Rapid Publication

Frequency synthesis of digital filters based on repeatedly applied unweighed moving average operations

Th. J. C. Faes[1] **H. G. Govaerts**[1] **B. J. Tenvoorde**[1] **O. Rompelman**[2]

[1] Laboratory of Medical Physics and Informatics, Vrije Universiteit, Amsterdam, Netherlands
[2] Laboratory of Information Theory, Technical University Delft, Netherlands

Abstract—*Simple formulae are presented for designing filters based on repeatedly applied moving average operations with unit coefficients. Design formulae are derived to synthesise the filter in a way that satisfies specified passband and stopband specifications. These filters are attractive because of the reasonable frequency characteristics, the computational efficiency of the design and filter algorithms, and the uncomplicated implementation in software.*

Keywords—*Digital filter, Moving average filter*

Med. & Biol. Eng. & Comput., 1994, 32, 698–701

1 Introduction

ALTHOUGH THE low-pass frequency property of moving average filters with unit coefficients (rectangular window) is generally known, the attractiveness of a *cascade* (multiple use) of these filters is underestimated, in particular

(i) the reasonable frequency characteristics due to the repeated application of the moving average operation.
(ii) the easy implementation due to the simple design formulae and filter algorithm.
(iii) the computational efficiency due to the small number of multiplications involved, and moreover, the strongly reduced total number of arithmetical operations due to a recursive formulation of the algorithm.

At present, the serviceableness of these filters is limited as formulae to design these filters from passband and stopband specifications are not widely available. Therefore, our aim is to discuss simple design formulae for synthesising filters based on repeatedly applied unweighed moving average operations in a way that satisfies passband and stopband specifications.

2 Design problem

In the unweighed moving average operation, the kth value of the output sequence $y[n]$ is produced by taking the average of $M + 1$ input values of the input sequence $x[n]$ in the vicinity of the kth input value. By repeating this operation N times, the *repeatedly* applied unweighed moving average filter is constituted. The transfer function $H(z)$ is

$$H(z) = \frac{1}{(M+1)^N}\left[\sum_{m=-M/2}^{M/2} z^m\right]^N \quad (1)$$

Consequently, the filter is completely defined with parameters N and M, where N is the number of times the average operation is repeated and $M + 1$ is the number of input points used in each average operation. Note that M is even.

The frequency response $|H(f)|$ of the filter is found by bringing the sum in a closed-form by using geometric series, and by substitution of $z = e^{-j2\pi fT}$, where T is the sampling interval (RABINER and GOLD, 1975; OPPENHEIM and SCHAFER, 1989). Hence, the magnitude $|H(f)|$ is

$$|H(f)| = \frac{1}{(M+1)^N}\left|\frac{\sin \pi (M+1) fT}{\sin \pi f T}\right|^N \quad (2)$$

and the phase is zero. The interpretation is restricted to the frequency range $(0, f_\sigma/2)$, where $f_\sigma = T^{-1}$.

Fig. 1 illustrates the dependence of $|H(f)|$ on N and M. The log magnitude is plotted against the normalised frequency fT at the interval 0–0·5. At low frequencies the magnitude reaches unity, and the attenuation increases strongly at higher frequencies. Clearly, the stopband attenuation is efficiently increased by a larger N, and the passband edge is shifted to lower frequencies by a larger M.

Although the dependence of $|H(f)|$ on N and M is illustrated in Fig. 1, we need a way of performing the inverse to synthesise the filter, i.e. getting N and M from a specified filter response.

3 Synthesis

The frequency properties of a filter are commonly specified in terms of attenuation and frequency location of

Correspondence should be addressed to Th.J.C. Faes, Laboratory of Medical Physics and Informatics, AZVU-Poli, De Boelelaan 1118, 1081 HV Amsterdam, The Netherlands.

First received 11 May 1993 and in final form 27 May 1994

This paper was originally submitted before the introduction of the Rapid Publication section.

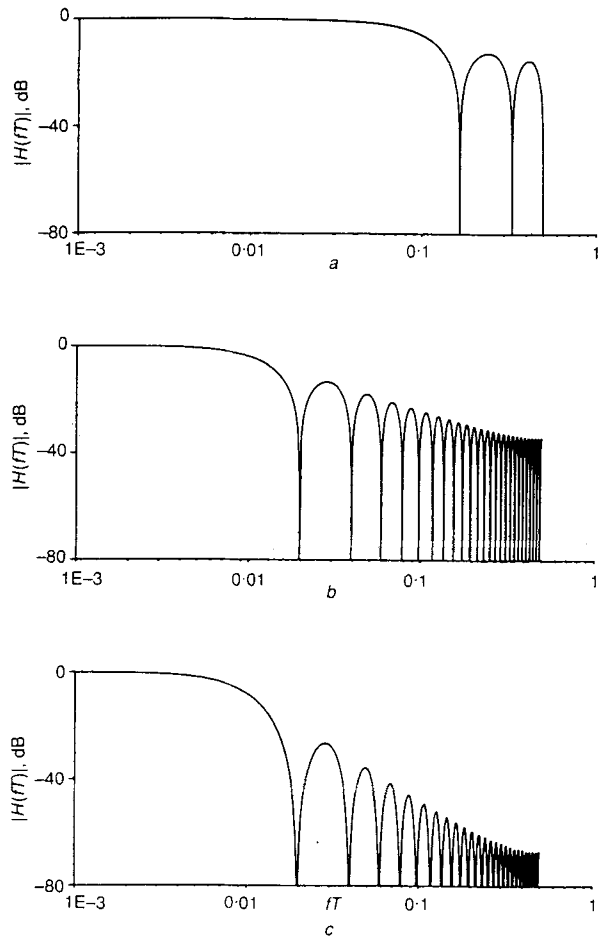

Fig. 1 *Typical examples of the log-magnitude frequency responses of N times repeatedly applied M + 1 unweighed moving average filter; (a) M = 5, N = 1; (b) M = 50, N = 1; (c) M = 50, N = 2*

the passband and stopband. In the case of low-pass filters (Fig. 2), the passband is specified by the passband edge f_p and the allowed passband ripple d_p (deviation from unity), whereas the stopband is specified by the stopband edge f_s and the required stopband attenuation d_s (deviation from zero). The passband and stopband are separated by the transition band. Hence, the design problem is the determination of the filter parameters N and M, given these passband and stopband specifications.

3.1 *Passband*

The normalised passband frequency f_pT is defined as the frequency where the response $|H(fT)|$ equals the specified passband ripple $1 - d_p$. Hence (eqn. 2)

$$|H(f_pT)| = \frac{1}{(M+1)^N}\left|\frac{\sin \pi(M+1)f_pT}{\sin \pi f_pT}\right|^N = 1 - d_p \quad (3)$$

To write the passband frequency f_pT explicitly, Taylor expansions of both sinus functions are used. Hence

$$|H(f_pT)| = \frac{1}{(M+1)^N}\left|\frac{\pi(M+1)f_pT - \frac{(\pi(M+1)f_pT)^3}{3!} + \frac{(\pi(M+1)f_pT)^5}{5!} - \ldots}{\pi f_pT - \frac{(\pi f_pT)^3}{3!} + \frac{(\pi f_pT)^5}{5!} - \ldots}\right| \quad (4)$$

To solve f_pT from eqn. 4, the first step is to estimate the number of terms contributing substantially in both Taylor expansions. For that, f_pT is approximated by substitution of the inequality

$$\left|\frac{\sin \pi(M+1)f_pT}{\sin \pi f_pT}\right|^N \leq \frac{1}{|\sin \pi f_pT|^N} \quad (5)$$

in eqn. 3, which yields

$$(1 - d_p) \leq \frac{1}{(M+1)^N}\frac{1}{|\sin \pi f_pT|^N} \quad (6)$$

Solving for f_pT provides a first-order approximation

$$f_pT \leq \frac{1}{\pi}\arcsin\left(\frac{1}{(M+1)\sqrt[N]{1-d_p}}\right) \approx \frac{1}{\pi}\frac{1}{M+1} \quad (7)$$

where the right-hand side holds, because an arcsin function approximates its argument for small values and the root is close to unity. Substitution of this approximation in eqn. 4 shows that the terms of both alternating series decrease rapidly. In fact, truncation of the series in the denominator after the first term and in the numerator after the second term results in errors of less than 1% for all $M > 1$ (For alternating series, the truncation error is less than the first term omitted). Consequently, eqn. 4 is reduced to

$$|H(f_pT)| = (1 - d_p) \approx \frac{1}{(M+1)^N}\left|\frac{\pi(M+1)f_pT - \frac{(\pi(M+1)f_pT)^3}{3!}}{\pi f_pT}\right|^N \quad (8)$$

Solving for f_pT yields an accurate approximation

$$f_pT \approx \frac{\sqrt{6}}{\pi(M+1)}\sqrt{1 - \sqrt[N]{1-d_p}} \quad (9)$$

As expected from Fig. 1, f_pT is more sensitive to M than N.

3.2 *Stopband*

The normalised stopband frequency f_sT is defined as the lowest frequency where the frequency response $|H(fT)|$

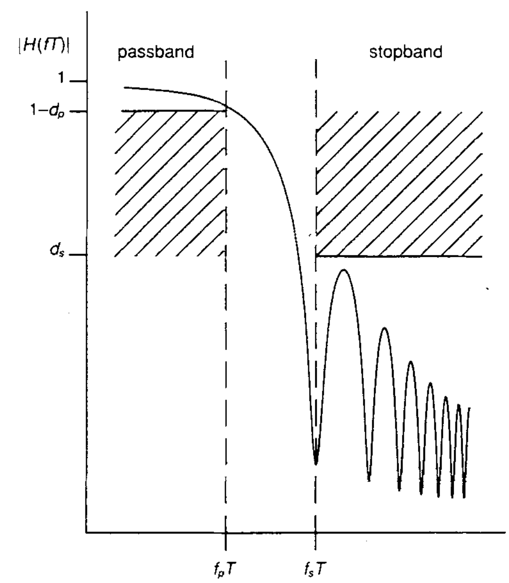

Fig. 2 *Specification of low-pass filter by ripple d_p and edge f_pT of passband and ripple d_s and edge f_sT of stopband; solid line represents a possible realisation*

equals zero, and the amplitude of the first side lobe is defined as equal to the required attenuation d_s (Fig. 2).
Hence

$$f_s T = \frac{1}{(M+1)} \tag{10}$$

because the numerator in eqn. 2 is zero at this frequency.
The maximum value of the first side lobe of $|H(fT)|$ is found at

$$f_m T = \frac{3}{2(M+1)} \tag{11}$$

because the numerator in eqn. 2 is 1 at this frequency. The amplitude of the first side lobe is found by substitution of eqn. 11 into eqn. 2, which yields

$$|H(f_m T)| = \frac{1}{(M+1)^N} \frac{1}{\left|\sin \frac{3\pi}{2(M+1)}\right|^N} = d_s \tag{12}$$

The independence of $|H(f_m T)|$ from M is shown by taking the limit $M \to \infty$ and using the Taylor expansion of the sinus function in eqn. 12. Hence

$$\lim_{M\to\infty} |H(f_M T)|$$

$$= \lim_{M\to\infty} \frac{1}{(M+1)^N} \frac{1}{\left|\frac{3\pi}{2(M+1)} - \frac{(3\pi)^3}{3!(2(M+1))^3} + \dots\right|^N}$$

$$= \left(\frac{2}{3\pi}\right)^N \tag{13}$$

For finite values of M, the right-hand side of eqn. 13 is an accurate approximation (error $< 5\%$ for $M > 6$). Consequently

$$|H(f_m T)| = d_s \approx \left(\frac{2}{3\pi}\right)^N \tag{14}$$

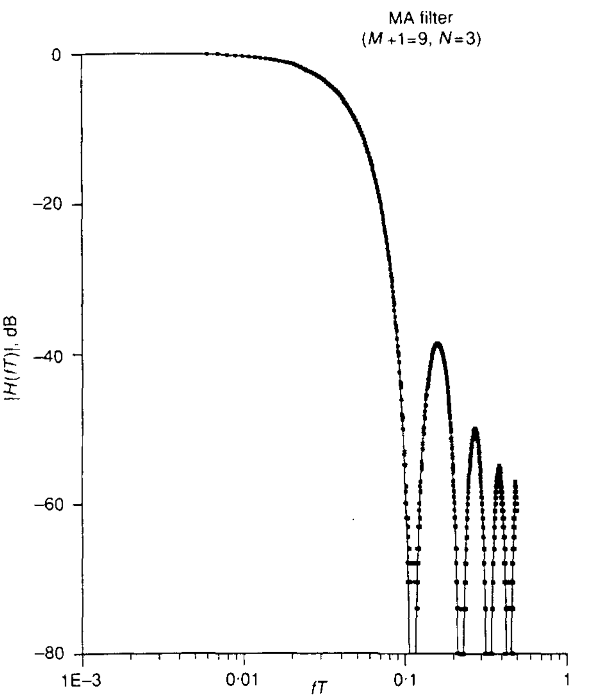

Fig. 3 *Log magnitude of filter response ($N = 3$, $M = 8$); estimated repsonse (■) from filtered sweep signal using cross-spectral techniques and theoretical response (solid line) from eqn. 2*

which is an accurate relation between N and d_s for $M > 6$. Note that d_s depends solely on N.

3.3 *Design formulae*

The filter parameters N and M can be calculated from the specified passband (f_p, d_p) and stopband (f_s, d_s) by inverting eqns. 9 and 14. More specifically, N follows from eqn. 14

$$N \geq \frac{\log(d_s)}{\log\left(\frac{2}{3\pi}\right)} \approx -1.5 \log(d_s) \tag{15}$$

whereas M follows from eqn. 9

$$M \leq \frac{1}{\pi f_p T} \sqrt{6(1 - \sqrt[N]{1 - d_p})} - 1 \tag{16}$$

Finally, eqn. 10 provides a check on the design by calculating the *realised* stopband frequency. This value should be smaller than the specified stopband frequency edge $f_s T$. In this case, the transition band is smaller than required, and consequently the filter performance is better than specified. If, however, the realised frequency exceeds the specified frequency, then the desired frequency response cannot be realised by this filter. In practice, the design formulae behave well if the transition band is allowed to be wider than about half a frequency decade, and the required passband frequency f_p is smaller than about 0·1 times the sample frequency.

So far, only low-pass filters have been discussed. High-pass filters, however, are easily obtained by subtracting a low-pass filter response from 1, i.e. $H_{high}(f) = 1 - H_{low}(f)$. The design formulae for this filter are found by substitution of the passband and stopband specifications of the low-pass filter in eqns. 10 and 14, i.e. the filter parameters N and M are found from $N = -1{\cdot}5 \log(1 - d_p)$ and $M + 1 = (f_p T)^{-1}$, where d_p and f_p are the passband specifications of the high-pass filter. Bandpass and bandstop filters can be designated by a combination of low- and high-pass filters.

4 Examples

The accuracy of the design formulae is demonstrated in an example. A filter is specified by an attenuation of at least 40 dB above 0·2 Hz and a distortion of less than 3 dB below 0·029 Hz (sample rate 1 Hz). Hence, the filter specifications are $f_p T = 0{\cdot}029$, $1 - d_p = 0{\cdot}7$ (passband), $f_s T = 0{\cdot}2$ and $d_s = 0{\cdot}01$ (stopband). Applying eqns. 15 and 16 yields the parameters $N = 3$ and $M = 8$. As the realised stopband frequency at 0·111 (eqn. 10) is less than the desired value of 0·2, the filter is found to be realisable.

The filter response was estimated empirically from a filtered signal (linear frequency sweep 0·005–0·5 Hz). The transfer function $|H(f)|$ was calculated by cross-spectrum analysis (rectangular window, no interpolation or smoothing). Input and output signals consisted of 2048 samples each. For reasons of numerical accuracy, the filter operation was performed on an 8 byte floating point representation.

Fig. 3 shows the estimated and theoretical responses obtained from eqn. 2. As required, the distortion is less than 3 dB below 0·029, and an attenuation of 40 dB is realised above 0·2. This example clearly demonstrates that the filter is accurately designed with the formulae 15, 16 and 10.

A biomedical application of the moving average (MA) filter is taken from the field of heart rate variability (HRV), i.e. the study of fluctuations in heart rate to provide non-invasive tools for the investigation of the autonomic nervous system (LATSON, 1994). The basic HRV signal is derived from the electrocardiogram (ECG) by the detection of the QRS complexes and by the measurement of the time between the successively QRS complexes (KOELEMAN, 1984; ROMPELMAN, 1987). QRS complexes are commonly detected, off-line, by searching the maximum in the ECG for each heart beat. This approach, however, is hindered by the fact that the ECG shows

(*a*) large low-frequency drift phenomena.
(*b*) large P- and T-waves.
(*c*) high-frequency noise.

The moving average filter can remove these three disturbances effectively.

Fig. 4 (unfiltered ECG) shows an ECG recording, which was sampled at 1000 Hz to allow an QRS occurrence estimation at an accuracy of 1 ms. Large drift phenomena and T-waves are clearly visible, by which the detection of the QRS complexes would fail. Fourier anlaysis of the ECG (not shown) indicated, however, that the drift phenomenon, as well as the P- and T-waves, could be reduced by a high-pass filter ($f_p = 15$ Hz, $d_p = 0{\cdot}3$), and a low-pass filter ($f_p = 20$ Hz, $d_p = 0{\cdot}3$, $d_s = 0{\cdot}001$) could reduce the noise. Moving average filters were designed, i.e. high-pass filter: $N = 1$, $M = 66$; low-pass filter: $N = 5$, $M = 8$. The filtered ECG in Fig. 4 shows that noise and drift phenomena, as well as the P- and T-waves, are effectively reduced. The subsequent detection of the QRS complexes should be trouble-free.

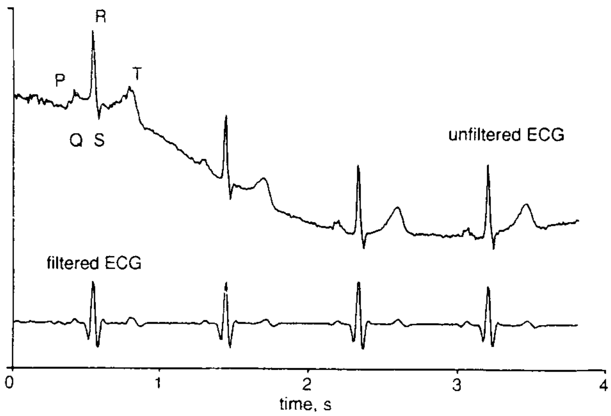

Fig. 4 *Unfiltered and filtered electrocardiogram recordings; dominant low-frequency drift phenomenon and T-waves, as well as high frequency noise and P-waves, effectively reduced by moving average filter; low-pass: $N = 5$, $M = 8$; high-pass: $N = 1$, $M = 66$*

5 Discussion

The N times repeatedly applied $M + 1$ unweighed moving average filters have been discussed. Uncomplicated and straightforward formulae were derived to calculate the filter parameters N and M from specified frequency edges and ripples of the passband and stopband.

The main advantage of the repeatedly applied unweighed moving average filter, as a special case of finite impulse response (FIR) filters, is the computational efficiency in combination with a reasonable frequency characteristic. In general, FIR filters are realised by moving average operations with non-unity coefficients. From a computational point of view, these coefficients might be discerned in real (RABINER and GOLD, 1975), integer (WILCOCK and KIRSNER, 1968; LYNN, 1977; THAKOR and MOREAU, 1987), powers-of-two (YONG CHING LIM and PARKER, 1983; QUANGFU ZHAO and TAKADORO, 1988) and unity coefficients. A $(M + 1)$-point filter requires, for each sample, M accumulations, in addition to $M + 1$ floating point multiplications for real coefficients; $M + 1$ fixed point multiplications for integer coefficients; and $M + 1$ shift operations for power-of-two coefficients; but only one real multiplication, for scaling, is required for unity coefficients. Thus, the use of unity coefficients increases computational efficiency.

Computational efficiency might be further increased by a recursive formulation of the filter algorithm with unity coefficients. In this case, only two accumulations per sample are required for each repetition of the unweighed moving average operation. Hence, the N times repeatedly applied $M + 1$ moving average filter requires only $2N$ accumulations and one real multiplication per sample (i.e. $(M + 1)^{-N}$, eqn. 1) for scaling. In addition to the computational efficiency of the filter algorithm, the algorithms for designing the repeatedly applied unweighed moving average filter (eqns. 15 and 16) are very simple to implement and very fast to compute in comparison to the Remez exchange algorithm, which is commonly used to calculate the coefficients for FIR filters (RABINER and GOLD, 1975). Thus, if indeed high-speed requirements are to be met, then the most simple, and therefore fastest, FIR implementation is a filter with unity coefficients.

The combination of reasonable filter characteristics and computational efficiency makes the repeatedly applied unweighed moving average filter interesting for high-speed and on-line applications of (adaptive) digital filtering.

References

KOELEMAN, A. S. M., ROMPELMAN, O., VAN DEN AKKER, T. J. (1984): 'Estimation accuracy of P wave and QRS complex occurrence times in the ECG: the accuracy for simplified theoretical and computer simulated waveforms', *Sig. Proc.*, **7**, (4), pp. 389–405

LATSON, T. W. (1994): 'Principles and applications of heart rate variability analysis' *in* LYNCH, C. III, (Ed.): 'Clinical cardiac electrophysiology: Perioperative considerations' (Lippincott Company, Philadelphia) pp. 307-349

LYNN, P. A. (1977): 'Online digital filters for biological signals: some fast designs for a small computer', *Med. Biol. Eng. Comput.*, **15**, pp. 534–540

OPPENHEIM, A. V., and SCHAFER, R. W. (1989): 'Digital signal processing' (Prentice-Hall, Englewood Cliffs, New Jersey)

QUANGFU ZHAO, and TAKADORO, Y. (1988): 'A simple design of FIR filters with powers-of-two coefficients', *IEEE Trans.*, **CS-35**, pp. 566–570

RABINER, L. R., and GOLD, B. (1975): 'Theory and applications of digital signal processing' (Prentice Hall, London)

ROMPELMAN, O. (1987): 'Accuracy aspects of ECG-preprocessing for the study of heart-rate variability' *in* KITNEY, R. I., and ROMPELMAN, O. (Eds.): 'The beat-by-beat investigation of cardiovascular function' (Clarendon Press, Oxford) pp. 103–125

THAKOR, N. V., and MOREAU, D. (1987): 'Design and analysis of quantised coefficient digital filters: application to biomedical signal processing with microprocessors', *Med. Biol. Eng. Comput.*, **25**, pp. 18–25

WILCOCK, A. H., and KIRSNER, R. L. G. (1969): 'A digital filter for biological data,' *Med. Biol. Eng.*, **7**, pp. 653–660

YONG CHING LIM and PARKER, S. R. (1983): 'FIR filter design over a discrete powers-of-two coefficient space,' *IEEE Trans.*, **ASSP-31**, pp. 583–590