ROOT MOMENTS: A NONLINEAR SIGNAL TRANSFORMATION FOR MINIMUM FIR FILTER DESIGN

Tania Stathaki   Ioannis Fotinopoulos   Anthony Constantinides
Signal Processing and Digital Systems Section, Imperial College, UK

ABSTRACT

In this work we propose to design a minimum phase FIR digital filter transfer function from a given linear phase FIR transfer function which has identical amplitude. We are concentrating on very high degree polynomials for which factorisation procedures for root extraction are unreliable. We also assume that the polynomials have roots on the unit circle. The approach taken involves the use of the Cauchy Residue Theorem applied to the logarithmic derivative of the transfer function. This leads into a set of parameters derivable directly from the polynomial coefficients that facilitate the factorisation problem. The concept is developed in a way that leads naturally to the celebrated Newton Identities. In addition to solving the above problem, the results of the proposed design scheme are very encouraging as far as robustness and computational complexity are concerned.

1. INTRODUCTION

The design of Finite Impulse Response (FIR) digital filters has attracted considerable attention [1-4]. An influential representative of the methods is based on the Remez exchange algorithm. However, the Remez and also most procedures assume a linear phase response with the consequence that the resulting filters do not have the lowest group delay. Direct design with prespecified phase response is possible. In this work we concentrate on the following problem.

"Given a linear phase FIR digital filter transfer function to determine an FIR digital filter which has identical amplitude response but is of minimum phase".

At first glance this may appear to be a trivial problem. Indeed a naive approach would be to factorise the given FIR transfer function and replace each of the zeros outside the unit circle with its reciprocal. This, in principle at least, would leave the overall amplitude response unaltered and would make the resulting transfer function of minimum phase. However for a large class of polynomials, factorisation is an unreliable process especially when some of the zeros are located on the unit circle.

A new approach for polynomial factorisation without root finding is attempted. The approach involves the use of the Cauchy Residue Theorem applied to the logarithmic derivative of the transfer function. This leads into a set of parameters, the *root moments*, derivable directly from the polynomial coefficients that facilitate the factorisation problem. In addition to solving the above problem, the results of the proposed design scheme are very encouraging as far as robustness and computational complexity are concerned.

2. PRELIMINARIES

We consider a linear phase FIR digital filter transfer function having the following form

$$H(z) = z^n + h_1 z^{n-1} + h_2 z^{n-2} + \cdots + h_n = \prod_{i=1}^{n} (z - r_i) \qquad (1)$$

Linear phase, real FIR digital filter transfer functions are symmetric or anti-symmetric and have non-minimum phase. It can be easily shown that if a transfer function of the above type has a zero at the location r_i, it will also have the zeros $1/r_i$, r_i^* and $1/r_i^*$ for $|r_i| \neq 1$.

Lets suppose that the roots of the polynomial $H(z)$ above are denoted by r_i. We employ the following notation:

- $r_j = r_{j\text{in}}$ if the root r_j is inside the unit circle.
- $r_j = r_{j\text{out}}$ if the root r_j is outside the unit circle.
- $r_j = r_{j\text{o}}$ if the root r_j is on the unit circle.

Thus we can write

$$H(z) = \left[\prod_j (z - r_{j\text{in}})\right]\left[\prod_j (z - r_{j\text{out}})\right]\left[\prod_j (z - r_{j\text{o}})\right] \quad \text{or}$$

$$H(z) = H_{\min}(z) H_{\max}(z) H_{\text{o}}(z)$$

where $H_{\min}(z)$ is the minimum phase part of $H(z)$ and $H_{\max}(z)$ is the maximum phase part of $H(z)$. The factor $H_{\text{o}}(z)$ contains all roots that are on the unit circle.

Some useful general points need to be made.

- The group delay of an n th order linear phase real FIR transfer function is $\tau(\theta) = n/2$. For a range of applications with stringent specifications as in telecommunications, a typical FIR digital filter transfer function may be of length 200 or more. For such filters the group delay may be undesirable particularly when it approaches large values, such that bidirectional human-to-human communication is not viable.
- Often in many applications the phase response is either unimportant or irrelevant. For example in some speech processing areas it is not significant. This form of freedom is not normally taken into consideration by existing FIR filter design methods.

At any rate, the design of minimum phase FIR filters from linear phase FIR filters with specific amplitude requirements would inevitably lead to a stage of factorisation in order to select the appropriate zeros, and hence problems with imprecisions would arise.

3. PROPOSED DESIGN ALGORITHM

At this section we aim to derive the required nonlinear phase FIR filter transfer functions from corresponding linear phase functions which are assumed to be designable by standard means as the Remez exchange algorithm.

Let the linear phase real FIR filter transfer function be $H(z) = H_{\min}(z) H_{\max}(z) H_{\text{o}}(z)$ as already indicated.

Let also n_{o} be the number of zeros of $H_{\text{o}}(z)$ and n_i the number of zeros of $H_{\min}(z)$ and to the extent of $H_{\max}(z)$.

If the zeros of $H_{\min}(z)$ are of the form $\rho_i e^{\pm j\theta_i}$ we can write

$$H_{\min}(z) = \prod (z^2 - 2\rho_i \cos\theta_i z + \rho_i^2).$$

Suppose that $H_{\min}(e^{j\theta})$ can be written as

$$H_{\min}(e^{j\theta}) = A(\theta)e^{j\phi(\theta)}$$

where $\phi(\theta)$ is a nonlinear function.

Because $H(z)$ is linear phase, the zeros of $H_{\max}(z)$ are of the form $\frac{1}{\rho_i}e^{\pm j\theta_i}$. Thus, we can write $H_{\max}(z)$ as

$$H_{\max}(z) = \prod(\rho_i{}^2 z^2 - 2\rho_i \cos\theta_i z + 1)$$

$$= z^{n_i}\prod(\rho_i{}^2 - 2\rho_i \cos\theta_i z^{-1} + z^{-2}) = z^{n_i} H^*_{\min}(z)$$

or

$$H_{\max}(e^{j\theta}) = e^{jn_i\theta}A(\theta)e^{-j\phi(\theta)}.$$

Hence

$$|H_{\min}(e^{j\theta})| = |H_{\max}(e^{j\theta})|$$

as already expected.

On the unit circle we have

$$H_o(e^{j\theta}) = \prod_{r=1}^{n_o/2}(z^2 - 2\cos\theta_r z + 1) = \prod_{r=1}^{n_o/2} z(z - 2\cos\theta_r + z^{-1})$$

$$= z^{n_o/2}\prod_{r=1}^{n_o/2}(2\cos\theta - 2\cos\theta_r) = e^{jn_o\theta/2}B(\theta)$$

Hence

$$H(e^{j\theta}) = e^{j(n_i + n_0/2)\theta}[A(\theta)]^2 B(\theta)$$

The group delay is

$$\tau(\theta) = (n_i + \frac{n_o}{2}) = \frac{n}{2}.$$

Thus, in principle, to obtain a minimum phase version of the given transfer function we can follow the steps below.

Step 1

Either determine $H_{\max}(z)$ and reflect its zeros into the unit circle, or determine $H_{\min}(z)$ and make each of its zeros of multiplicity 2.

Step 2

Find $H_o(z)$

Step 3

Construct the transfer function as $T(z) = [H_{\min}(z)]^2 H_o(z)$.

Then we shall have $|T(e^{j\theta})| = |H(e^{j\theta})|$.

Both Step 1 and Step 2 imply at first glance that a root finding procedure may be required. However, as already pointed out, root finding procedures are known to be inaccurate and unreliable for large order polynomials. Factorisation without root finding forms also the basis of the procedure developed in [1],[3-4],[9]. In [1],[4-5] use is made of the real cepstral parameters, where the cepstral aliasing problem is recognised and careful procedures are recommended to reduce its effects. In [3] they approach the factorisation problem from the Lagrange interpolation point of view. In the above procedures it is assumed that the zeros of the transfer function on the unit circle are a priori known. We make no such assumption in our present work.

An alternative and direct polynomial construction procedure without having to go through root estimation procedures is possible through the root moments of a given polynomial [7-8].

4. ROOT MOMENTS

In relation to polynomials typically given as in equ.(1) Newton defined a set of parameters given by

$$S_m = r_1{}^m + r_2{}^m + \cdots + r_n{}^m = \sum_{i=1}^{n} r_i{}^m \qquad (2)$$

where r_i is the ith root of (1). The roots of (1) are not needed explicitly to compute S_m in that these parameters can be determined directly from the coefficients h_i. The parameters S_m are known as the root moments of the polynomial $H(z)$. They are related to many signal processing operations, dominant amongst which is the differential cepstrum. However it would be limiting to think of them purely in this sense since a wider perspective enables us to provide answers to many digital signal processing problems that have been, hitherto, unattainable [7].

4.1 Iterative Estimation of Root Moments

By writing the polynomial (1) as a product of factors we can write $H'(z) = \sum_{i=1}^{n}\frac{H(z)}{z - r_i}$ and given that $H(r_i) = 0$ we have

$$H'(z) = nz^{n-1} + (S_1 + nh_1)z^{n-2} + (S_2 + h_1S_1 + nh_2)z^{n-3} + \cdots + (S_m + h_1S_{m-1} + h_2S_{m-2} + \cdots + nh_m)z^{n-m-1} + \cdots$$

By direct differentiation of equation (1) we have

$$H'(z) = nz^{n-1} + (n-1)h_1z^{n-2} + (n-2)h_2z^{n-3} + \cdots + (n-m)h_m z^{n-m-1} + \cdots \qquad (3)$$

Hence by equating the last two expressions we obtain the following fundamental relationships known as *Newton Identities*

$S_1 + nh_1 = (n-1)h_1$ or $S_1 + h_1 = 0$

$S_2 + h_1S_1 + nh_2 = (n-2)h_2$ or $S_2 + h_1S_1 + 2h_2 = 0$

and generally

$S_m + h_1S_{m-1} + h_2S_{m-2} + \cdots + nh_m = (n-m)h_m$ or

$$S_m + h_1S_{m-1} + h_2S_{m-2} + \cdots + mh_m = 0 \qquad (4)$$

When the signal treated by this means is infinitely long, the above equation is repeatedly used to calculate successive values of the root moments. If the signal is of finite duration then for $m > n$ $S_m + h_1S_{m-1} + h_2S_{m-2} + \cdots + h_nS_{m-n} = 0$.

The same relationship as above can be used to calculate S_m for $m < 0$ by inserting successively values for m equal to $n-1$, $n-2$, $n-3, \cdots$ etc. It should be noted that S_m for either positive or negative values of m are evaluated recursively from the coefficients of equation (1) alone.

The above relationships also follow from the definition of the differential cepstrum and are essentially included in [6]. However in [6], n is assumed to be finite and a priori known. This is only a minor point as the iteration in equ.(4) do not require n to be finite and, hence, it can be applied to infinite duration signals. It is sufficient at this juncture to observe that both finite duration signals and infinite duration signals of exponential entire function type interpretation can be treated in the same way [7]. To facilitate the exposition, the parameters in (2) are referred to as the root moments. This terminology emphasises the deviation from the differential cepstrum.

Essentially one can interpret the set of equations (4) as a transformation of the coefficients $\{h_r\}$ to the parameter set $\{S_m\}$ of the same cardinality. The transformations are one-to-one and hence we can have the following existence corollaries.

Corollary 1 Given a set of coefficients $\{h_r\}$ of the nth degree polynomial in equation (1) which has roots

$\{r_i\}$ $i = 1, \cdots, n,$ there exists a set of parameters $\{S_m\}$ $m = 1, \cdots, n,$ $S_0 = n$, given by equ.(2).

Corollary 2 Conversely given a set of root moments $\{S_m\}$ there exists a set of coefficients $\{h_r\}$ $r = 1, \cdots, n,$ for a polynomial as in equ.(1) determinable recursively through equ.(4). The proofs are self evident from the above analysis.

4.2 Non Iterative Estimation of Root Moments

The Newton Identities yield the root moments of the entire signal, encompassing not only those roots that lie within the unit circle but also those that are outside the unit circle. However, it is often the case that a specific factor of a given polynomial $H(z)$ is required, such as the minimum phase factor and in this case its root moments can be determined in a different manner.

Let a closed contour Γ defined as $z = \rho(\theta)e^{j\theta}$ contain the roots of the required factor of $H(z)$. Then it follows from the Cauchy residue theorem that the root moments of this factor are given by

$$I_\Gamma(m) = S_m^\Gamma = \frac{1}{2\pi j}\oint_\Gamma \frac{H'(z)}{H(z)} z^m dz \tag{5}$$

This is evident from the fact that

$$S_m^\Gamma = \frac{1}{2\pi j}\oint_\Gamma \sum_i \frac{1}{(z - r_i)} z^m dz$$

and the contribution to the integration are those due to those r_i' s that lie within Γ . (It is assumed that we have no zeros on Γ .)

In practice the contour integration will have to be effected directly from the coefficients of $H(z)$ and this can be done quite conveniently through the use of the DFT as it is shown below.

Equation (5) becomes for $z = \rho(\theta)e^{j\theta}$

$$S_m^\Gamma = \frac{1}{2\pi j}\int_{-\pi}^{\pi} g(\theta) e^{j(m+1)\theta} d\theta \tag{6}$$

where

$$g(\theta) = \frac{H'(\rho(\theta)e^{j\theta})}{H(\rho(\theta)e^{j\theta})}\left(\frac{d\rho(\theta)}{d\theta} + j\rho(\theta)\right)\rho^m(\theta) \tag{7}$$

Discretisation of equation (6) suitable for DFT use requires values $\theta_k = \frac{2\pi}{N}k,\ k = 0,1,\cdots, N-1$ for an N -point transform. Therefore, we have the inverse DFT

$$S_m^\Gamma \approx \frac{1}{jN}\sum_{k=0}^{N-1} g(\theta_k) e^{j(m+1)\theta_k} \tag{8}$$

If the contour of integration is the unit circle $C: |z| = 1$ then the resulting root moments from the above, correspond to those of the minimum phase component of $H(z)$. In this case we have the special form of (8)

$$S_m{}^{f_{\min}(z)} \approx \frac{1}{N}\sum_{k=0}^{N-1} \frac{H'(\theta_k)}{H(\theta_k)} e^{j(m+1)\theta_k} \tag{9}$$

For either equ.(7) or equ.(8) the computation of $g(\theta_k)$ can be done through the use of the DFT also.

It is observed that on $z = \rho(\theta)e^{j\theta}$ we can write

$$H'(\rho(\theta)e^{j\theta}) = e^{j(n-1)\theta}\sum_{i=0}^{n-1}(n-i)h_i\rho^{n-i-1}(\theta)e^{-ji\theta}$$

which for $\theta = \theta_k$ can be computed as

$$H'(\rho(\theta_k)e^{j\theta_k}) = e^{j(n-1)\theta_k}\,\mathrm{DFT}\{(n-i)h_i\rho^{n-i-1}(\theta_k)\} \tag{10}$$

Similarly we have

$$H(\rho(\theta_k)e^{j\theta_k}) = e^{jn\theta_k}\,\mathrm{DFT}\{h_i\rho^{n-i}(\theta_k)\} \tag{11}$$

and hence

$$g(\theta_k) = e^{-j\theta_k}\frac{\mathrm{DFT}\{(n-i)h_i\rho^{n-i-1}(\theta_k)\}}{\mathrm{DFT}\{h_i\rho^{n-i}(\theta_k)\}}\left(\frac{d\rho(\theta)}{d\theta} + j\rho(\theta)\right)\rho^m(\theta)$$

With N a power of 2 we can use the Fast Fourier Transform (FFT) algorithm.

5. THE ALGORITHM

The algorithm relies on the direct and accurate extraction of the appropriate factors from the FIR linear phase transfer function $H(z)$ needed to implement $T(z)$ described above. The FIR linear phase transfer function is designed using the Remez or any other similar existing algorithm according to the specifications of the user. The designed filter has roots on the unit circle. For that reason we cannot obtain the minimum phase part by integrating around the unit circle. The approach we take follows the steps below.

Step 1:
We integrate around a circle centered at the origin and of radius less than unity. With a careful choice of the contour radius, the integration gives the root moments $S_1(m) = S_{\text{in}}(m)$ that correspond to that part of the original FIR transfer function which has its zeros inside the unit circle, namely the minimum phase part.
Obviously the radius of the contour is of crucial importance. We have to ensure that the selected circle includes all the roots of the minimum phase part of the original polynomial.
This is examined separately below.

Step 2:
We integrate around a circle centred at the origin and of radius greater than unity. Again the radius of the contour must be selected carefully. A good selection in Step 1 yields a correspondingly good selection for Step 2, that is the reciprocal of the radius selected in Step 1. The integration produces the parameters $S_2(m) = S_{\text{in}}(m) + S_0(m)$ where $S_0(m)$ are the root moments of that factor of the original FIR digital filter transfer function which has its zeros on the unit circle.

Step 3:
The required transfer function has the root moments $S(m) = 2S_{\text{in}}(m) + S_0(m)$, or $S(m) = S_1(m) + S_2(m)$.

Step 4:
From Step 3 and from the Newton's Identities we form the required minimum phase FIR filter transfer function which is the minimum phase version of the initial mixed phase transfer function. The order of this is the same as the order of the original polynomial and equal to $S_1(0) + S_2(0)$.

5.1 Estimation of the radii of the integration

The radii of integrations in the above algorithm must be chosen so as to enclose the appropriate zeros of the given FIR digital filter transfer function. Thus for $S_1(m)$ the radius of the integration contour r must be such that $1 > r > \max(|r_{\text{in}}|)$, while for $S_2(m)$ the radius of the integration contour r must be chosen such that $1 < r < \min(|r_{out}|)$.

For equiripple filters the required radii can be estimated as follows.

Let us remove the linear phase factor from the frequency response to yield only a real function. This function now we shift vertically half way between its maximum and minimum values.

An approximate representation of this zero pattern, is given almost everywhere by

$$C(z) = (z^n - a^n)(a^n z^n - 1) \qquad (12)$$

The above transfer function is equiripple, linear phase and its zeros are located on two circles controlled by the parameter a. We now take the magnitude of it. Since the initial transfer function is equiripple the result of these operations will be a real positive function of equiripple modulus almost everywhere. The ripple variation remains unchanged, namely, a normalised response will vary between $1+\delta$ and $1-\delta$ almost everywhere except in the transition band. The amplitude characteristic is equiripple between the values

$$C_{\max} = a^{2n} + 2a^n + 1 \text{ and } C_{\min} = a^{2n} - 2a^n + 1$$

The mean between the minimum and the maximum value is $(C_{\max} + C_{\min})/2 = a^{2n} + 1$. To calculate the ripple, the amplitude response is normalized, by deviding with the mean $(a^{2n} + 1)$, so that the maximum value now becomes $C_{\max} = \dfrac{a^{2n} + 2a^n + 1}{a^{2n} + 1}$. We use the relationship

$$C_{\max} = \frac{a^{2n} + 2a^n + 1}{a^{2n} + 1} = 1 + \delta \Rightarrow \delta = \frac{2}{a^n + \dfrac{1}{a^n}}$$

The quantity δ can be estimated from the initial transfer function created by the Remez algorithm.

Hence we can estimate the radii of the circles within the zeros of the required polynomials are expected to be located, using the relationships

$$a_{1,2} = \left(\frac{1}{\delta} \pm \sqrt{\left(\frac{1}{\delta^2} - 1\right)} \right)^{\frac{1}{n}}.$$

For very small ripple width, $a_1 << 1$ and a_2 can be approximated to $a_2 = \left(\dfrac{2}{\delta}\right)^{\frac{1}{n}}$.

5.2 Discussion for Future Work

It may be the case in a system application that the reduction in the group delay obtained by the above algorithms is more than the required amount. Then we can improve the non-linearity in the phase response as follows.

- The root moments corresponding to the stop band transmission zeros remain the same as above.
- From the rest of the zeros we can select an appropriate number in conjugate form, for real transfer functions, in an arbitrary fashion.

Extraction of factors of polynomials that have zeros in certain non-circular regions can also be put into effect by integrating around these regions. Specifically with respect to the filter design problem we can determine the factor corresponding to the zeros on the unit circle by integrating along a segment of the unit circle.

Further work is necessary to explore the options open here.

6. EXPERIMENTAL RESULTS

The following example is a low pass equiripple FIR digital filter of order 80, with linear phase, designed using the Remez algorithm. The pass band occupies the frequency samples from 0 to 204, the transition band from 205 to 225 and the stop band from 226 to 511 samples (with 512 being the Nyquist frequency).

Figures 1 and 2 show the amplitude response and the location of the roots within the z – plane, of the mixed (and linear) phase FIR system transfer function, designed using the Remez algorithm.

Figures 3 and 4 show the amplitude response and the location of the roots within the z – plane, of the minimum phase version of the same system using the proposed designed algorithm.

Figure 5 shows the phase response for both systems.

Figure 6 shows the group delay for both systems.

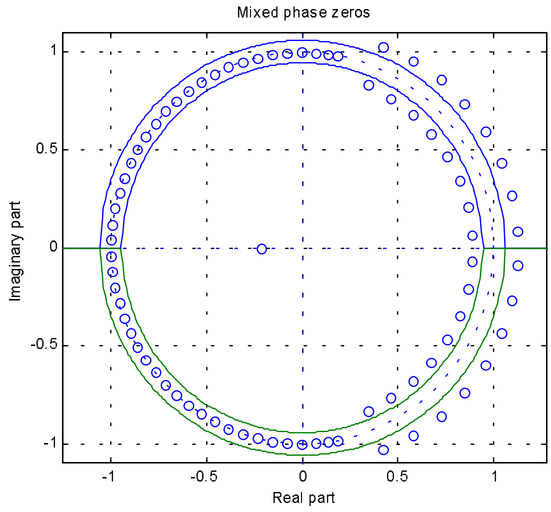

Figure 1

In the above figure, the dotted circle corresponds to the unit circle and the two solid circles to the ones that will be used for the contour integrations in the proposed algorithm.

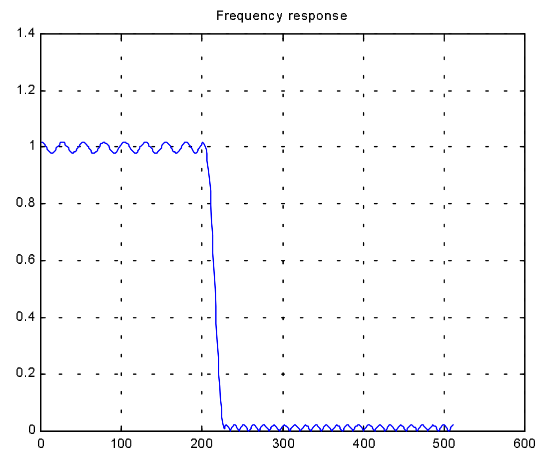

Figure 2: Amplitude response of the original linear phase FIR filter transfer function.

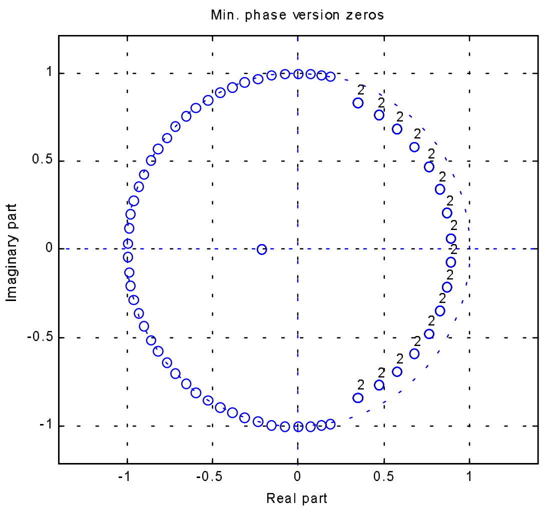

Figure 3

The roots outside the unit circle have inverted inside, exactly on their reciprocal roots. The number two on the figure indicates the double roots.

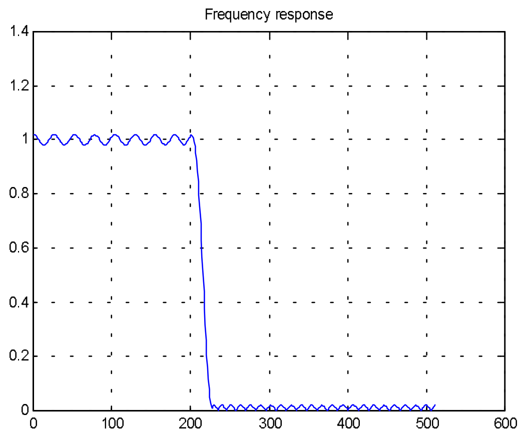

Figure 4

The amplitude response of the reconstructed minimum phase version of the transfer function is almost identical as the one obtained using the Remez algorithm.

The two following graphs show the phase response of the original and the reconstructed filter respectively. The range of frequency samples of interest is the pass band which goes from 0 to 204 samples.

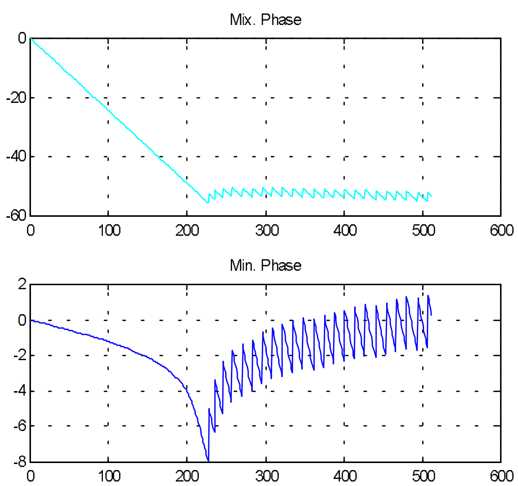

Figure 5

The original filter has a linear phase response as designed by the Remez algorithm. The minimum phase version of it has a non linear phase response as expected.

At the last figure the group delay of the original and the reconstructed filters are shown. Again the range of frequency samples of interest is the pass band. There is a significant decrease of the group delay within the range of frequencies of interest, namely, the passband.

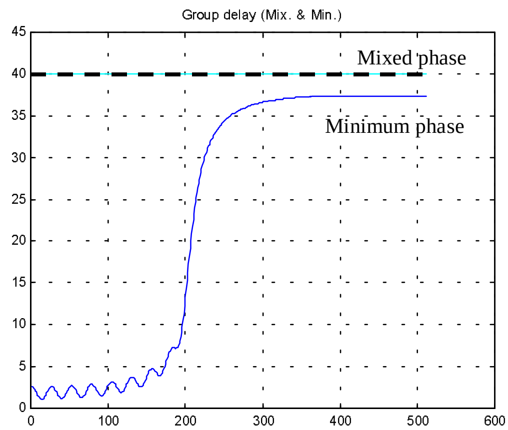

Figure 6

7. REFERENCES

[1] Boite R. and H. Leich, "A New Procedure for the Design of High Order Minimum Phase FIR Digital or CCD FIlters", *Signal Processing*, vol. 3, pages 101-108, 1981.

[2] Herman O. and H.W. Schussler, "Design of Non Recursive Digital Filters With Minimum Phase", *Electronics Letters*, vol. 6, pages 329-330, 1970.

[3] Kamp Y. and Wellekens C.J., "Optimal Design of Minimum Phase FIR Filters", *IEEE Trans. on ASSP*, vol. 31, pages 922-926, 1983.

[4] Mian G.A. and A.P. Naider, "A Fast Procedure to Design Equiripple Minimum Phase FIR FIlters", *IEEE Trans. on Circuits and Systems*, vol. 29, no. 5, pages 327-331, 1982.

[5] Nikias C.L. and A. Petropulu, *Higher Order Spectra Analysis*, Prentice Hall, 1993.

[6] Polydoros A. and A.T. Fam, "The Differential Cepstrum. Definition and Properties", *Proc. IEEE Int. Symposium on Circuits and Systems*, pages 77-80, April 1981.

[7] Stathaki T., "Cumulant Based and Algebraic Techniques for Signal Modelling", *PhD Thesis*, University of London, 1994.

[8] Todhunter I., *Theory of Equations*, McMillan & Co, London 1882.

[9] Tania Stathaki, Anthony Constantinides, Georgios Stathakis, “Equiripple Minimum Phase Piecewise Flat Fir Filter Design From Linear Phase Systems Using Root Moments”, *Proc. of the IEEE ICASSP* 98, 15-19 May, Arizona, USA.