Two 1000 pulse per revolution encoders record position at the motor shaft and the system output. A DEVA 004 motion control card processes this information and increases their resolution to the equivalent of 4000 pulses per revolution.

A standard squirrel cage induction motor supplied by an inverter, operating in variable voltage variable frequency (VVVF) mode, drives the load. A PC is used to control the system and the software so far developed is capable of implementing a large number of ILC schemes within a single executable program. The package that has been developed uses a Graphical User Interface (GUI) similar to other mathematical programming environments, and uses specially designed ILC subsystems that use traditional ILC notation, each being capable of using every present and past signal. The system is sampled at 2.5 KHz.

Two approaches have been used to model the system. The first is a time-based simulation approach which involves deriving theoretical expressions for groups of components which are then verified against experimental data. The resulting simulation is then used to model the effect of algorithms before they are applied to the system. The second, frequency-domain model, is less accurate but can be used with classical techniques to derive and analyse ILC algorithms. A linear

model was fitted to the Bode plot obtained by frequency tests on the plant. The plant transfer function of the linear model is

$$G_{stage1}(s) = \frac{123.853 \times 10^4 (3.5 - s)}{(s^2 + 6.5s + 42.25)(s + 45)(s + 190)} \qquad (rads^{-1}v^{-1}) \qquad (3)$$

The experimental test facility has been used to evaluate the performance of three simple structure ILC algorithms. The results obtained are presented and discussed in the remainder of this paper.

3 PID Tuning

A feedback controller may be used in order to stabilize the plant prior to the use of ILC. For this work a PID controller was implemented and initially tuned using the Zeigler-Nichols method but this was found not to produce satisfactory results, especially when using rapidly changing demands. The tuning procedure was therefore conducted experimentally using a program that ran a given demand and recorded the error using a cost function, J_k, that consisted of the sum of the modulus of the error at every sample instant, i.e

$$J_k = \sum_{i=1}^{N} |y_d(i) - y_k(i)| \qquad (4)$$

Where N is the number of samples in a single trial. Minimising this with respect to the PID gains produced extremely oscillatory results and therefore the cost function was amended to

$$J_k = \sum_{i=1}^{N} |y_d(i + \gamma) - y_k(i)| \qquad \gamma = 1, 2, \ldots N - i \qquad (5)$$

This permits the demand to be shifted relative to the output and therefore the process of minimization ensures that the output is tuned to follow the shape of

the demand, whatever time delay that incurs. This was tuned for a sinewave and two repeating sequence demands using three different unit rates. The slowest version of each demand is shown in Figure 3. In the remainder of this paper the repeating sequence demands b) and c) will be referred to as R1 and R2 respectively.

4 P-Type ILC

The algorithm is given in its discrete form at sample i by Equation 6, where a sample delay is required to counter the one time step delay through a differential equation when fed by a zero order hold. Since u_k is initially equal to the demand y_d, this equation can be interpreted as the demand for trial $k+1$ being made from the original demand plus the integral of all the errors up to and including trial k.

$$\begin{aligned} u_{k+1}(i) &= u_k(i) + \Gamma e_k(i+1) \\ e_k(i) &= y_d(i) - y_k(i) \end{aligned} \tag{6}$$

Normalised Error (NE) has been used as a measure of how well the demand is followed. This is calculated according to Equation 4 for each trial but this value is then divided by the same equation with $y_k(i) = 0$ so that the NE counteracts the misleadingly large error of a long demand profile. The demand profile is positional but the tests are stopped when the output velocity makes it unsafe to continue.

Figure 4 shows the normalized error against the number of trials, for a sinewave demand at 10 Units Per Minute (UPM). The initial error is very large but is gradually reduced as the number of trials increases, thereby demonstrating

the benefit of well-designed ILC. As expected, divergence occurs as the trials continue. Reducing the gain Γ leads to a slower rate of convergence but it also takes far longer before divergent performance becomes significant. Similar results were obtained for higher unit rates and the repeating sequence inputs. Three important influences on performance were observed:

- As the unit rate is increased the number of trials until instability (T_{ins}) decreases and the minimum error (NE_{min}) increases
- For repeating sequences and high unit rates T_{ins} decreases and NE_{min} increases
- After a certain value is reached, the effect of further decreasing the gain does not result in any further decrease in NE_{min}

Figure 5 shows results from the trial which performed the best; a sinewave at 10 UPM. The output before learning, 'PID output', is included for comparison and is shown to significantly lag the demand. The updated demand is the value of u_k for the cycle shown. Use of P-Type ILC removes the lag of the output but the original demand is not followed well. The updated demand is very oscillatory. Figure 6 shows how the error evolves for a 10 UPM sinewave input by showing the output at every 10^{th} trial. Instability is manifested by large oscillations in the output. This is characteristic of all the results seen.

5 D-Type ILC

The discrete D-Type algorithm is given by Equation 7, in which T is the sampling period.

$$u_{k+1}(i) = u_k(i) + \Gamma(e_k(i+1) - e_k(i))/T \tag{7}$$

Figure 7 shows how the cycle error changes as the trial number increases. Similar results were again obtained for increased unit rates and for the repeating sequence demands. The same observations as noted for P-Type concerning changing the gain, unit rate and demand are again relevant. Comparing D-Type to P-Type ILC also reveals some important features:

- The value of T_{ins} is significantly reduced for all demands, especially those at high unit rates
- The value of E_{min} is slightly reduced for all demands, more so for repeating sequences and higher unit rates

Therefore D-Type ILC is found to improve the error at the expense of the stability. Figure 8 shows the signals from the trial with the optimum performance, again with the PID output (equal to the output at trial 0) shown as a reference. The output oscillates around the demand, but at the expense of a highly oscillatory updated demand. The demand also suffers from a large amount of noise due to the differentiation of the error. Figure 9 shows how the output becomes unstable as the trial number increases due to increasingly high amplitude oscillations. This is a feature common to all the results recorded.

6 Phase-lead ILC

Both P-Type and D-Type algorithms use some measure of the error signal in the previous trial at one sample instant ahead of the correction. Although several algorithms have been proposed which use some measure of the error from a greater number of sample instants ahead, they generally use only a very small number of samples [10], or enough samples to remove an explicit delay in a

system prior to implementing an ILC controller as in [11]. There has been some discussion relating to the control law shown in Equation 8 with regards to the phase-lead, λ, which provides the best performance [9, 12].

$$u_{k+1}(i) = u_k(i) + \Gamma e_k(i + \lambda) \tag{8}$$

Practical work [13] has been conducted in which λ was assumed to be the number of samples which, when shifting the demand forward, minimized the difference between the demand and the system output, this type of ILC is referred to as Delay-Type ILC in this paper.

Figure 10 shows error results for a sinewave demand at 10 UPM. Instead of different gains, the graph shows a variety of phase leads. When $\lambda = 0$, phase-lead ILC equates to P-Type ILC. At another phase-lead it equates to Delay-Type ILC, for the 10 UPM demand this delay equals 3100 samples. Figure 10 shows that this is not the optimum delay and furthermore, for the repeating sequence and variety of unit rates the delay calculated, is never found to be optimum (even though it varies from 2225 - 3635 samples depending on the demand). The experiments comprising Figure 10 were conducted over 400 cycles. Figure 11 helps illustrate the way in which more challenging demands effect the system performance. The choice of phase lead required for convergence is narrowed and T_{ins} severely reduced.

The results from all the demands, each for several choices of gain, show that certain features are consistently true:

- Phase-lead ILC outperforms P-Type and D-Type by over 20 times in terms of NE_{min}
- The optimum phase-lead does not change when the demand is altered and

is approximately 1500 samples

- The effect of changing the gain, unit rate and demand is consistent with comments made previously

In order to explain this lead, it is necessary to consider the typical impulse response of various systems, as detailed in Figure 12.

The first shows a generic first order response, the second a higher order response, and the third is the response of the non-minimum phase system considered in this paper. If the response were to be so simplified as to be, itself, a single impulse, they would occur at 0, m and n seconds for a), b) and c) respectively. Therefore the most accurate single impulse model of the inverse of these systems occurs at the times 0, m and n seconds before the output. Although this is an imprecise inversion, it approximates the method by which the simple structure algorithms function. P-Type ILC works well on first order systems because they have the property that the error at sample i is most directly due to the input at the same instant. The success of phase-lead ILC is therefore evident; if the time taken for the maximum impulse response peak can be found and used as λ in the phase-lead law, then it should be as successful as P-Type is for first order systems. Unfortunately this is not the case. The value n for the non-minimum phase system is found to be 1950 samples using the system model, well above the experimentally achieved optimum of 1500. When modelling phase-lead ILC on higher order systems, the optimum in terms of both convergence speed and minimum error has also consistently been found to be slightly below the value of m. The minimum error is also never zero, and divergence always occurs.

Choosing the phase-lead in accordance with the maximum impulse response value is a simplification of a more general update: the case in which corrections

are made at all sampling instants before the error, and the amount of correction dictated by the magnitude of an impulse response produced by the correction point at the error. In order to remove an error, corrections are made at every point that could possibly have been responsible for the error. The correction is equal to the error multiplied by a gain multiplied by this 'measure of responsibility' for the error. If only the single 'most responsible' of data points is considered, this method is the same as the phase-lead law described. This general method effectively reflects the impulse response in the time axis, and uses it to approximate the impulse response of the plant inverse. This is of course fundamentally flawed as the impulse that would be generated at the correction point is not an impulse at all.

Figure 13 shows data recorded during the best performing cycle of phase-lead ILC. The demand is followed closely, although the updated demand is quite oscillatory.

Figure 14 shows how the output signal changes as the number of trials increase up until the maximum convergence of the test. Data from other experiments shows that oscillations go on to grow in the updated demand and in the output until their velocity becomes too great for the testbed.

The failure of phase-lead ILC to converge to zero and remain there can, however, be explained. Figure 15 illustrates the failure mechanism that occurs when using phase-lead ILC; oscillations of a certain frequency grow gradually until they force the output position, and hence velocity, to become unmanageable. Analysis of results using different gains and phase-leads yields the following conjecture:

- The frequency of the destabilizing oscillations (f) is only dependent on the phase-lead used, and can be estimated using

$$\frac{1}{f}(\angle f - 180) = \frac{\lambda}{f_s} \tag{9}$$

where $\angle f$ is the phase-lag at f and f_s the sampling frequency. This states that f is the lowest frequency that can be propagated by the phase lead λ, and Figure 16 illustrates how this occurs.

If q represents an instant of an oscillatory new demand input, it will directly affect the value r of the output (with some gain change). If r is larger than the originally specified demand then, by the nature of phase-lead ILC (with lead λ), q will be made increasingly negative. This only succeeds in increasing r in the next trial. The growth of the oscillations in the updated demand is a function of the gain, Γ, and the magnitude of the gain at f.

Figure 17 shows how the phase-lead oscillations can be predicted from the Bode plot of the system. Equation 9 is plotted for a range of phase-leads and their intersections with the phase plot show the frequencies of instability that would arise. Since, for the system considered here, instability is caused by the output velocity, then the gain plot of $sG(s)$ should be examined instead of the gain plot of $G(s)$. This shows that as the phase-lead reduces from 2500 to 1250 samples, the gain of the velocity decreases from 1.5 to 0.63. This explains why the optimum lead is reduced from 1950 to 1500 samples; there is a compromise between the rate of learning and the rate of increase in the magnitude of oscillations caused by phase-lead ILC. Figure 18 shows the inability of phase-lead ILC in coping with rapidly changing demands, a shortcoming which motivates the use of the filters in the following sections. Even with the optimum phase-lead, the R2 demand cannot be followed accurately for very many trials. Whilst instability can occur rapidly, Figure 19 shows how closely the output matches the 20 UPM R2 demand before the previously described oscillations cause instability.

7 Causal Filters

A filter can either be applied to the error, seen in Equation 6, or to the input of the plant, u_{k+1}. The only difference is whether the demand itself is filtered, the effect of which will later be investigated. The open-loop system will therefore be considered to be the plant $G(s)$ in series with the filter, $F(s)$. The simplest way to reduce the destabilizing oscillations that have been observed is to use a causal low-pass filter to reduce the magnitude of the Bode plot of $F(s)G(s)$ at the frequency of oscillation. The act of adding a causal filter to the plant, however, changes the phase plot of the system and therefore the frequency at which a given phase-lead intersects with it. Furthermore, it is likely that the impulse response of the system will change also. It is therefore an iterative process to design a causal filter for use with phase-lead ILC. Firstly a cut-off is selected below the frequency of unstable oscillations, and a class of filter to implement it. The usual criteria of a sharp cut-off and minimal phase-lag are favorable, although, as yet it is not clear as to their relative importance. Little emphasis has been placed on ripple in the stop-band. The impulse response of $F(s)G(s)$ is then obtained and the number of samples to its maximum determined. A Bode plot of $F(s)G(s)$ is drawn together with a line representing those frequencies that can be propagated by a phase-lead of the number of samples calculated, in the same manner as that shown in Figure 17. The frequency of unstable oscillations is found by the intersection of this line with the phase plot. This frequency should correspond to the local minima seen on the magnitude plot caused by the filter. This ensures that no undue magnitude (and hence bandwidth) has been sacrificed below the unstable frequency. For a given filter it also ensures that no additional low frequency lag has been added other than what is necessary

according to the filter chosen. If no such correspondence occurs, the cut-off must either be moved slightly, or the filter order changed in order to produce more lag and the design process repeated.

Three causal filters have been designed and tested on the system. The first is a 5^{th} order Chebychev lowpass filter with a cut-off of 2.5Hz, and 270° phase-lag and 60dB attenuation at the unstable frequency. This filter is a compromise between sharpness of cut-off and lag. The second filter is more aggressive with an extra 90° lag but an additional 22dB attenuation centered on the unstable frequency. The third filter is a 4^{th} order Butterworth bandstop filter which has been selected for its high attenuation over a very small range of frequencies. The attenuation of 60dB is centered on the unstable frequency at a cost of just 90° lag beforehand. This filter was designed in order to maximize the system bandwidth whilst still reducing the effect of the unstable frequency.

Figure 20 shows an unexpected and illuminating effect observed when using the bandstop filter; there appears to be at least two higher frequencies than the unstable frequency which also progressively increase in magnitude as the cycle number increases. The figure suggests, and a frequency analysis confirms it, that along with the unstable frequency identified using Equation 9 and a Bode plot of $F(s)G(s)$, there are two other unstable frequencies. The unstable frequencies are 2.3, 2.65, and 5.2Hz. These can readily be explained if the cause of the original unstable oscillation (the largest frequency that can be propagated given the time of the phase-lead) be extended to include all frequencies that can be propagated. Equation 9 can then be rewritten as

$$\frac{1}{f_i}(\angle f_i - 180(1+2i)) = \frac{\lambda}{f_s} \qquad i = 0, 1, 2, \ldots \tag{10}$$

where again f_i is the i^{th} frequency of oscillation, $\angle f_i$ is the phase lag at f_i, and f_s the sampling frequency. The first three instability frequencies, f_1, f_2 and f_3, will be referred to as the primary, secondary, and tertiary frequencies.

Figure 21 shows the Bode plot of $F(s)G(s)$ using the bandstop filter with the first three phase-lead lines, generated using Equation 10. The gains at the frequencies of intersection are highlighted on the magnitude plot for clarity.

The primary, secondary and tertiary frequencies are found to be 2.3, 2.68, and 4.8Hz respectively, closely matching those experimental values observed. The reason for the prominence of these first three unstable frequencies can be seen from the gain plot; the magnitudes that correspond to these frequencies are all similar and close to -40dB. It is because the bandstop filter reduces the primary frequency alone to such a degree that the secondary and tertiary are so visible. Frequencies higher than the tertiary have been rarely observed due to their high attenuation.

Further tests have shown that altering the demand profile used does not alter the findings by any great degree. If the demand contains a sizable component of one or more of the unstable frequencies then instability progresses sooner, the updated demand containing components that would have otherwise taken many cycles to build up. Looking at the frequency components present in those demand used, shown in Figure 22, it is clear that there are only very small quantities of these frequencies are present in the demands.

Although only the 20 UPM demands are shown, the 15 and 10 UPM cases are obtained by multiplying the frequency scale by 0.75 and 0.5 respectively. Because these frequencies are close to the primary frequency, the filter is best located at the input to the plant. Results have confirmed that performance is

far worse if the demand is left unfiltered. The exceptions to this rule occur when using non-causal FIR filtering and are discussed in section 8.

The theoretically best lead using the bandstop filter (found from the impulse response) is 2250 samples. As with the unfiltered case, this is reduced when carried out in practice due to the higher attenuation of the unstable frequencies (the phase-lead lines in Figure 21 move to the right), and becomes 1750. Two extra objectives can now be put forward in order to improve the design of future causal filters:

- To ensure that the best possible phase-lead, derived from the impulse response, is as close as possible to that experimentally determined
- To seek to move the intersection of the $F(s)G(s)$ phase plot and the optimum phase-lead line further towards the right and thus at a higher frequency

The former task involves reducing the magnitude at the fundamental frequency (and beyond) sufficiently to allow the convergence and stability advantages of using the most favorable phase-lead to become more important than the extra attenuation gained by increasing it. Until this is true instability will always govern the process. The second task depends on the first; extra lag produced by a high-order causal filter causes a given phase-lead to give rise to unstable frequencies which are slightly higher, and therefore more attenuated, than otherwise. Unfortunately a system with more low frequency lag will usually have an impulse response with a larger number of samples to its maximum value. The bandstop filter raised the fundamental frequency at optimum lead from 1.66Hz to 1.8Hz, and the two lowpass filters both raise it to 2.3Hz, which helps

to account for their success. It should also not be forgotten, however, that a surfeit of lag before the cut-off point will destabilize the system. The two lowpass filters were designed with these points in mind. The first has a magnitude plot very similar to the bandpass filter, enabling performance comparisons to be made in terms of lag and attenuation above the cut-off alone. The second lowpass filter is similar to the first but with more lag and more attenuation, enabling comparisons with the first to be made on that basis only.

Figure 23 illustrates the shortcomings of the bandstop filter. Its lower attenuation of the instability frequencies, especially the secondary and tertiary, causes instability. As discussed, it also has lower instability frequencies than the other two filters. The superior performance of the second lowpass filter shows that frequency attenuation is more important than low frequency lag. Instability frequencies only account for a certain amount in explaining the lack of convergence, and it is has been found that the removal of frequencies below the primary improves convergence. This is due to two factors;

- The influence of the primary frequency extends a certain amount below that frequency where, instead of continually growing, it merely disrupts learning

- High frequencies naturally destabilize the process of learning, more so if they are present in the demand. This makes intuitive sense since all the ILC algorithms seen in this paper are effectively built on the notion of a heavily simplified plant. Attenuating increasingly low frequencies in the plant is a method of simplifying it. The simplified plant then more closely matches that required by the ILC algorithm, and learning is improved.

Without being able to substantially change the unstable frequencies it is impossible to separate these two factors. The emphasis for the need of a precise cut-off is also diminished, the only certain requirement being a large amount of attenuation at the unstable frequencies.

Figure 24 shows the output of the plant during a very unsatisfactory period of learning. No unstable frequencies are seen and the phase-lead is the experimental optimum, although this has only been found to a resolution of 125 samples. From this and other similar cases it is clear that higher frequencies than those present in the demand disrupt the learning process, and it may be beneficial to select a cut-off frequency only marginally above the highest frequency present in the demand. Since the approximation to the actual plant that exists at the heart of these simple ILC schemes is most accurate at low frequencies, it is likely that these techniques are only capable of learning low frequencies. As the trial number grows, either the integration of the error at each sample caused by this ILC inaccuracy causes instability, or the instability frequencies overcome the attenuation which has been imposed on them to cause instability themselves.

Having focused on the inadequacies of phase-lead ILC, Figure 25 shows the success of the causal filters that have been implemented. For the second low-pass filter, the learning process is nearly always stable over the 400 cycles that are undertaken with no divergence seen, and convergence is faster than in the non-filtered case. This is true for all the demands used.

8 Non-causal Filters

In order to assess whether additional lag in the system worsens the process of learning and subsequent stability, it is necessary to examine non-causal filters in place of the causal ones already tested. The filter design process is simplified as the oscillation frequencies are unchanged by the addition of the filter. The maximum impulse response is unlikely to be altered and so the design of the filter simply involves reducing the gain at these frequencies. Two classes of non-causal filter have been selected for use; one filter that can be implemented in batch mode, and one that has no such restriction. Although there are several techniques available for batch-mode filtering, the zero-phase IIR filter has been chosen for its simplicity and effectiveness.

8.1 Linear phase FIR filter with offset

A linear phase FIR filter is produced by creating a non-causal filter of order n, symmetrical about its mid-point(s), and then shifting it $\frac{n}{2}$ $\left(\frac{n+1}{2}\right)$ samples in order to make it causal. If this last stage is omitted then a zero-phase FIR filter is obtained which has no limitation on having to be performed in batches. This price of the non-recursive operation is a very large order compared the IIR equivalent. Four filters of this type have been implemented, two lowpass filters and two bandstop. The first lowpass filter is of order $n = 2101$ and has a gain of -36dB at the primary frequency, its magnitude before the cut-off is extremely aggressive, taking a value of -18dB at 0.8Hz. The second lowpass filter has the same magnitude at the fundamental frequency but only -13dB at 0.8Hz, it also has greater attenuation at higher frequencies. The first bandstop filter is of order $n = 2325$ and has extremely high attenuation at low frequencies. At

1.6Hz this is -60dB which reduces slightly to -48dB at the primary frequency. The second bandstop filter has a higher cut-off point making it less aggressive at low frequencies. At 1.6Hz the attenuation is -13dB, increasing to -38dB at the primary frequency. Both bandstop filters have similar characteristics above this frequency, the upper cut-off being 8Hz. The order of these filters approaches the maximum achievable with the hardware and sampling frequency used, therefore, although the attenuation is satisfactory, the cut-offs are not sharp. It is also advantageous that the filter should not be applied on data that is in the process of being updated, that is $n < 2\lambda$.

Figure 26 shows how large attenuation causes slow convergence, the first bandstop filter taking double the number of cycles to converge in every test performed. Its low frequency attenuation effectively gives it a lower learning gain, Γ, and increases T_{ins} at the cost of convergence. The effect of aggressive low frequency filtering extends beyond this, however; the large peaks that occur in the plots of NE against trial number are much reduced, even below the values seen in the less aggressive filters with much lower learning gains. This means that greater low frequency attenuation produces less deviation in the cycle error. The first bandstop filter is more successful than the first lowpass filter, as, with a limited order, it is able to supply greater attenuation at low frequencies. The more aggressive filters are the only ones to allow the test to last 400 cycles in Figure 26.

Without differences in the phase characteristic confusing the issue, lowpass and bandstop filters can also be compared. All the results obtained show that performance is determined by the amount of the low frequency attenuation. This suggests two points:

- Frequencies above 8Hz play an insignificant role in influencing the performance in the tests conducted
- Short term performance is mainly dictated by the magnitude plot of the system below the primary frequency

Short term performance is taken to include convergence rate and changes in the cycle error between trials. It differentiates between long term effects such as unstable frequencies and the effect of integrating high frequency error inherent in the simple ILC laws. Long term performance is therefore mostly influenced by the attenuation at the primary frequency and above. Figure 27 illustrates these points, showing the most successful non-causal and causal filters of those tried. The non-causal bandstop 1 filter has less cycle error deviation due to its high frequency attenuation and lack of destabilizing phase-lag. This overcomes the advantage of the increased instability frequencies that occur when using causal filters. The less aggressive filters are seen to suffer from divergence during the test.

8.2 Zero-phase IIR filter

A filter is designed in the normal way, but is run back and forth along a section of either the error or the ILC input to the plant. Unless the signal is divided into sections of less than $\frac{N-\lambda}{2}$ samples in length (where N is the samples per trial) and each one filtered separately, there will be insufficient time between the signal being recorded and the need for its use in the input to the plant. This length can be increased by using sections that overlap, and filtering them in parallel. This, however, only increases the allowable batch size to $N - \lambda$ samples. Because the ends of each batch are subject to error in the filtering

process, which causes them to recombine imperfectly, a longer batch length is desirous. However, longer sized batches have the effect that there is at least a cycle's duration between the error being recorded and its use. It will be seen that this causes problems as well as very slow convergence. In order to increase convergence speed, it is tempting to allow learning to recommence whilst a section of the error is being filtered. This means that the error will then form the update that follows on from a different input to the one which caused it. Experiments have shown that this always leads to large oscillation of the cycle error and no further convergence. Therefore two methods have been used which keep the same input to the plant during the filtering process. The first filters a single cycle-length of error as it arrives, then repeats the input while it is filtered in reverse. The error is extended in both directions to avoid transients. Convergence is twice as slow due to the cessation of learning. Learning must be halted during the reverse filtering stage as it would then form half of the next update and thus, in part, create the cycle error oscillations described. The second method takes lengths equal to three cycle-lengths and forward and reverse filters them in the same manner. The input is held for six cycles and only the middle cycle-length of error is used in the learning process. This helps reduce the filtering transients. The filter that has been used with these methods has a cut-off of 1.5Hz, following recommendations made in the last section. The small cut-off frequency/Nyquist frequency ratio has limited the filter order available for the class of filter chosen, and the cut-off is not ideal. The attenuation, however, will be double due to the dual filtering.

Figure 28 shows results obtained using the two methods of non-causal filtering with a zero-phase IIR filter that have been described. The best result obtained

with an FIR filter has been included for comparison. Figure 29 shows results obtained with the same filters, but using a different demand.

The results show that batch mode filtering processes are ill-suited to ILC implemented in repetitive form (with no resetting of initial conditions). Despite large attenuation, their performance leads to transitory cycle error, especially with high unit-rates and challenging demands. This is made more obvious when it is remembered that the triple segment IIR filter used only updates every 6 cycles, and the single segment, every two. Discrepancies that arise at the extremities of the system output for a fixed demand are the cause of these irregularities. The updates fit together imperfectly and cause oscillations. The value of E_{min}, however, is in some cases the lowest seen due to the choice of the cut-off frequency. It was found that the IIR filter's performance was improved in terms of less transient cycle error when the demand was not filtered. This differs from all the other tests performed in this respect, and is a consequence of both its increased attenuation, and ability to make corrections only every 6 cycles.

9 Conclusions and Further Work

Phase-lead ILC has been found to out-perform both P-Type and D-Type ILC when applied to the non-minimum phase test-bed described in this paper. Phase-lead ILC has been examined and reasons for its success, and indeed failure, have been put forward. A method of arriving at the phase-lead that produces the best performance has been proposed, using knowledge of the plant model. This also helps predict the likely success of the P-Type algorithm on a given plant. The existence and effect of unstable frequencies caused by phase-

lead ILC has been discussed and a method of predicting their value and harmfulness given. A number of both causal and non-causal filters have been tested and design procedures described in order to maximize performance. The role of attenuation of various frequencies has been discussed and results presented to illustrate the conclusions drawn. Results have indicated that demands can only be learnt up to a certain frequency. It has also been seen that, due to unstable frequencies, stability of phase-lead ILC cannot be assured as the cycle number progresses. It is therefore necessary to combine the process with a mechanism that halts the learning process when a certain level of error has been achieved. It would then be restarted when the cycle error dictates that a change in the system dynamics has occurred. Although this is a regrettable situation, the great success of the simple ILC schemes shown should help compensate for it. Further work will concentrate on the application of simple structure ILC algorithms on more complex non-minimum phase systems.

References

[1] S. Arimoto, F. Miyazaki, and S. Kawamura. Bettering operation of dynamical systems by learning: A new control theory for servomechanism or mechatronics systems. In *Proceedigs of the 23rd Conference on Decision and Control*, pages 1064–1069, December 1984.

[2] S. Arimoto, F. Miyazaki, S. Kawamura, and S. Tamaki. Learning control theory for dynamical systems. In *Proceedings of the 24th Conferenece on Decision and Control*, pages 1375–1380, December 1985.

[3] H. Ahn, C. Choi, and K. Kim. Iterative learning control for a class of nonlinear systems. *Automatica*, 29(6):1575–1578, 1993.

[4] Z. Bien and K. Huh. Higher-order iterative learning control algorithm. In *IEE Proceedings*, volume 136, pages 105–112, May 1989.

[5] M. Norrlof and S. Gunnarsson. A frequency domain analysis of a second order iterative learning control algorithm. In *Proceedings of the 38th Conference on Decision and Control*, pages 1587–1592, December 1999.

[6] T. Kuc, J. Lee, and K. Nam. An iterative learning control theory for a class of nonlinear dynamic systems. *Automatica*, 28(6):1215–1221, 1992.

[7] Y. Chen, J. Xu, and T. Lee. Current iteration tracking error assisted iterative learning control of uncertain nonlinear discrete-time systems. In *Proceedings of the 35th Conference on Decision and Control*, pages 3038–3043, December 1996.

[8] Y. Chen, C. Wen, and M. Sun. A robust high-order p-type iterative learning controller using current iteration tracking error. *International Journal of Control*, 68(2):331–342, 1997.

[9] R. W. Longman. Iterative learning control and repetitive control for engineering practice. *International Journal of Control*, 73(10):930–954, 2000.

[10] D. Wang. On d-type and p-type ilc designs and anticipatory approach. *International Journal of Control*, 73(10):890–901, 2000.

[11] K. Park, Z. Bien, and D. Hwang. Design of an iterative learning controller for a class of linear dynamic systems with time delay. In *IEE Proceedings - Control Theory Applications*, volume 145, pages 507–512, Nov 1998.

[12] S-L. Wirkander and R. W. Longman. Limit cycles for improved performance in self-tuning learning control. *Advances in the Astronautical Sciences*, 154:763–773, 1999.

[13] A. Barton, P. Lewin, and D. Brown. Practical implementation of a real-time iterative learning position controller. *International Journal of Control*, 73(10):992–999, 2000.

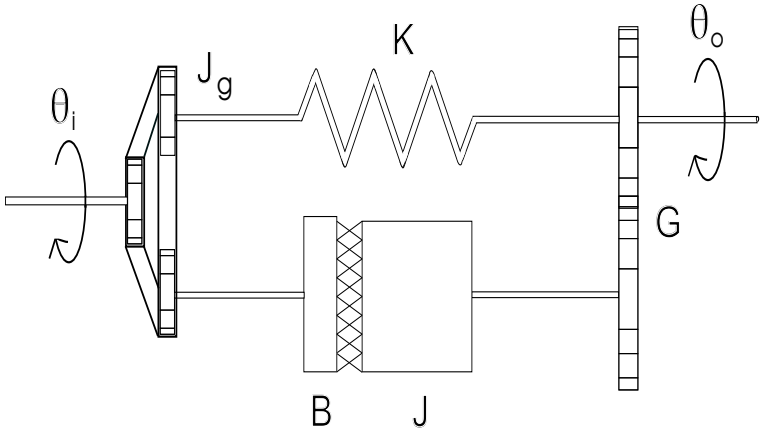

Figure 1: Mechanical realization of non-minimum phase component

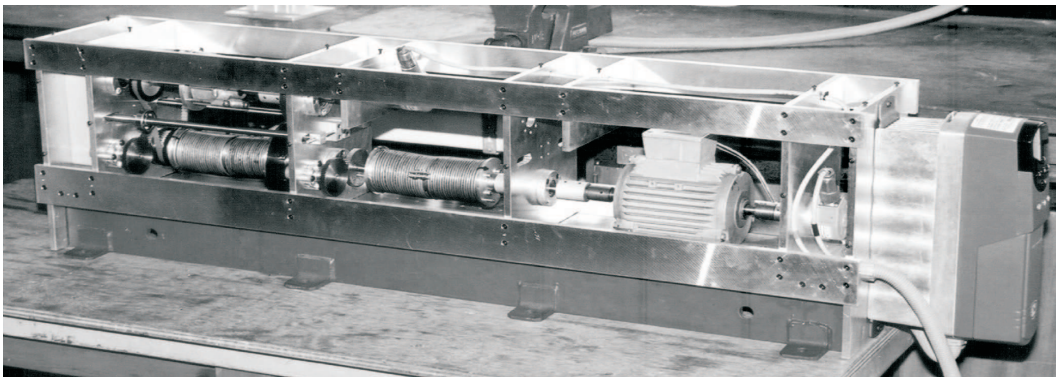

Figure 2: Entire mechanical testbed

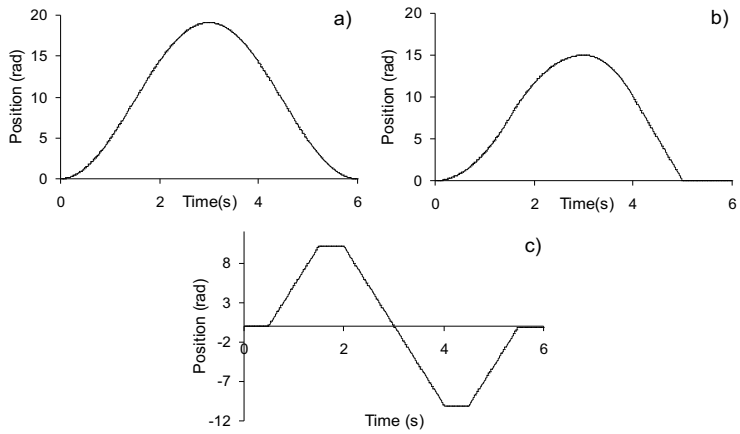

Figure 3: 10 UPM a) sinewave, b) R1, and c) R2 demands

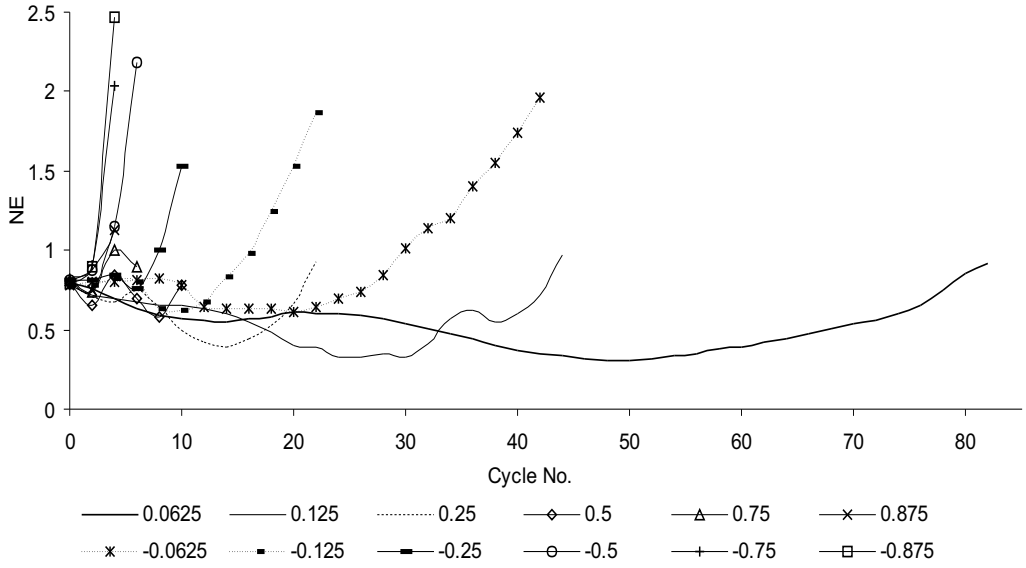

Figure 4: Error results for a 10 UPM sinewave demand for a variety of gains, using the P-Type algorithm

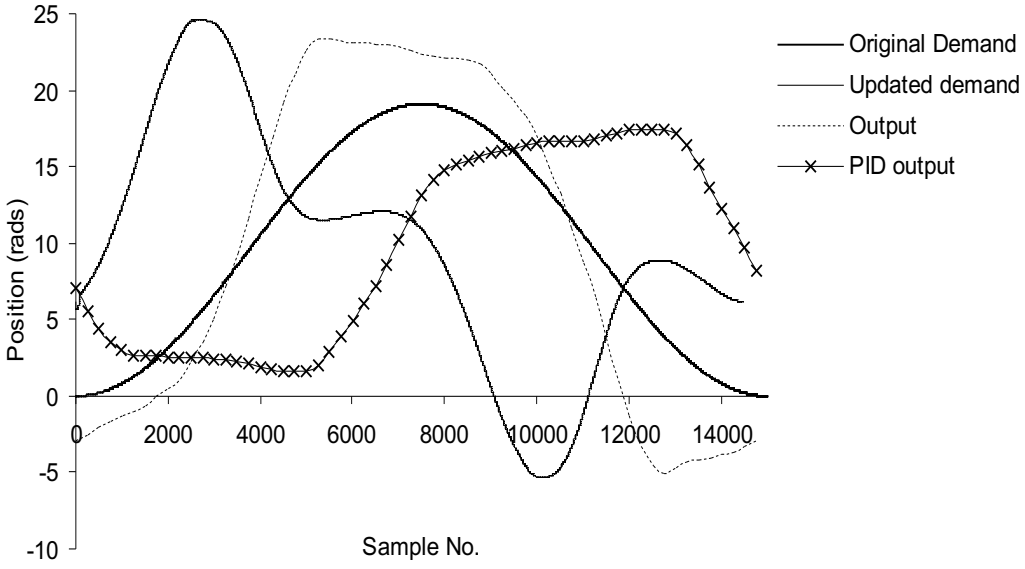

Figure 5: Data recorded during cycle 50 of a 10 UPM sinewave with $\Gamma = 0.0625$

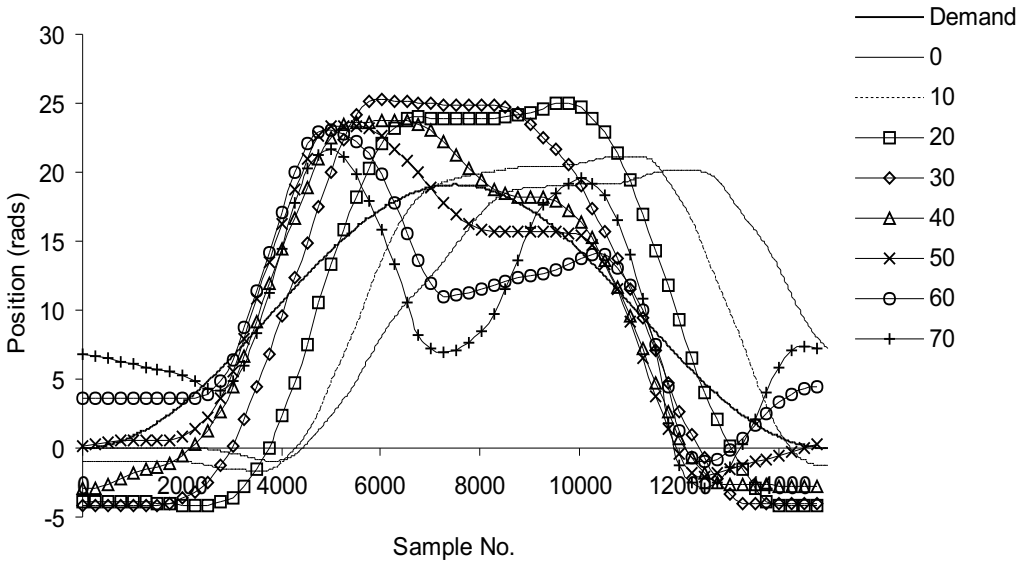

Figure 6: Output evolution of 10 UPM sinewave demand with $\Gamma = 0.0625$

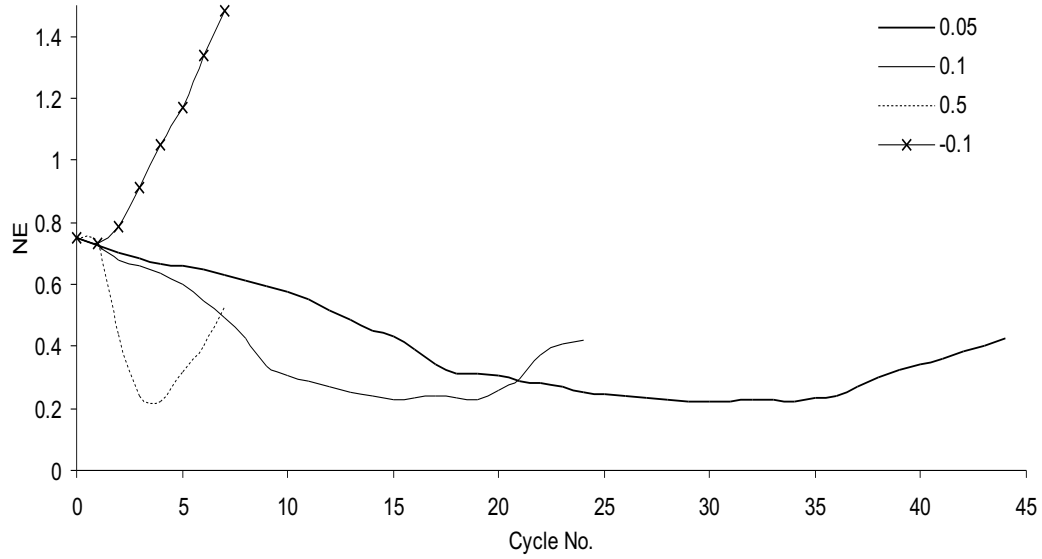

Figure 7: Results for a 10 UPM sinewave demand for different gains, using the D-Type algorithm

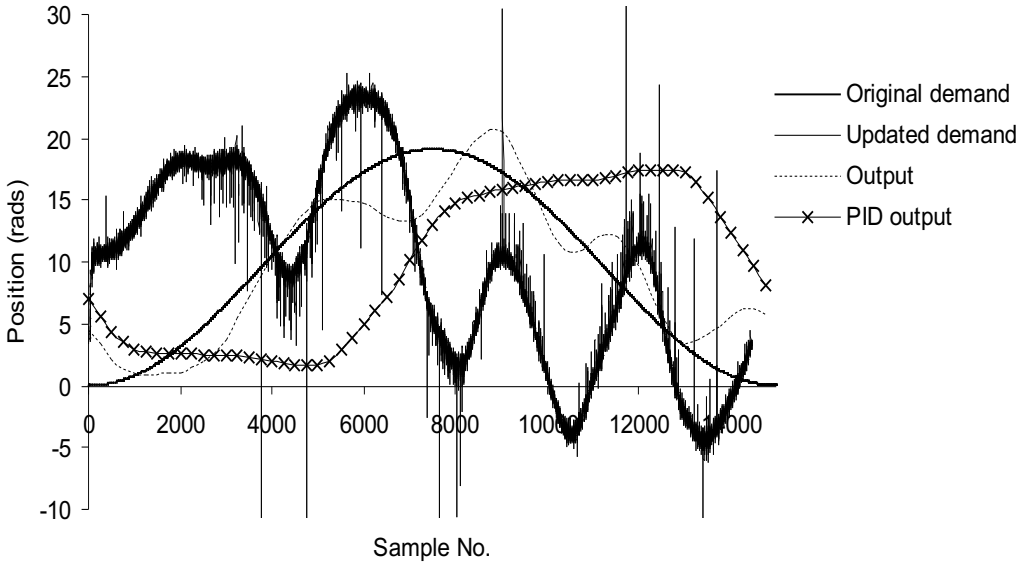

Figure 8: Data recorded during cycle 30 of a 10 UPM sinewave with $\Gamma = 0.05$

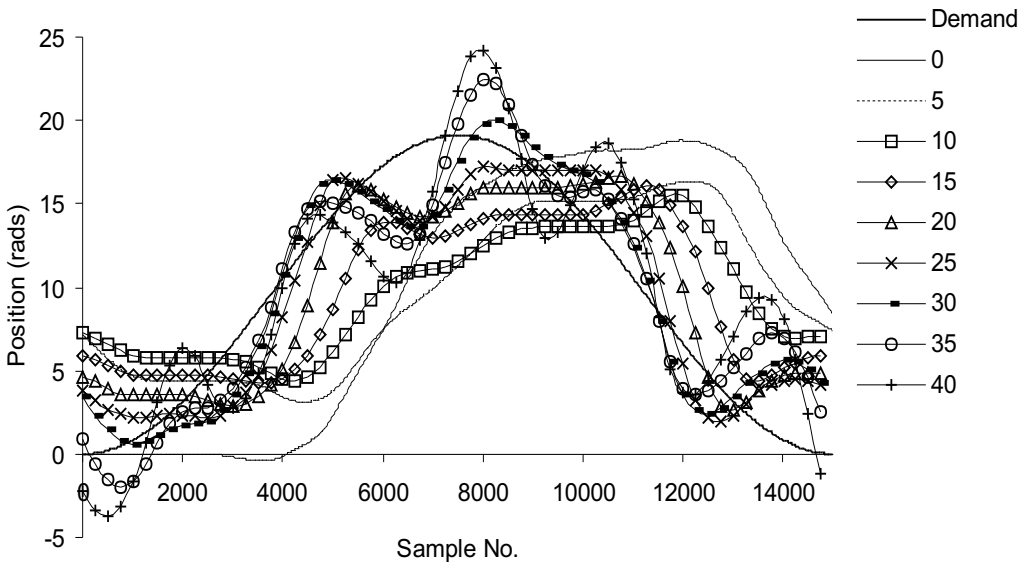

Figure 9: Output evolution 10 UPM sinewave demand with $\Gamma = 0.05$

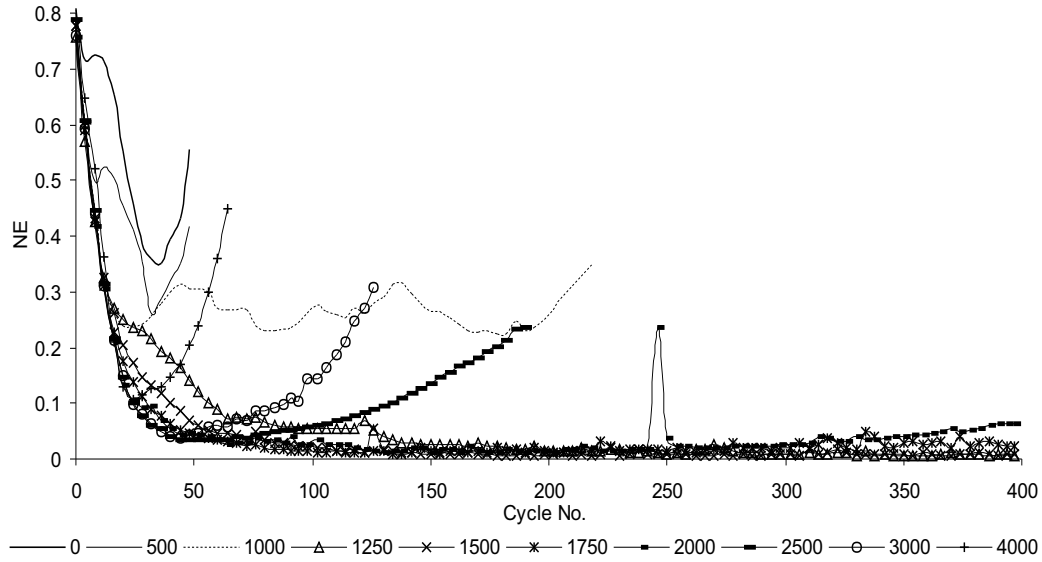

Figure 10: Error results for a 10 UPM sinewave demand for $\Gamma = 0.1$, using a variety of phase-leads

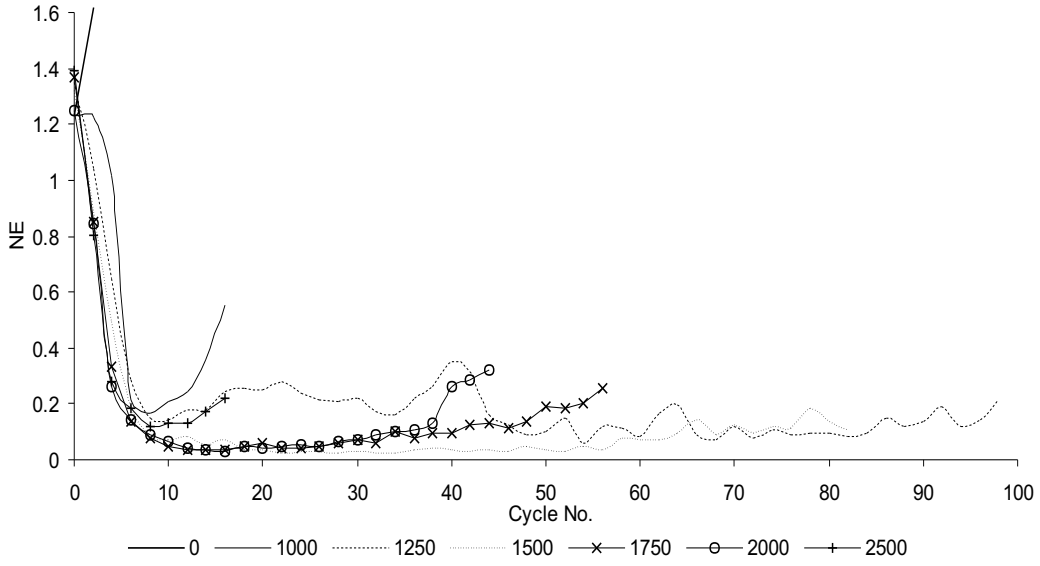

Figure 11: Error results for a 20 UPM R1 demand, $\Gamma = 0.5$ and a variety of phase-leads

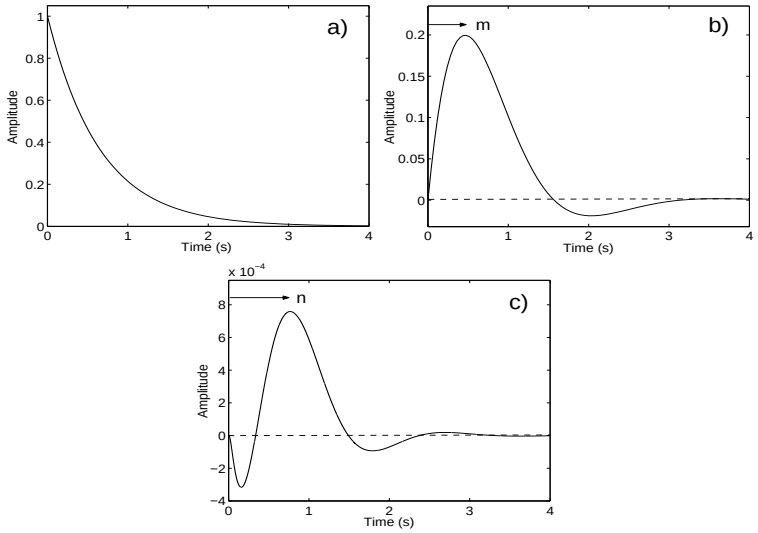

Figure 12: Impulse responses of a) a generic first order plant, b) a higher order plant, and c) the non-minimum phase test facility

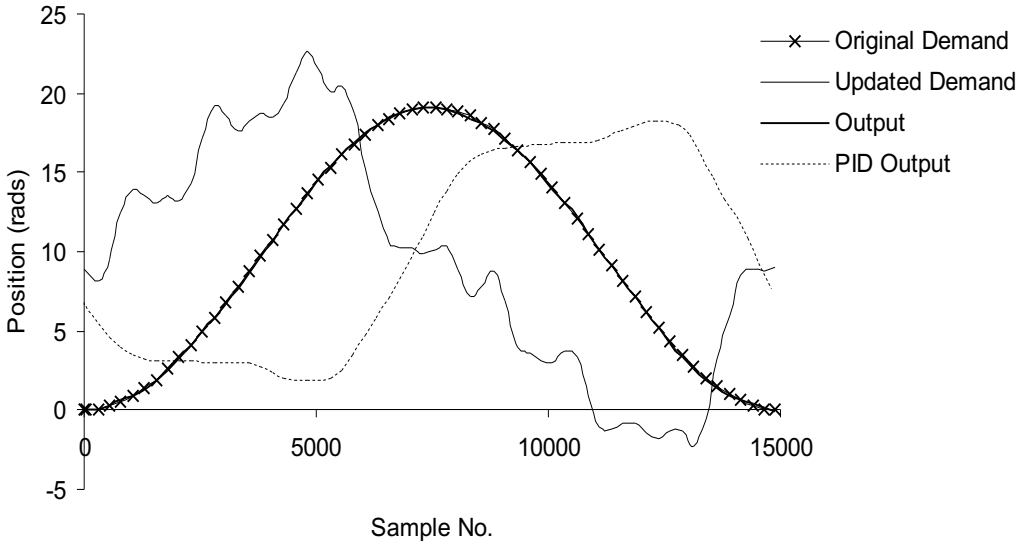

Figure 13: Data recorded during cycle 296 of a 10 UPM sinewave demand with $\Gamma = 0.1$, $\lambda = 1250$

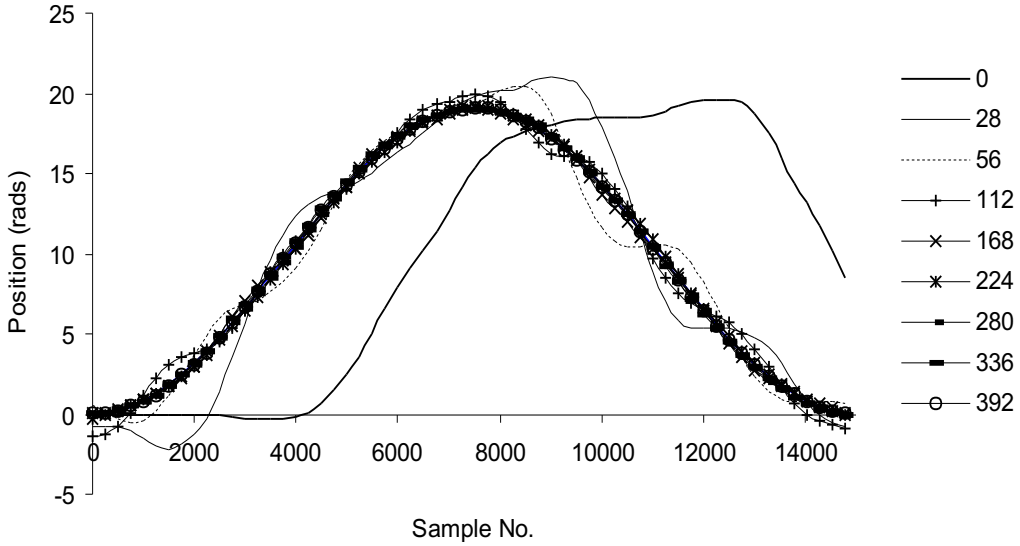

Figure 14: Output evolution 10 UPM sinewave demand with $\Gamma = 0.1$, $\lambda = 1250$

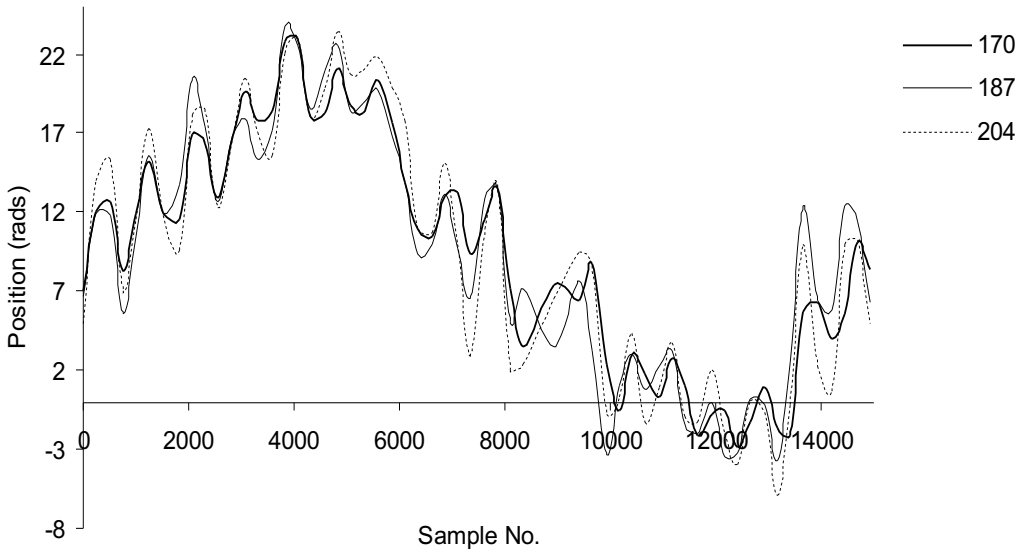

Figure 15: Updated demand of 10 UPM sinewave with $\Gamma = 0.5$, $\lambda = 1250$

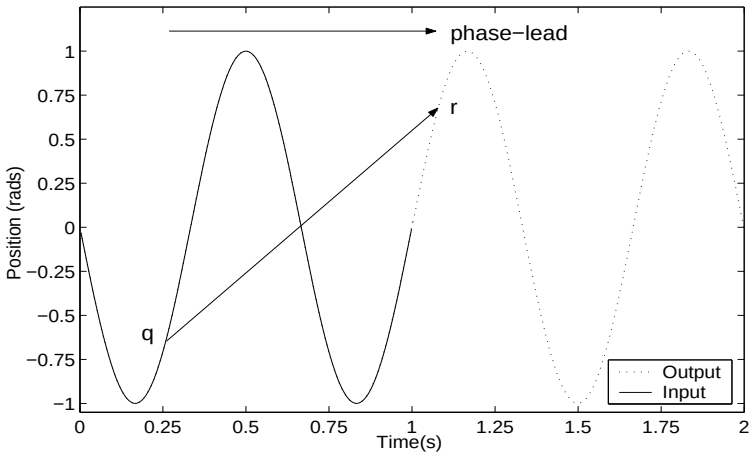

Figure 16: Propagation of an oscillation

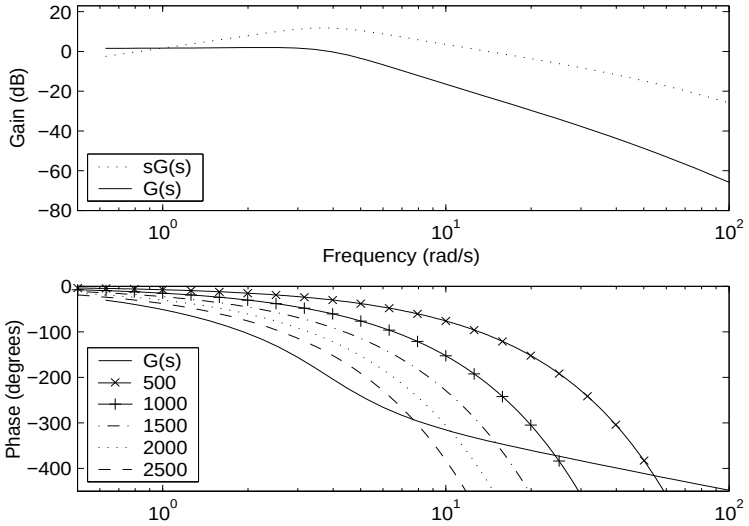

Figure 17: Bode plot showing intersections of phase-lead lines

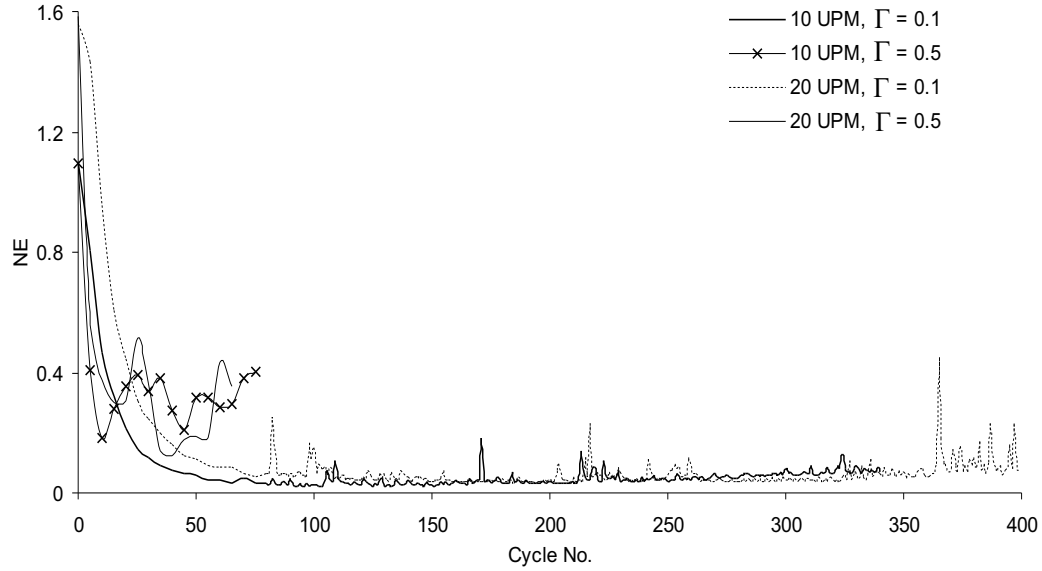

Figure 18: Error results for the R2 demand at two unit rates, both using the optimum phase lead ($\lambda = 1500$)

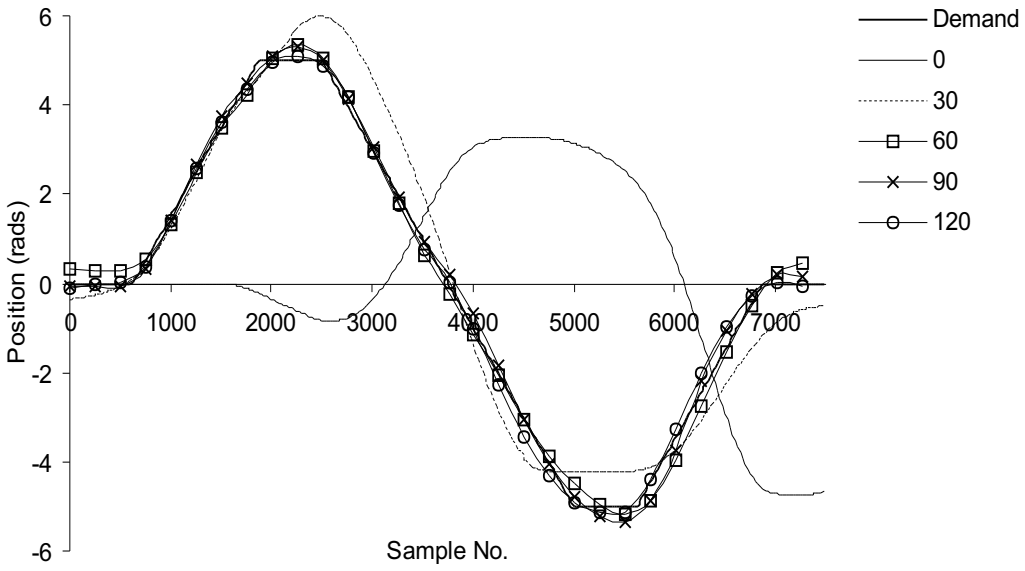

Figure 19: Output evolution of 20 UPM R2 demand with $\Gamma = 0.1$, $\lambda = 1500$

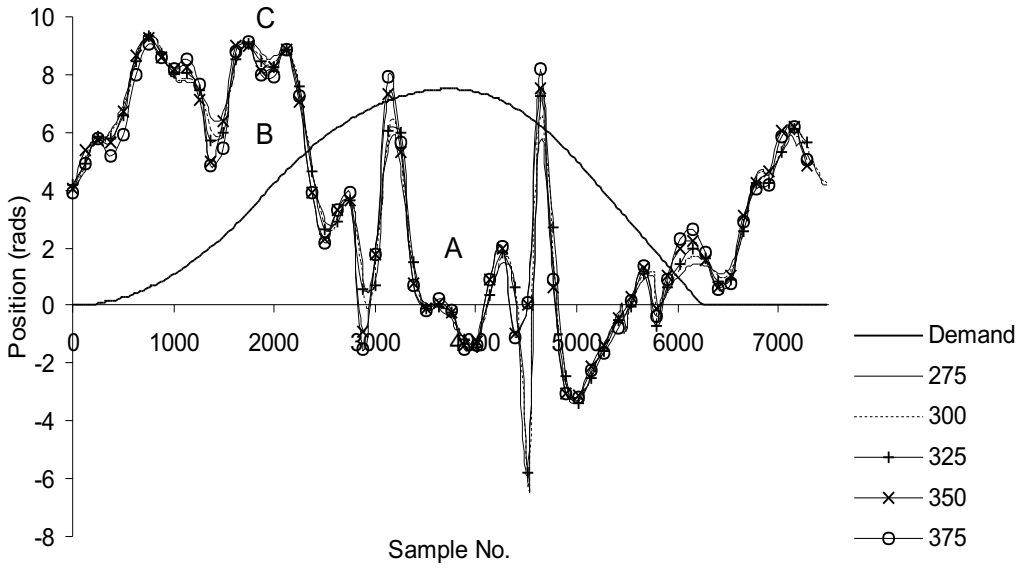

Figure 20: Updated demand using 20 UPM R1 demand with $\Gamma = 0.3$, $\lambda = 1500$ and the bandstop filter

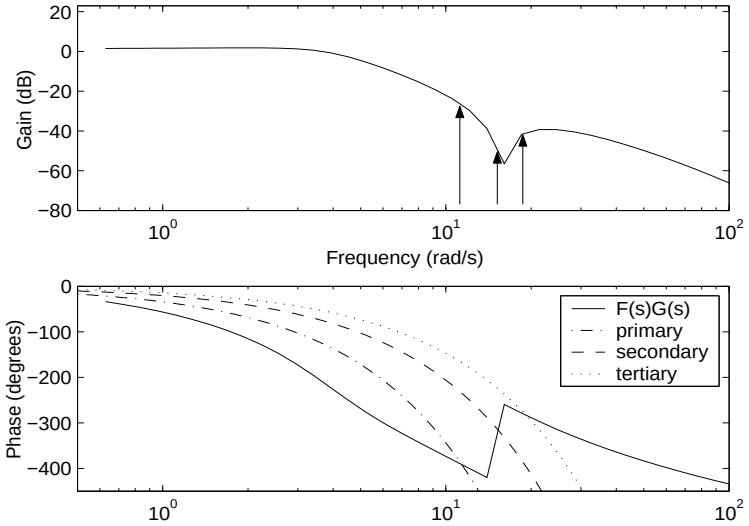

Figure 21: Bode plot showing intersections of $F(s)G(s)$ and primary, secondary and tertiary frequencies for $\lambda = 1750$

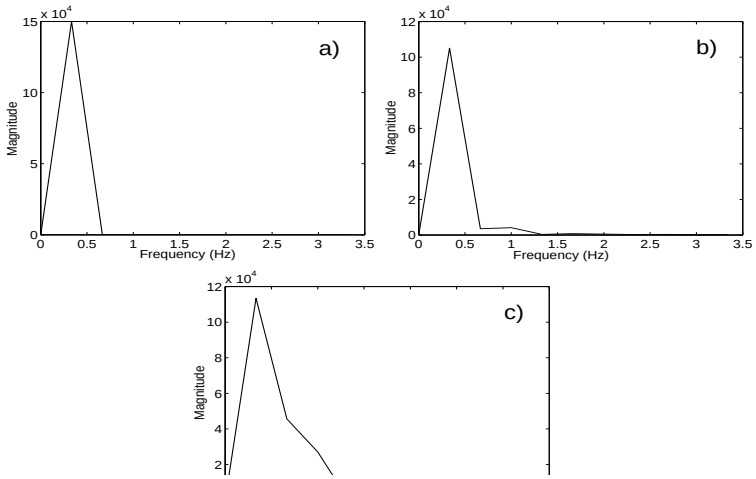

Figure 22: Power spectrums of the 20 UPM a) sinewave, b) R1, and c) R2 demands

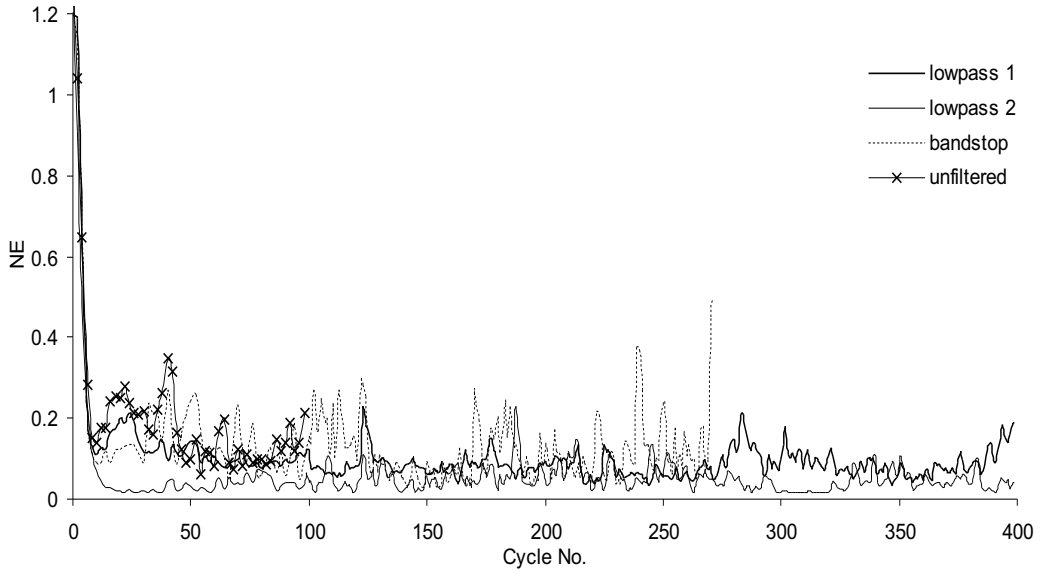

Figure 23: Results for 20 UPM R1 using a variety of causal filters, all with their optimum phase-lead and $\Gamma = 0.5$

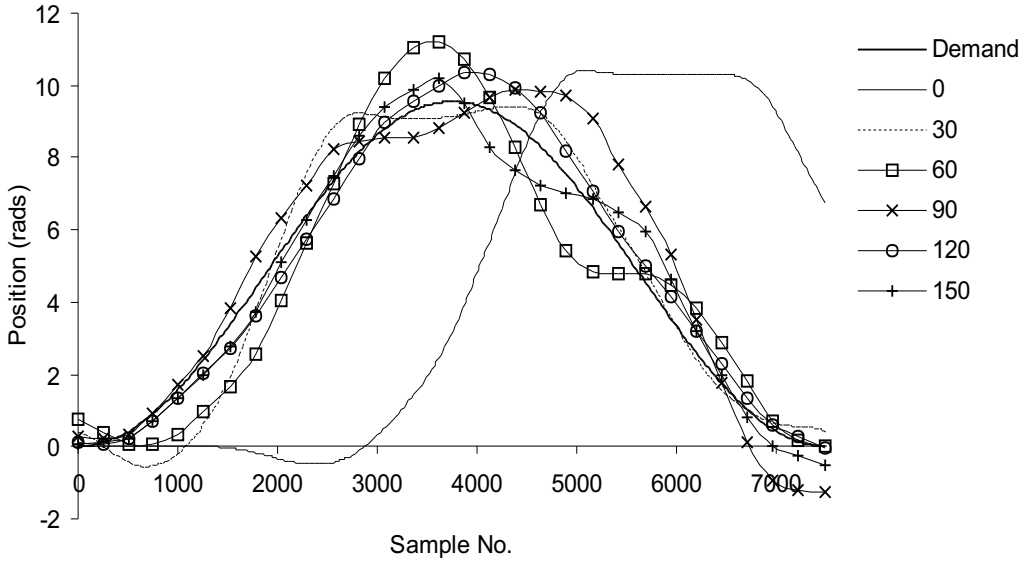

Figure 24: Output evolution of 20 UPM sinewave demand with $\Gamma = 0.1$ and optimum phase-lead using the bandstop filter

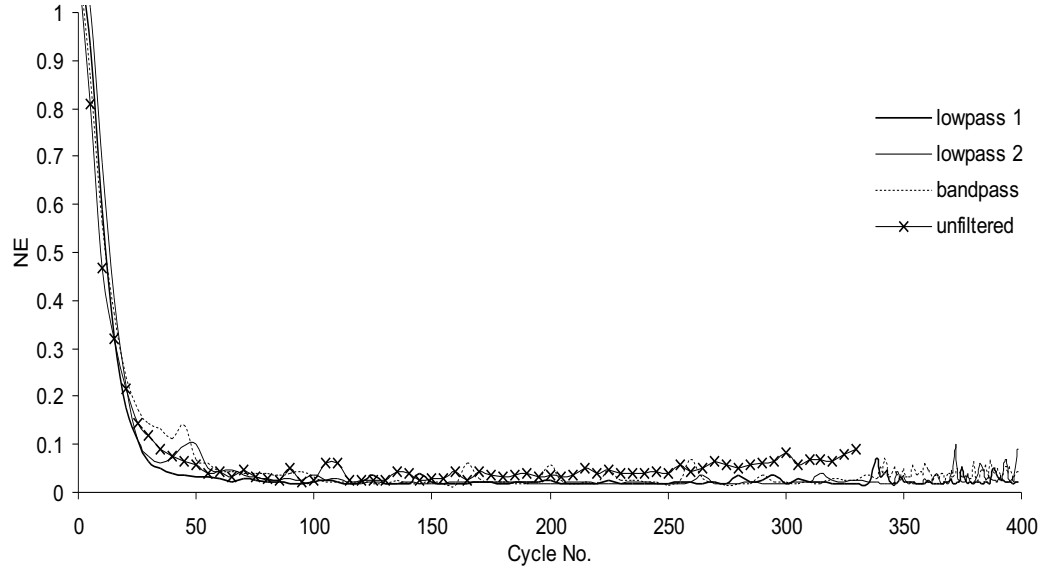

Figure 25: Error results for 10 UPM R2 using a variety of causal filters, all with their optimum phase-lead and $\Gamma = 0.1$

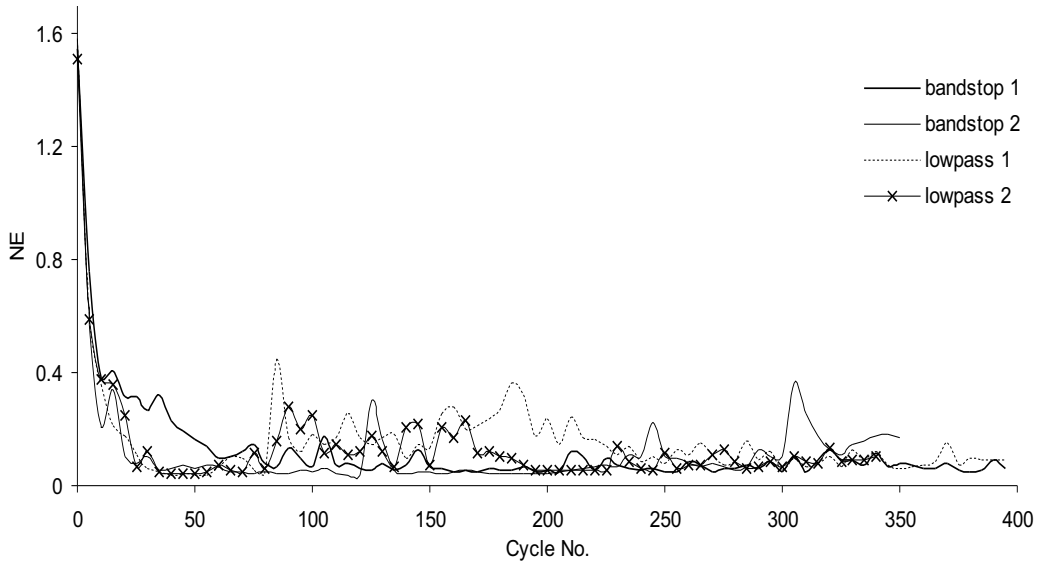

Figure 26: Error results for 20 UPM R2 using a variety of non-causal filters, all with their optimum phase-lead and $\Gamma = 0.5$

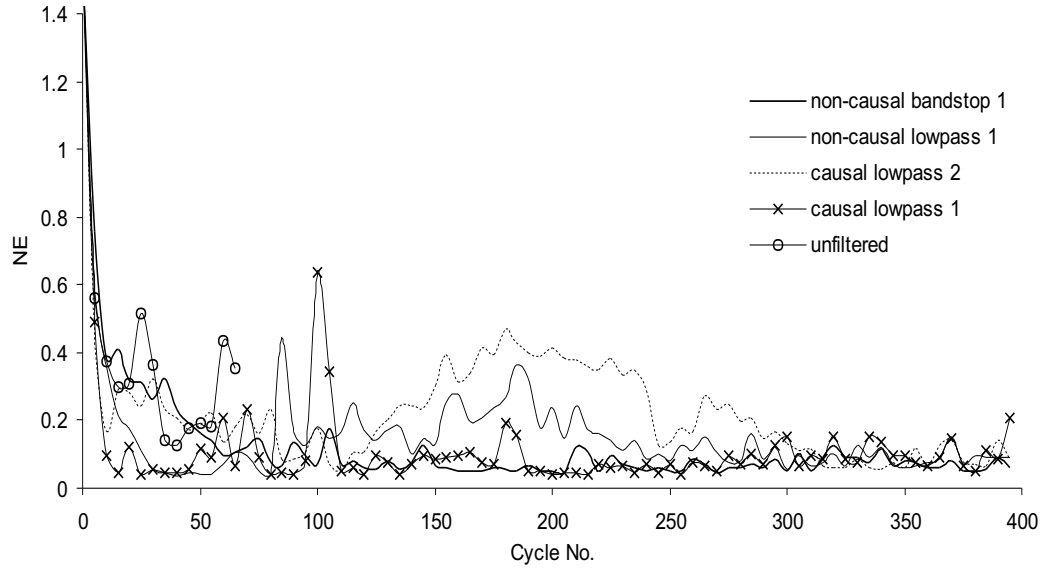

Figure 27: Error results for 20 UPM R2 using the most successful filters, all with their optimum phase-lead and $\Gamma = 0.5$

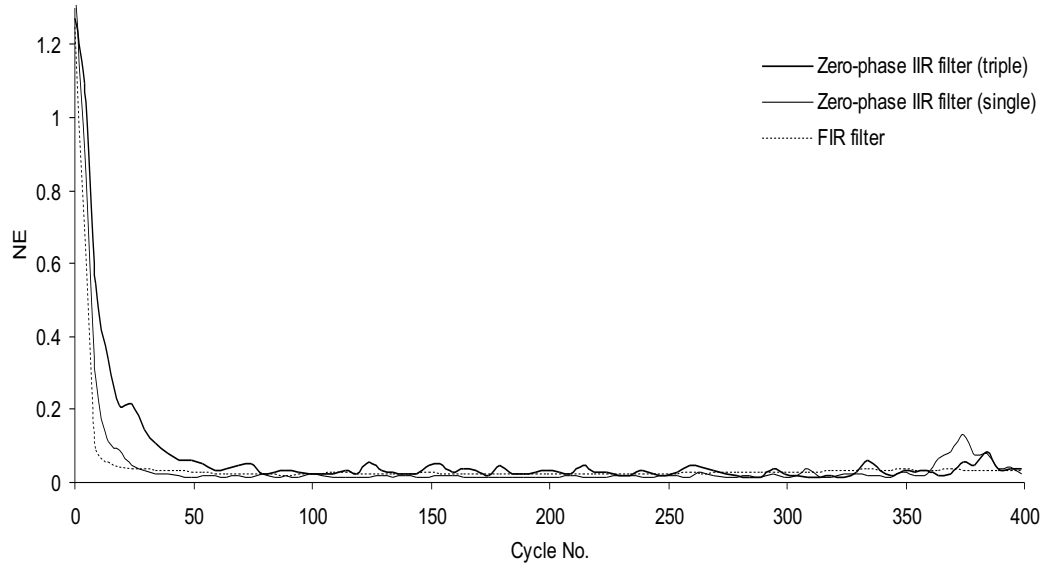

Figure 28: Error results for a 20 UPM R1 demand using non-causal filters, all with their optimum phase-lead and $\Gamma = 0.5$

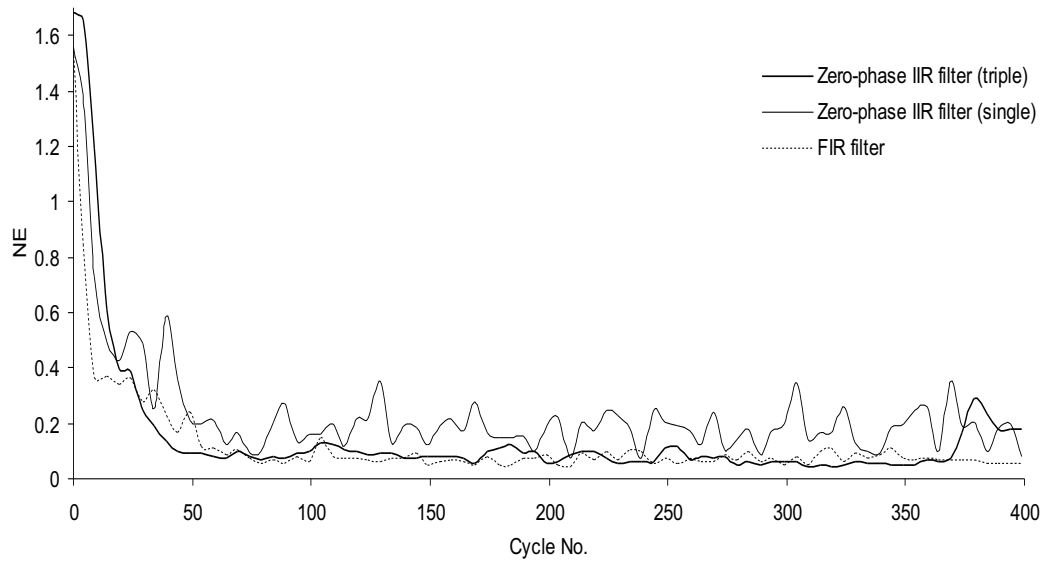

Figure 29: Error results for a 20 UPM R2 demand using non-causal filters, all with their optimum phase-lead and $\Gamma = 0.5$