Proc. of the Conference on Digital Signal Processing Florence Italy, Sept, 1984

DESIGN OF DIGITAL FILTER BANKS WITH FLAT COMPOSITE RESPONSE

A. Dembo and D. Malah

Department of Electrical Engineering
Technion - Israel Institute of Technology
Technion City, Haifa 32000, Israel

Two new methods for designing uniform and non-uniform digital filter banks with flat composite response are presented. The individual filters in the bank are FIR (finite impulse response) digital filters with linear phase.
The first method is based on the conventional window method, but obtains a nearly equiripple response of the individual filters in the banks by using an optimal window specifically designed for the task. This window is obtained by properly modifying the coefficients of a differentiator filter designed by the Remez exchange algorithm.
The second method is based on a weighted minimum mean square error criterion, which assures a flat composite response by properly incorporating this requirement into the design process. The new optimal window method is demonstrated by design examples and is compared with earlier known approaches.

1. INTRODUCTION

The conventional window method, for the design of linear phase FIR filters, can be used to achieve a flat composite response for both uniform and non-uniform filter banks [1,2]. However, the magnitude response of the individual filters in the bank is not optimal (in the min-max sense) and a longer filter is needed to meet the magnitude response specifications, as compared with a min-max (equiripple) design [3]. On the other hand, application of the Remez Exchange method [3,4], for the design of equiripple linear phase filters, does not result, in general, in a flat composite response of the filter bank. For non-uniform filter banks an automated trial and error approach of iterated designs using the Remez Exchange method is applicable, but it is practical only for simple filter banks (those with a few wideband filters) [5]. For uniform filter banks one needs only to design a single lowpass prototype filter, and a reasonable approximation to a flat composite response can be obtained by iterating the design several times (with slightly different band edges specification). Alternatively, adopting the approach in [6], every N-th term (except the center one) in the impulse response of the prototype filter can be forced to zero (a constraint that must be satisfied if the uniform filter bank is to have a flat composite response [1]). This causes however, a degradation in the filter performance and longer filters are needed to meet the specifications.

The optimal window method presented in the next section assures a flat composite response. It results in a better magnitude response (in the min-max sense) of the individual filters than the conventional window method. For uniform filter banks it results in a similar performance to the approach in [6], and for non-uniform filter banks it results in a similar performance to the method in [5]. It is however simpler than the method proposed in [5], and can be used for both uniform and non-uniform filter banks (unlike the approach in [6]).

The Weighted Minimum Mean Square Error (WMMSE) criterion was used in [7,8] for the design of bandpass filters. A new design method for non-uniform filter banks, which are composed of optimal WMMSE individual filters, is presented in the last section. This design method assures a flat composite response of the filter bank and allows the individual filters to be of different lengths.

2. THE OPTIMAL WINDOW METHOD

The conventional window method can be used for designing filter banks with flat composite response, by properly setting the cut-off frequencies of the individual filters in the bank [1]. The same window must be used for all the filters in the bank and hence they all have the same length [1].

The main disadvantage of the conventional window method is in the use of classical windows which result in a non-optimal (in the min-max sense) magnitude response of the individual filters.

Our aim is to design a special window that will result in an optimal response of the individual filters in the filter bank. We will describe here in detail the design of the optimal window sequence for odd length filters ($M=2L+1$) only. For even length window sequences a similar method exists.

Since all the filters are designed to have linear phase they have a symmetric impulse response, and the window sequence will be assumed to have the symmetric form: $\{w(L),\ldots,w(1),w(0),w(1),\ldots,w(L)\}$.

An analysis of the magnitude response error in the design of multibandpass/bandstop filters using the window method is given in [9,10]. From this analysis two important results follow:

(A) Proper gain of the filters imposes $w(0)=1$.

(B) For wideband filters (i.e., for which the widths of the pass/stop bands are greater than the width of the transition bands), the magnitude response error near a cut-off frequency of the desired filter (F_c) will have approximately the shape of the function $J_w(|f-F_c|)$, where

$$J_w(f) \triangleq (0.5-f)w(0) - \sum_{n=1}^{L} \frac{w(n)}{\pi n} \sin 2\pi n f \qquad (1)$$

Due to result (B), the maximal deviation in the stop/pass bands, whose edges are at $(F_c \pm \Delta F/2)$, will be:

$$\delta_w(\Delta F) \cong \max_{f \in [\Delta F/2,\, 0.5]} |J_w(f)| \qquad (2)$$

Where ΔF is the width of the transition band.

The optimal window sequence $\{w^*(n)\}_{n=-L}^{L}$ is defined to be the sequence that minimizes $\delta_w(\Delta F)$ over all symmetric window sequences of length $2L+1$ satisfying $w(0)=1$, i.e.,

$$\delta_{w^*}(\Delta F) = \min_{\substack{w \text{ sequences} \\ \text{with } w(0)=1}} \{\delta_w(\Delta F)\} \qquad (3)$$

Combining (1), (2) and (3) results in the following minimization problem

$$\min_{\{w(n)\}_{n=1}^{L}} \{ \max_{f \in [\Delta F/2, 0.5]} |(0.5-f) - \sum_{n=1}^{L} \frac{w(n)}{\pi n} \sin 2\pi n f | \} \qquad (4)$$

This problem is a Chebyshev approximation problem, and can be solved either by means of linear programming or by the Remez exchange method.

The solution of the above problem gives an equiripple response to $J_w(f)$ in the frequency band $[\Delta F/2, 0.5]$. Since $J_w(f)$ is only an approximation of the exact magnitude response error, the designed filter will only nearly have an equiripple response.

By a simple change of variables the minimization problem in (4) becomes the design problem of an odd length optimal (min-max) differentiator. This latter problem can be solved using the well known program of McClellan, Parks and Rabiner in [4]. The change of variables is as follows:

$$f = 0.5-\vartheta \ ; \ F_H \triangleq 0.5 - \Delta F/2 \qquad (5)$$

$$w(n) = 2\pi n(-1)^{n+1} a(n) \qquad n=1,2,\ldots,L \qquad (6)$$

With this change of variables, (4) becomes:

$$\min_{\{a(n)\}_{n=1}^{L}} \{ \max_{\vartheta \in [0, F_H]} |\vartheta - 2\sum_{n=1}^{L} a(n) \sin 2\pi n \vartheta | \} \qquad (7)$$

This is exactly the design problem of an odd length differentiator, with a desired slope of unity, and a weight function of unity for $|\vartheta| \le F_H$ and zero elsewhere. The first L coefficients in the impulse response of the differentiator are $\{a(L),\ldots,a(1)\}$ and a simple modification of them (according to (6)) gives the optimal window sequence $\{w^*(n)\}_{n=-L}^{L}$ (with $w^*(0)=1$ due to result (A) above).

The deviation of the differentiator filter from the ideal differentiator (which is an output of the program in [4]), is exactly $\delta_{w^*}(\Delta F)$, and this is (approximately) the deviation of the individual filters in the filter bank from the ideal response.

To illustrate the design of a uniform filter bank we consider a bank of 16 filters, having each an impulse response of length 123 samples. The sampling rate is 8kHz, so that each individual filter is to have a bandwidth of 500Hz. The transition band width is specified to be 137.5Hz. Table I compares the design results using four different design methods with three different values of $k=\delta_p/\delta_s$.

It is seen that while the method denoted by DMX has the smallest values of δ_p, δ_s, it results in a poor composite response, especially for $k \gg 1$. Therefore, DMX is preferred only for $k \approx 1$. However, in this case, TMX, OW, and KW have values of δ_p, δ_s no larger than twice those of DMX. For $k \gg 1$, TMX and OW have similar performance, while KW is inferior to them. Note that TMX is applicable only to the design of uniform filter banks and for a limited range of k $((N-1) \ge k \ge 1$, see [8]).

For non-uniform filter banks both KW and OW can be used. However, OW is preferred on KW. In Ref [5] there are two examples of designing non-uniform filter banks using an iterative approach. The proposed OW method which requires no iterations achieved the same performance using the same filter length. Figs. 1 to 5 illustrate the shape of the frequency response of the prototype filter which results when the TMX, OW, KW, and DMX methods are used. All the figures refer to the design example in Table I, and for $k=10$. Fig. 1 shows the response of the equiripple filter designed by DMX, as well as the composite response ripple of the designed filter bank. The column under A_c in Table I refers to the composite response ripple obtained with DMX, which is here 3dB. Fig. 2 is the response of the filter designed with TMX. The frequency bands around $(0.5+m)/N$, $m=1,\ldots,N-2$, (i.e. the transition bands of the other filters in the bank) are excluded from the stopband, and therefore, in these bands, the attenuation is quite low (in this example it is about 30dB; see also the column under $\tilde{A}_s$ in Table I). Figs. 3 and 4 illustrate the shape of the filters designed with KW and OW methods, respectively. Note that for $f \gg F_c$, the deviation with KW drops below -80dB, while with OW it remains about -46dB. Fig. 5 shows the responses of the filters designed by KW and OW methods at the frequency band near the transition band. It is seen that at the edge of transition band OW results in an attenuation of 46dB where with KW it is only 38dB.

Note: the optimal window design method presented above refers to $k=1$ only (i.e. $\delta_p=\delta_s$). For $k \ne 1$ similar equations result. However, if $k \ne 1$ the differentiator (in (4)) has two bands, having the same slope but different weight function values. The cut off frequencies of these bands are related to k via the empirical design equations in [9,10].

3. THE WMMSE METHOD FOR FILTER BANKS

In the previous section we proposed a new design method and compared it with other known methods for filter bank design. In those methods the performance of the individual filters in the filter bank was measured under the min-max criterion. A direct method for optimal (in the min-max sense) filter bank design, satisfying the constraint of flat composite response, is not known. Therefore, those design methods were sub-optimal methods. We shall now introduce another performance measure for the individual filters in the filter bank, and interpret it. Then, we shall present a direct optimal design method under that measure, combined with the flat composite response constraint. The performance measure of Weighted Minimum Mean Square Error (WMMSE) was used in [7,8] in the design of multi-bandpass/bandstop FIR filters. This performance measure was interpreted either statistically [8] or deterministically [7]. For the individual filters in the filter bank, the classical deterministic interpretation of the WMMSE criterion is as follows:

The WMMSE measure seeks to minimize the weighted energy of the magnitude response error of each individual filter. The weight function sets the relative importance of the error in the stopbands and in the passband. The transition bands are assumed to be of no importance, and hence are given a weight of zero. Minimization of the WMMSE measure gives therefore the best separation of bands in the filter bank in terms of residual energies.

The statistical interpretation of the WMMSE criterion in filter banks is as follows:

Each individual filter refers to the input signal frequency components in its passband as the desired signal, and to the frequency components in its stopbands as the noise. Since the transition bands are of no importance, the frequency components in these bands are ignored. If both the desired signal and the noise are assumed to be samples of two uncorrelated, zero-mean, wide-sense stationary continuous random processes, then the optimal WMMSE filter will maximize the signal-to-noise output ratio (SNR).

The setting of the weight function is based on the a-priori knowledge of the relative magnitude of the desired signal in the input in reference to the magnitude of the noise in the input.

We shall now state the design problem. The filter bank is composed of N individual filters. The filters are FIR filters with linear phase, and impulse responses $\{\underline{a}_i\}_{i=1}^{N}$ of lengths $\{M_i\}_{i=1}^{N}$. All filters have real coefficients ($\underline{a}_i \in R^{M_i}$), although generalization to complex filter banks is easy. If the filters are of non-equal length, a proper delay is assumed to be inserted in each channel, so that the overall delay in all channels is the same. The composite response of the filter bank is the sum of the responses of all N channels, and it ought to be flat, in some frequency band of interest. The WMMSE performance measure of the individual filters in the filter bank is given by:

$$\delta_i^2 = \int_{-0.5}^{0.5} W_i(f)^2 |D_i(f) - H_{a_i}(f)|^2 df \qquad i=1,\dots,N \quad (8)$$

Where $W_i(f) \geq 0$ is the weight function, $D_i(f)$ is the desired frequency response of this filter and $H_{a_i}(f)$ is the actual frequency response, related to the impulse response $\underline{a}_i$ (and including the inserted additional delay). The composite response performance measure is defined in a similar manner by:

$$\delta_{N+1}^2 = \int_{-0.5}^{0.5} W_{N+1}(f)^2 |D_{N+1}(f) - H_{a_{N+1}}(f)|^2 df \qquad (9)$$

Where $W_{N+1}(f) \geq 0$ is the weight function related to the composite response, and $D_{N+1}(f)$ is the desired composite response (here flat response). $H_{a_{N+1}}(f)$ is the actual composite response of the filter bank, and is given by:

$$H_{a_{N+1}}(f) = \sum_{i=1}^{N} H_{a_i}(f) \qquad (10)$$

Equations (8) and (9) define ($N+1$) different WMMSE design problems, but due to (10) these problems are tied together and must be solved simultaneously.

In order to complete the specifications, each of the ($N+1$) different problems is given a relative weight $K_i \geq 0$, $i=1,2,\dots,N+1$. The optimal WMMSE filter bank is given by the set $\{\underline{a}_i^*\}_{i=1}^{N}$ of impulse responses that achieves:

$$\min_{\{\underline{a}_i\}_{i=1}^{N}} \left(\sum_{i=1}^{N+1} K_i^2 \delta_i^2\right) \qquad (11)$$

The detailed solution of (8)-(11) is given in the Appendix. Three different types of solutions exist, as following:

(A) When $K_{N+1}=0$ there is no specification on the composite response and each individual filter is designed separately using the standard WMMSE method [7,8].

(B) For $K_{N+1}=\infty$, if the specification on the composite response can be fulfilled by any filter bank composed of FIR linear phase filters (i.e., if $\delta_{N+1}=0$ is possible for at least one set $\{\underline{a}_i\}_{i=1}^{N}$), it will be fulfilled by the solution $\{\underline{a}_i^*\}_{i=1}^{N}$.

(C) For a positive, finite value of K_{N+1}, the composite response specification is just an additional performance measure. The solution will not necessarily be with the desired composite response. However, it will probably have a composite response which is close to the desired one under the WMMSE criterion.

The application of this method to the design of various non-uniform filter banks is now under examination.

References

[1] L.R. Rabiner & R.W. Schafer, Digital Processing of Speech Signals, Prentice Hall, N.J., 1978 (Ch 6)

[2] R.W. Schafer, L.R. Rabiner & O. Herrman, "FIR Digital Filter Banks for Speech Analysis", BSTJ Vol. 54, No. 3, March 1975, pp 531-544.

[3] L.R. Rabiner & B. Gold, Theory and Application of Digital Signal Processing, Prentice Hall, N.J., 1975 (Ch.3).

[4] J.H. McClellan et al., "A Computer Program for Designing Optimum FIR Linear Phase Digital Filters", IEEE on Audio and Electroacoustics Vol. AU-21, No. 6, Dec. 1973, pp. 506-526.

[5] J.T. Rubinstein & H.F. Silverman, "Some Comments on the Design & Implementation of FIR Filterbanks for Speech Recognition", ICASSP 83, pp. 812-815.

[6] F. Mintzer, "On Half-Band, Third-Band and N-th-Band FIR Filters and their Design", IEEE Trans on Acoustics, Speech and Signal Processing, ASSP-30, No. 5, Oct. 1982, pp. 734-738.

[7] H. Clergot & L.L. Sharf, "Connections Between Classical & Statistical Methods of FIR Digital Filter Design", IEEE Trans. on Acoustics, Speech and Signal Processing, ASSP-26, No. 5, Oct. 1978, pp. 463-465.

[8] D.C. Farden & L.L. Sharf, "Statistical Design of Non-Recursive Digital Filters", IEEE Trans. on Acoustics, Speech and Signal Processing, ASSP-22, No. 2, June 1974, pp. 188-196.

[9] A Dembo & D. Malah, "Generalization of the Window Method for FIR Digital Filter Design", IEEE Proc. 13th Convention of Electrical & Electronics Engineers in Israel, Tel-Aviv, March 1983, pp. 4.4.2-1 to 4.4.2-4.

[10] A Dembo & D. Malah, "Generalization of the Window Method for FIR Digital Filter Design", Technion EE Pub. No. 456, Aug. 1983 (47pp).

APPENDIX

Here we shall derive the solution of the WMMSE design problem formulated in Eqns. (8)-(11). First we introduce some notation:
Let

$$M_{N+1} \triangleq \max_{1\le i\le N}\{M_i\} \tag{A1}$$

The augmentation of a vector $\underline{v} \in R^l$, denoted by $\underline{v}_a \in R^{M_{N+1}}$ $(l \le M_{N+1})$, is composed of $\underline{v}$ elements in the center l coordinates and zeros at the edges.

The $M_{N+1} \times M_{N+1}$ augmented matrix of R (an $l \times l$ matrix), using R in the l rows and columns around the center, surrounded by zeros, is denoted by R_a.

$\underline{v}|_l \in R^l$ is the reduction of $\underline{v} \in R^{M_{N+1}}$, $(l \le M_{N+1})$, taking only the l elements of $\underline{v}$ around the center ($M_{N+1}/2$ coordinate).
Define the following values (for $i=1,\ldots,N+1$):

$$\alpha_i^2 \triangleq \int_{-0.5}^{0.5} W_i(f)^2 |D_i(f)|^2 df \tag{A2}$$

$$d_i(k) \triangleq \int_{-0.5}^{0.5} W_i(f)^2 D_i(f) e^{j2\pi fk} df, \quad k=0,1,\ldots,M_i-1 \tag{A3}$$

$$R_i(k,m) \triangleq \int_{-0.5}^{0.5} W_i(f)^2 e^{j2\pi f(k-m)} df, \tag{A4}$$

$$k,m=0,1,\ldots,M_i-1$$

Let $\underline{d}_i \in R^{M_i}$ be the vector whose elements are $d_i(k)$ and R_i be $M_i \times M_i$ matrix whose elements are $R_i(k,m)$. Now we define:

$$\delta_{i_o}^2 \triangleq \alpha_i^2 - \underline{d}_i^T R_i^{-1} \underline{d}_i \tag{A5}$$

Under the following transformation of variables:

$$\begin{cases} \underline{a}_i = R_i^{-1}\underline{d}_i + \underline{b}_i & i=1,\ldots,N \\ \underline{a}_{N+1} = R_{N+1}^{-1}\underline{d}_{N+1} - \underline{b}_{N+1} \end{cases} \tag{A6}$$

The WMMSE criterion for each filter is:

$$\delta_i^2 = \delta_{i_o}^2 + \underline{b}_i^T R_i \underline{b}_i \qquad i=1,\ldots,N+1 \tag{A7}$$

The dependence of the $(N+1)$ optimization problems is manifested by the fact that $\underline{b}_{N+1}$ is a linear combination of $\underline{b}_i$ $1 \le i \le N$ (since $\underline{a}_{N+1}$ represent the composite response) or.

$$\sum_{i=1}^{N+1} \underline{b}_{i_a} = \underline{u} \tag{A8}$$

$$\underline{u} = -\sum_{i=1}^{N} R_{i_a}^{-1}\underline{d}_{i_a} + R_{N+1}^{-1}\underline{d}_{N+1} \tag{A9}$$

The optimal WMMSE filter bank is derived by the solution of:

$$\min_{\{\underline{b}_i\}_{i=1}^{N+1}} \sum_{i=1}^{N+1} K_i^2 \underline{b}_i^T R_i \underline{b}_i \tag{A10}$$

subject to (A8). If at least one K_{i_o} (with $M_{i_o} = M_{N+1}$) equals zero, we can minimize (to zero) the criterion (A10) by setting all $\underline{b}_i$-s (except of $\underline{b}_{i_o}$) to zero, while $\underline{b}_{i_o} = \underline{u}$ satisfies (A8). For $i_o = (N+1)$ this case coincides with the conventional WMMSE design in [7], [8]. When all K_i are positive, a direct solution of (A10) is.

$$\underline{b}_i = \frac{1}{K_i^2} R_i^{-1} \underline{v}|_{M_i} \qquad i=1,\ldots,(N+1) \tag{A11}$$

$$\underline{v} = P^{-1}\underline{u} \tag{A12}$$

where:

$$P \triangleq \sum_{i=1}^{N+1} \frac{1}{K_i^2} R_{i_a}^{-1} \tag{A13}$$

We assumed that $\{R_i\}_{i=1}^{N+1}$ as well as P are non-singular (for linear phase filter bank $\{R_i\}_{i=1}^{N+1}$ are p.s.d. When all of them are p.d., P is also p.d. and non-singular). For $K_i \to \infty$, $i \le N$, the i-th filter in the bank will be optimal regardless of the composite response shape ($\underline{b}_i = 0$). Whereas for $K_{N+1} \to \infty$ the composite response error will be the minimal attainable by FIR filter bank $(\delta_{(N+1)_o}^2)$, while possibly increasing the individual filters errors.

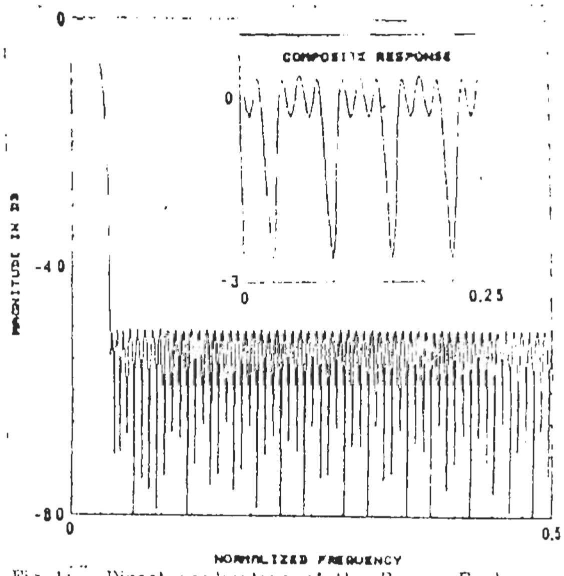

Fig. 1: Direct application of the Remez Exchange method (DMX):

(a) Frequency response of lowpass prototype filter.
(b) Composite frequency resonse.

Table I

Uniform Filterbank Design

$K = \delta_p/\delta_s$;

A_p, A_c - Ripple in dB;

$A_s, \tilde{A}_s$ - Attenuation in dB.

	Kaiser Window		Optimal Window				min-max					
	Conventional Method [1]		$\delta_w^*(\Delta F)$ from eqn (3)		Actual Deviations		"Truncated" [6]			"Direct" [4]		
	KW		OW^e		OW		TMX			DMX		
	A_p	A_s	A_p	A_s	A_p	A_s	A_p	A_s	$\tilde{A}_s$	A_p	A_s	A_c
K = 1	0.22	38.10	0.13	42.65	0.25	38.45	0.14	41.65	41.57	0.14	41.68	0.03
K = 10	0.22	38.10	1.47	47.46	1.10	46.68	1.03	45.66	30.25	0.54	50.08	4.61
K = 50	0.22	38.10	3.40	54.27	3.09	51.22	-	-	-	1.08	58.18	8.73

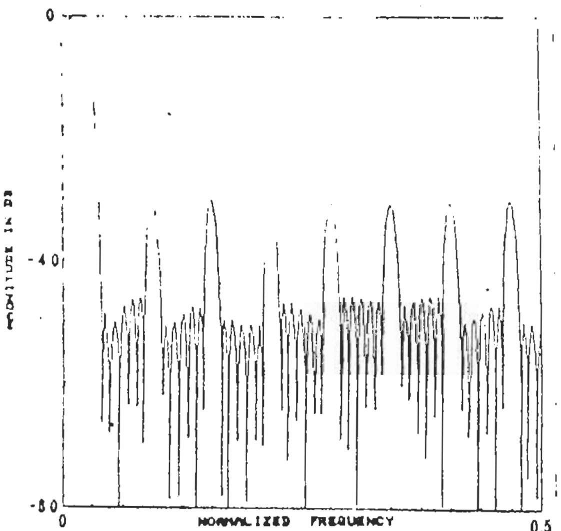

Fig. 2: Frequency response of lowpass prototype filter using the TMX method (forcing every N-th term in the impulse response to zero).

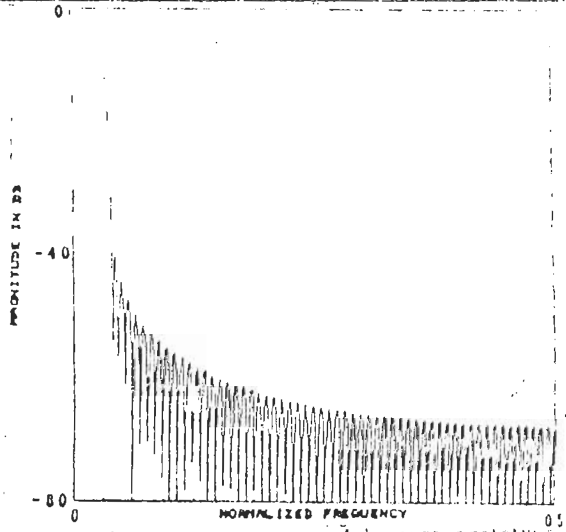

Fig. 3: Frequency response of lowpass prototype filter using the conventional window method with a Kaiser window (KW method).

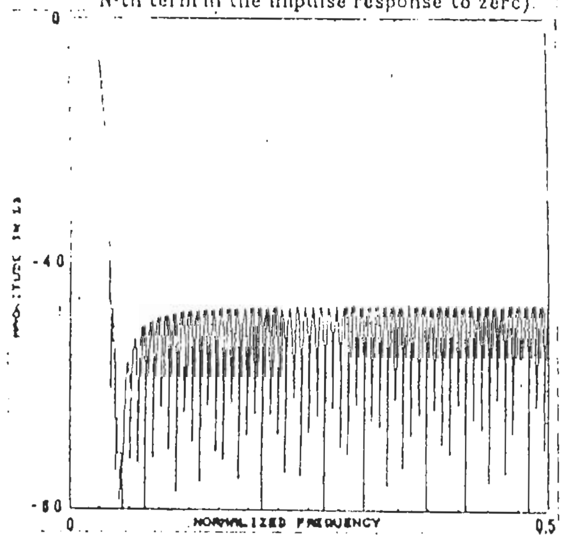

Fig. 4: Same as Fig. 3 but for the Optimal window method (OW).

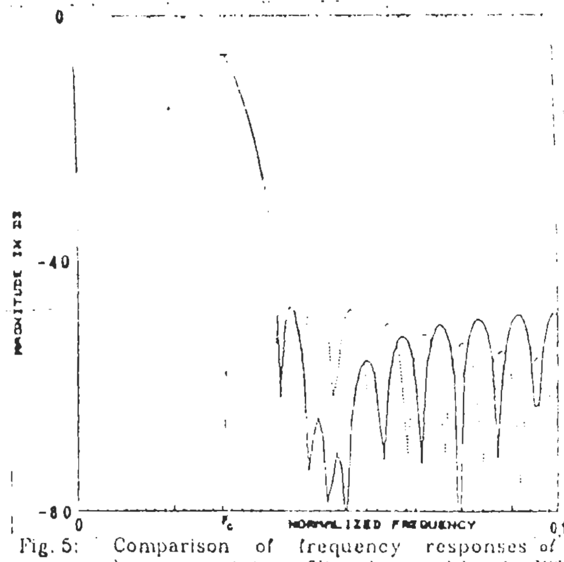

Fig. 5: Comparison of frequency responses of lowpass prototype filters designed by the KW and OW methods.