

Anti-alias and anti-image filtering: The benefits of 96kHz sampling rate formats for those who cannot hear above 20kHz.

Julian Dunn

Nanophon Limited

Cambridge, UK

Email: Jdunn@iee.org

Web: <http://www.nanophon.com>

ABSTRACT

Reports that 96kHz sampled digital audio systems have greater transparency than those sampling at 44.1kHz apparently conflict with knowledge of the capability of human hearing. The band-limiting filters required are examined for a role in producing these differences. Possible mechanisms are presented for these filters to produce audible artefacts and filter designs avoiding these artefacts are illustrated.

1.0 INTRODUCTION

Anti-alias filters are used to limit the bandwidth of a signal being input to a sampled system, such as a digital audio system, so that it is less than one half the system sample rate. This is required to avoid ambiguity resulting from the sampling process - which represents all signals as being within a frequency bandwidth defined by a 'folding' frequency of one half of the sample rate. Signals sampled at a frequency outside this range are 'aliased' so that they are indistinguishable from frequencies within it.

For digital audio analogue to digital converter (ADC) sub-systems with an output sample rate of 44.1kHz this filter would, ideally, be transparent to signals in the region below the folding frequency of 22.05kHz and totally reject signals above that frequency. In practice the passband response will not be transparent and the stopband may have inadequate rejection. This may be a particular problem just above 22.05kHz due to the finite attenuation gradient over the transition (band) between the passband and rejection band.

In the reverse process, anti-imaging (or reconstruction) filters are used at the output of a sampled system to eliminate spectral images of the desired signal. These are generated at integer multiples of the sample rate. The ideal filter specification is the same as that for the anti-aliasing filter. Any passband errors will have the same effect for both types of filter. The consequences for the rejection band deviating from the ideal are not the same, however, and depend on the response of the following equipment to the images that may be generated.

Both types of filter can be implemented in the analogue or digital domains. Digital sample rate converters also have bandlimiting filters to perform the functions of anti-alias and anti-image filtering. These are implemented only in the digital domain. The issues and compromises made are similar whether the processing is digital or analogue.

In most current audio ADC and DAC sub-systems this filtering is performed by a gentle, non-critical, analogue low-pass filter of low order in conjunction with an oversampled converter and a high order digital 'brickwall' filter. This allows the sharp cut-off filtering to be performed in the digital domain using a finite impulse response (FIR) filter of one or more stages. In a multi-stage filter only one of the stages performs the 'brickwall' function and relates to the issues raised in this paper.

This paper is concerned with the compromises often made in the specification of this filter. (Many of the issues raised also apply to analogue and digital infinite impulse response (IIR) filters.) In particular, it shall examine the effects of passband frequency response error, or ripple, and of having the rejection band start at a frequency above half the sample rate.

Could these effects - producing very small errors in the sub-20kHz audio band - explain any of the reported differences between 44.1kHz and 96kHz sampled systems? [1, 2, 3, 4]

If these aspects are relevant then how should the design of anti-alias and anti-image filters for 96kHz sampled systems differ from those for 44.1kHz sample rate systems? Could some of the benefits reported for 96kHz systems be achieved for lower sample rate systems with different filters?

2.0 PASSBAND DEVIATIONS

2.1 Ripple

The passband response should ideally be flat (or an exact complement of other deviations so that the total system response is flat).

The digital FIR filters used in most ADC and DAC sub-systems approximate to this ideal in a way that minimises the maximum distortion of the real filter response from this ideal. They are 'equiripple' filters, meaning that their responses are characterised by the presence of ripples which are all the same size.

The REMEZ [5] filter design program works in this way. Figures 1 to 4 illustrate the response of some example 2x oversampling filters designed using REMEZ. Notice from graph B of each of these that the deviation of the passband (like the stopband) has a constant ripple. This passband response can be considered as the desired - flat - response with an additional error response. This error can be approximated by a cosine wave. This is shown in graph D. For figure 1 this cosine has cycles that are 3.3kHz apart, and an amplitude of 0.0006 (relative to the average gain of 1). This is not to be confused with a cosine wave in the time domain; it is a transfer function that has a cosinusoidal shape in the frequency domain. What does it represent in the time domain?

Given a frequency response, the time response can be found by Fourier transformation. If we approximate the error (of the figure 1 filter) to a sine wave then the transform is simply an impulse pair of total amplitude -67dB (-70dB each) displaced by $\pm 0.3\text{ms}$ (14.5 samples) relative to the main signal. (The signals are delayed by at least this amount in the filter so causality is maintained.) This illustrates that the periodic passband ripple indicates pre and post-echoes in the time domain.

This result - that for equiripple filters the passband time dispersion results in discrete time displacements - is contrary to many expectations. The spread of the impulse response is normally quoted as being indicative of the amount of 'smearing'. A comparison of graphs C and E for each figure show that the pre and post echoes occur at times corresponding to the beginning and end of the impulse response - where the amplitude is smallest.

The filter of figure 1 filter is not unique. Table 1 is calculated from the characteristics of various integrated ADCs, DACs, and digital filters. The amplitude and the pre and post echo time displacement are calculated from the passband ripple amplitude and periodicity estimated from data sheets and published graphs. (The filters of figures 1 and 2 are designed to have similar responses to devices 1 and 2 in this list)

	Type	Ripple deviation	Ripple Amplitude	Echo displacement
1	ADC	± 0.005 dB	-68 dB	± 0.3 ms
2	DAC filter	± 0.000015 dB	-118 dB	± 0.8 ms
3	DAC	± 0.035 dB	-51 dB	± 0.3 ms
4	ADC	± 0.001 dB	-82 dB	no data
5	ADC	± 0.01 dB		
	(last stage)	± 0.003 dB	-72 dB	± 0.7 ms
6	DAC	± 0.0005 dB	-88 dB	no data

Table 1

An inspection of the columns for echo amplitude and ripple rate shows a wide variation with amplitudes of -51dB to -118dB and timing variations of between 0.3ms and 0.8ms. (Calculated for a sample rate of 48kHz)

The production of pre-echoes from filter ripple was reported by Lagadec and Stockham [6]. They found the pre-echo due to a filter ripple of ± 0.2 dB with a span of 23Hz corresponded with echoes of -32dB at ± 40 ms - which was found to be quite perceptible even with untrained listeners. A question is raised of how much echo, and in particular how much of the un-masked pre-echo, can be permitted without producing a degradation in the highest quality reproduction system? The psychoacoustics results, including those reported in [7] for example, do cast some light on this. However, none of the results seem to be directly applicable. It would seem that some kind of threshold test on the audibility of pre and post-echo pairs is required.

Some of the work on the evaluation of 96kHz sampling systems reports an improved resolution of the timing of impulsive test signals [3], and of an improved perception of musical transient attacks and fast staccato passages [2].

Timing is also related to localisation cues but research into the improvement in phantom source perception differences between 48kHz and 96kHz [4] has not shown an improvement in localisation accuracy. The experiments may be worth repeating with less ideal filters.

2.2 Filters with ripple-free passbands

There are methods of FIR filter design that produce filters without passband ripple, and therefore without the discrete echoes, such as the high-fidelity filter [8] or the constraint based FIR filter design program, METEOR [9]. These approaches are not normally used for digital audio applications - presumably because of the requirement for greater computation. As the cost of processing falls this situation may change.

2.3 Clipping

It may seem a trivial matter to eliminate the possibility of a signal at the ripple peak causing a clip by reducing the passband gain by the amount of the ripple. This is adequate for static sine-wave signals but does not allow for the worst case of input.

For an FIR filter the very maximum output value, or overshoot, will occur with an input signal that causes the magnitude of all the coefficients to be added. This input signal would consist of positive and negative full-scale values that all have the same sign (or all have the opposite sign) as the coefficient with which they are aligned. This pattern is very unlikely to exist in a real signal but represents an upper bound to the amount of overshoot that could occur. This value is shown for each of the filters in figure 1 to 4 and can be seen to vary between 2.9 and 5.7dB.

Clipping of DAC interpolation filters can often be observed when presenting them with a square wave peaking at close to full scale. (This may look clean but it can be shown to be clipping by reducing the level slightly and a clipped ringing overshoot will be observed on the filter output.) This clipping represents a sharp, or high order, non-linearity that should be avoided. The step response shown in graph F of each figure illustrates the extent of step overshoot for these example designs. It can be seen that there is not much variation between them so this may not be a useful indicator.

One approach to addressing the overshoot problem is to reduce all the coefficient values by the same proportion. This results in a gain reduction and will normally reduce the dynamic range of by the same amount.

Another method of eliminating the overshoot is to design the filter so that there is less ringing. For example, design it with a wider transition region between passband and rejection band (this is obviously easier with a higher sample rate). Compare, for example the maximum overshoot values for figures 3 and 4.

3.0 INADEQUATE REJECTION

The purpose of the anti alias and anti image filters is to reject signals at frequencies above the folding frequency ($0.5f_s$). It has also become accepted wisdom that for high quality applications these filters should have a flat response to 20kHz and that immediately above that frequency the response should plummet into the rejection band. For system operating at 44.1kHz this requires a transition region from 20kHz to 22.05kHz ($0.454f_s$ to $0.5f_s$) in which the filter should develop the full rejection of the filter if it is to avoid the generation of aliases or images. In practice - as an examination of table 2 reveals - many of the (otherwise) highest performance integrated circuits do not achieve this.

	Type	Stopband attenuation	Stopband edge	Attenuation at 0.5fs
1	ADC	100dB	0.604fs	3dB
2	DAC filter	110dB	0.5465fs	6dB
3	DAC	60dB	0.55fs	10dB
4	ADC	110dB	0.5465fs	no data
5	ADC	117dB	0.4979fs	117dB
6	DAC	90dB	0.55fs	6dB

Table 2

3.1 Filter design compromises

The minimum length of an FIR low pass filter is related to three parameters. It increases with reduced transition region width, reduced maximum pass-band error (or ripple), and with increased minimum stop-band rejection. (Empirical formulae relating these parameters to filter length are given in [10])

A decimating FIR filter with N non-zero coefficients requires N multiplications to be performed per output sample. For an interpolating stage many of the input samples will be zero filled: In this case the filter multiplication rate is N per input sample. (There are also methods of saving on this computation but these do not change the importance of the three parameters of the previous paragraph.)

The length of the filter also infers a storage requirement with its associated costs.

The processing and storage requirements normally limit the size of the filter, but in some cases the filter group delay is important. In applications such as artist return feeds, the delay is more important than other aspects of the filter response. In that case the filter size may be restricted (as in the low group delay mode of the CS5397 [11]), or an asymmetrical FIR (with a non-flat group delay) may be used.

3.2 Half-band filters

One of the methods of saving complexity in the low pass filter is to use half-band filters. [12] These filters are a special case of a low pass FIR filter with a frequency response that has the following symmetric property (where ffs is the filter sampling frequency, equivalent to 2fs for a 2x oversampling filter):

$$H(f) = 1 - H(0.5ffs - f)$$

The difference in the frequency response from 1 below 0.25ffs and from 0 above 0.25ffs is symmetric about that frequency.

Given this property in a filter of odd length then every even coefficient - apart from the central coefficient - is exactly 0. Graph C of figure 2 illustrates this. For a given length of FIR this saves almost 50% in the number of multiply-accumulates that are required.

This type of filter is used for interpolation or decimation by factors of two (then the symmetry is around one half the lower sample rate).

There is, however, an important disadvantage to using this type of filter to form the final decimation stage or first interpolation stage in a multi-stage oversampling filter. The symmetry property requires that at a frequency of 0.25ffs (0.5fs) the attenuation

is 0.5. This means that it cannot provide adequate rejection to avoid images or aliases close to that frequency.

It is also a restriction that the passband ripple and minimum rejection cannot be specified separately. This may not be seen as much of a problem given the ability to have a filter of twice the length for the same processing power.

4.0 IMAGING

For the listener who cannot hear signals above 20kHz it may be argued that anti-image filters for DACs operating with an input sample rate of 44.1kHz are not required. All images will be above 22.05kHz and inaudible. (This case may be argued even more strongly for 96kHz systems, when images are more than an octave higher.)

Such arguments ignore the potential for non-linear behaviour in the electronic and electromechanical stages following the DAC in the signal path. This non-linearity will cause high frequency images above the audio band to intermodulate with signals within the audio band. This would produce audio band intermodulation distortion artefacts that could fall in the band.

For example, consider the half-band interpolation filter shown in figure 2.

Take, for an extreme example, a full scale input signal of 1kHz below the half sample frequency. The image of this tone (to be suppressed by this filter) will be at 1kHz above the half sample frequency, and at full scale. The filter attenuation at that frequency is 25dB.

The output of the DAC following this filter will therefore have two tones 2kHz apart. A second-order non-linearity in an amplifier or loudspeaker would produce intermodulation products at the frequency sum and frequency difference of these tones. The amplitude of the products would be equivalent to the product of the two amplitudes multiplied by the coefficient of non-linearity.

If the second order non-linearity in the following stages (amplifiers and loudspeakers) is worse than 1% at that frequency and level then the resulting distortion product at 2kHz will be approximately -70dB below the signal. Do not forget that - for our listener at least - the signal itself is inaudible and will not have any masking effect.

Few people would choose to listen to this signal. It merely illustrates how the image rejection performance of an interpolation filter could have a significant impact on the audio band below 20kHz.

5.0 ALIASING

This is caused by sampling and is the reflection of the spectrum of a signal about the $0.5f_s$ folding frequency. The effect of inadequate aliasing in a digital audio ADC sub-system is to produce frequency shifted signal in the audio band.

Most digital audio ADC sub-systems have poor alias rejection at close to the folding frequency. Figure 1, graph A, is an illustration of the effect. The dotted line shows the folded stopband rejection. This corresponds to the attenuation applied to aliases that would be folded to the respective frequency. In this example the direct effect of

the poor alias rejection in the transition region would be inaudible for those who cannot hear above 18kHz.

Signal just above the folding frequency will be hardly attenuated at all. The intermodulation mechanisms mentioned in the previous section could also apply to signals in this region. The frequency shifting in that region would change the frequency of any lower frequency difference-frequency distortion. Such a change may be perceived as unnatural and hence more noticeable. As before, this effect depends on the linearity of the electronics (and mechanics) of the following equipment.

6.0 FILTERS FOR 96KHZ SAMPLED SYSTEMS

The filters of figures 1 and 2 are modelled on products that can operate at 44.1kHz with an equiripple passband to 20kHz. At that sample rate the equiripple band upper edge is at 90% of the half sample rate.

The filters of figures 3 and 4 are specified for the equiripple region of the passband to extend to 32% and 52% of the folding frequency respectively. These designs illustrate the use of the extra bandwidth (provided by doubling the sample rate) for providing a wider, or more relaxed, transition region. (Please note the distinction between the passband and the equiripple region of the passband. For a low pass filter with a very low ripple and a gradual transition region the traditional definition for filter bandwidth (the -3dB frequency) will be significantly above the top of the equiripple region, as illustrated in graph B of figures 3 and 4.)

6.1 31 tap decimation/interpolation filter

The 192kHz filter of figure 3 requires approximately the same multiply-accumulate rate as the 63 tap filter of figure 1 operating at 96kHz (for a 48kHz output fs).

Graph A shows that this filter reaches the full attenuation of the stopband at half the sample rate. This compares with figure 1 graph B where the attenuation of aliases at the equivalent point is only 3dB.

The upper edge of the equiripple region is at 0.08 of the 192kHz sample rate, 15.4kHz. Above that frequency the roll-off is very gentle so that the attenuation at 20kHz is only about 0.3dB and the -3dB frequency is at 26.5kHz. This should be compared with the filter of figure 1 which has 20kHz within the passband equiripple region and the -3 dB frequency is at 25kHz.

The periodicity of the ripple, illustrated in graph D, at 10kHz per cycle transforms to a timing of ± 0.1 ms and the ripple amplitude of ± 0.003 dB corresponds with a level of -73dB. This compares with the ± 0.3 ms at -67dB for figure 1.

The headroom that could be required for the worst possible input stimulus - the maximum overshoot - is just under 3dB. This compares with 5.5dB and 5.7dB for the first two filters.

6.2 63 tap decimation/interpolation filter

Figure 4 shows a filter with the same number of taps as that of figure 1. As the sample rate is doubled this means that the computational rate is increased by the

same amount and that the filter length is halved in time. Like the filter of figure 3 the stopband lower edge is aligned with half the sample frequency.

The passband upper edge of this filter is at 0.13 of the 192kHz sample rate, or 25kHz, and the graph B shows the transition region rolls off to the -3dB frequency at 34.7kHz.

This filter is long enough to reduce the passband ripple to $\pm 0.000\ 006$ dB which transforms to a pre-echo amplitude of -126dB. The periodicity of this ripple transforms to echo timing of ± 0.18 ms.

The maximum overshoot is worse than that of the previous filter, at 4.8dB, and so there is minimal dynamic range advantage for this over the filters of figures 1 and 2.

7.0 CONCLUSION

Linear and non-linear distortion mechanisms within digital audio FIR low pass filters used for decimation and interpolation have been described. Four filters have been designed using the REMEZ algorithm. Two of these are based on commercially available devices in common use and these are used to illustrate the issues.

The typical cosinusoidal passband ripple characteristic has been analysed to estimate the time-dispersion characteristics of the filter to signals within the audio band. This analysis is used to show significant differences in the pre and post-echo performance between the example filters.

Susceptibility of the filters to overshoot has been illustrated and the effect of compromises in stopband performance close to the folding frequency are discussed.

Filters designed for the higher sampling frequency of 96kHz are used to show how all these distortion mechanisms can be reduced by filters of similar computational requirements but with a more relaxed transition region. This results in either reduced ripple (and hence pre-echo) amplitude or in a lower time displacement for the echo. In both cases the effect is likely to be a reduction in the audibility of the echo. A direct effect of the higher sampling rate is that for an identical filter design the time displacements will scale inversely with sample rate. Hence an improvement can be made just from raising the sample rate - even for those who cannot hear above 20kHz.

More work is required to evaluate the limits on the perception of the echo effects described here. This should cover both the audibility of echoes and their effect on the localisation of sound sources. One direct benefit of the increased use of 96kHz for recording is that there will be an increasing amount of source material suitable for this work.

The effects described here indicate that it may be difficult to distinguish any beneficial effects of an increase in sampling frequency from the different filter behaviour. This should be considered when making comparisons between different rates.

8.0 REFERENCES

- [1] S. Yoshikawa, S. Noge, M. Ohsu, S. Toyama, H. Yanagawa and T. Yamamoto 'Sound Quality Evaluation of 96 kHz Sampling Digital Audio' Preprint 4112 of the 99th AES Convention, New York, October 1995.
- [2] Takeo Yamamoto 'Sound Quality of 96 kHz Sampling Digital Audio' Presentation to workshop 'High Bandwidth High Quality Digital Audio' at the 101st AES Convention, Los Angeles, November 1996
- [3] S. Yoshikawa, S.Noge, T. Yamamoto and K. Saito 'Does High Sampling Frequency Improve Perceptual Time-Axis Resolution of Digital Audio Signal?' Preprint 4562 of the 103rd AES Convention, New York, September 1997.
- [4] Bernd Theiß and Malcolm Hawksford 'Phantom Source Perception in 24 Bit @ 96 kHz Digital Audio' Preprint 4561 of the 103rd AES Convention, New York, September 1997.
- [5] J. H. McClellan, T. W. Parks and L. Rabiner 'A Computer Program for Designing Optimum FIR Linear Phase Digital Filters' IEEE Transactions on Audio and Electroacoustics, Vol. AU-21, No. 6, pp. 506-526, December 1973.
- [6] R. Lagadec and T. G. Stockham, 'Dispersive Models for A-to-D and D-to-A Conversion Systems' Preprint 2097 of the 75th AES Convention, Paris, March 1984.
- [7] Thomas Sporer and Holger Schröder, 'Measuring Tone Masking Noise', Preprint 3349 of the 93rd AES Convention, San Francisco, October 1992
- [8] R. H. Wilkinson, 'High-fidelity finite-impulse-response filters with optimal stopbands' IEE Proceedings-G, Vol. 138, No. 2, pp. 264-272, April 1991.
- [9] K. Steiglitz, T.W. Parks and J.F. Kaiser, 'METEOR: A Constraint-Based FIR filter design program' IEEE Trans. Signal Processing, Vol. 40, No. 8, pp.1901-1909, August 1992.
- [10] R.E. Crochiere and L.R. Rabiner 'Interpolation and Decimation of Digital Signals - A Tutorial Review' Proceedings of IEEE, vol. 69, no. 3, pp.417-448, March 1981.
- [11] Crystal Semiconductor data sheet 'CS5396 CS5397 120dB, 96kHz Audio A/D Converter' September 1997.
- [12] Fred Mintzer, 'On Half-Band, Third-Band, and Nth-Band FIR Filters and Their Design' IEEE Transactions on Acoustics, Speech and Signal Processing, Vol. ASSP-30, No. 5, pp. 734-738, October 1982.

9.0 LIST OF FIGURES

Figure 1	(2 pages)	ADC decimation filter for 48kHz output sampling rate
Figure 2	(2 pages)	DAC interpolation filter for 48kHz input sampling rate
Figure 3	(2 pages)	A 31 tap filter for 96kHz sampling rate
Figure 4	(2 pages)	A 63 tap filter for 96kHz sampling rate

```

63 tap decimation filter
Input fs = 96kHz, Output fs = 48kHz
REMEZ design parameters:

```

	BAND 1	BAND 2
LOWER BAND EDGE	0.00000000	0.29200000
UPPER BAND EDGE	0.22900000	0.50000000
DESIRED VALUE	1.00000000	0.00000000
WEIGHTING	1.00000000	10.00000000

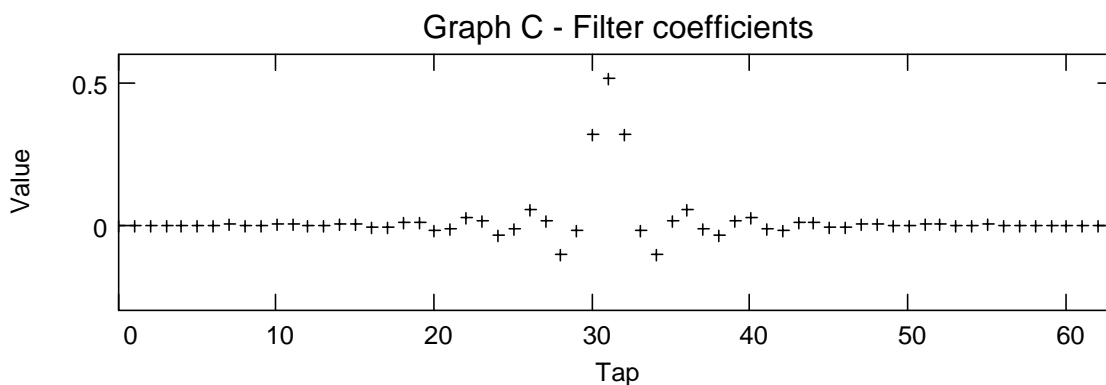
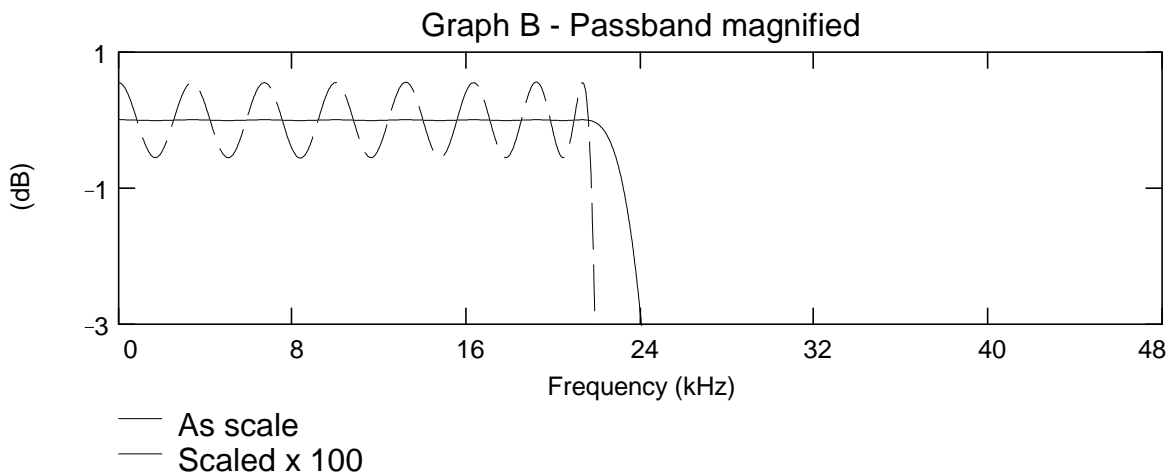
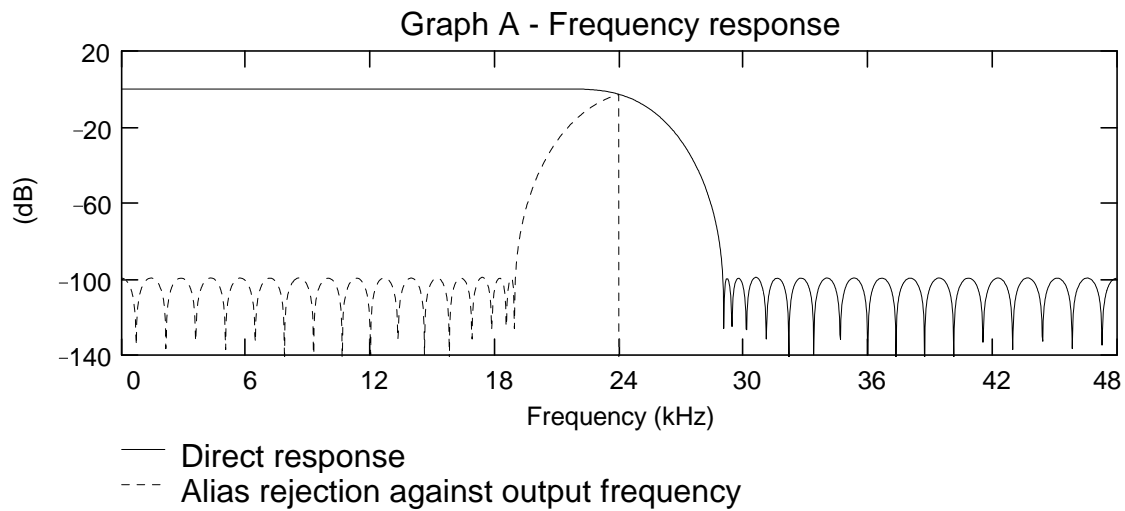
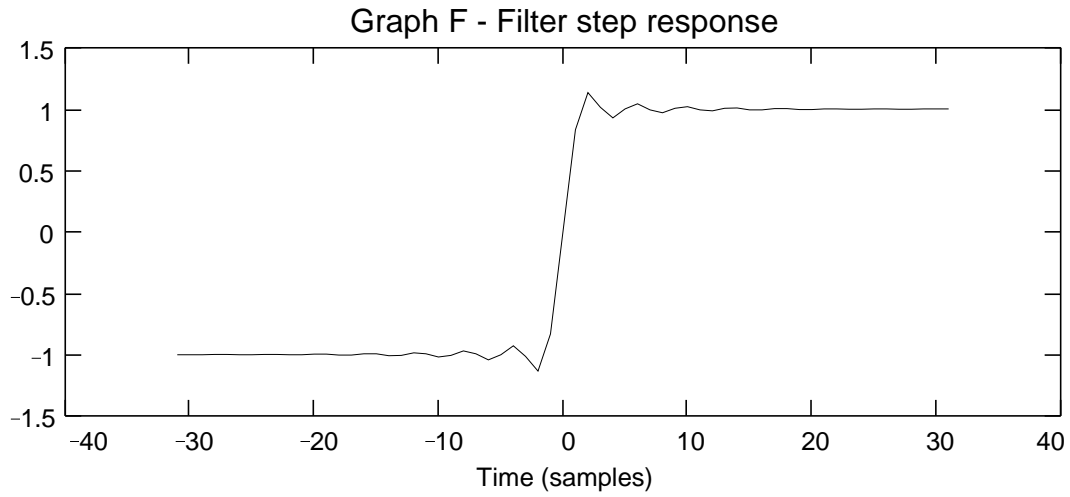
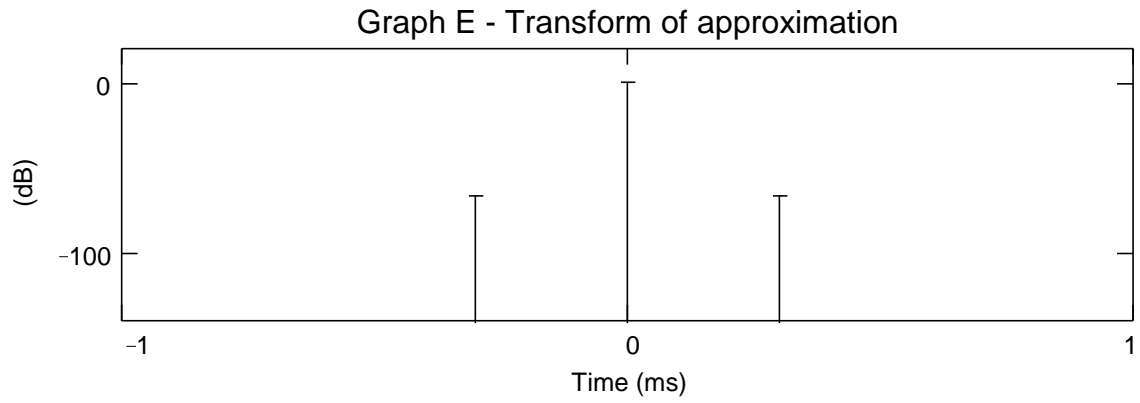
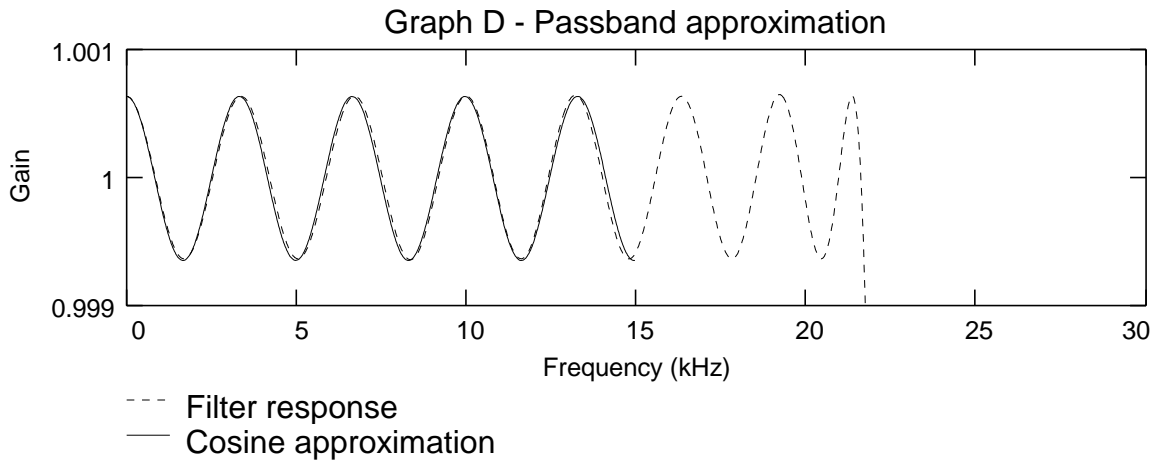


Figure 1
© Julian Dunn 1998



Step_overshoot = 1.092 · dB

Fsample = 96 · kHz

Maximum_overshoot = 5.497 · dB

153 tap half-band interpolation filter
 Input fs = 48kHz, Output fs = 96kHz
 REMEZ design parameters:

	BAND 1	BAND 2
LOWER BAND EDGE	0.00000000	0.27325000
UPPER BAND EDGE	0.22675000	0.50000000
DESIRED VALUE	1.00000000	0.00000000
WEIGHTING	1.00000000	1.00000000

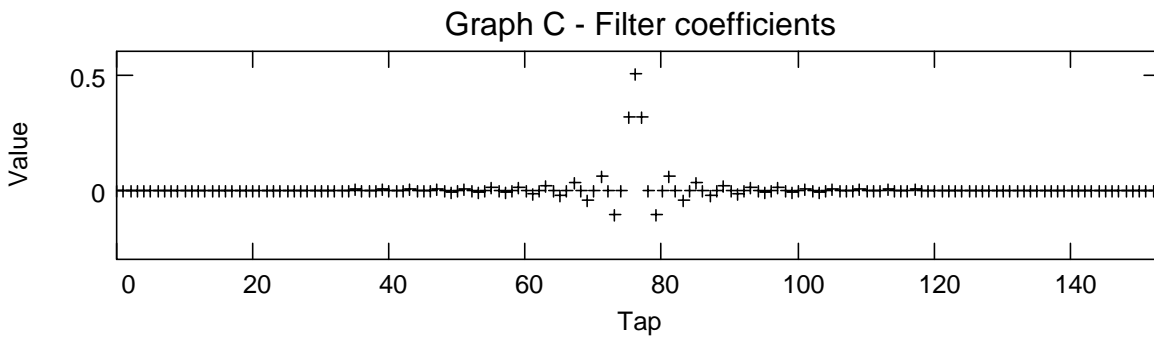
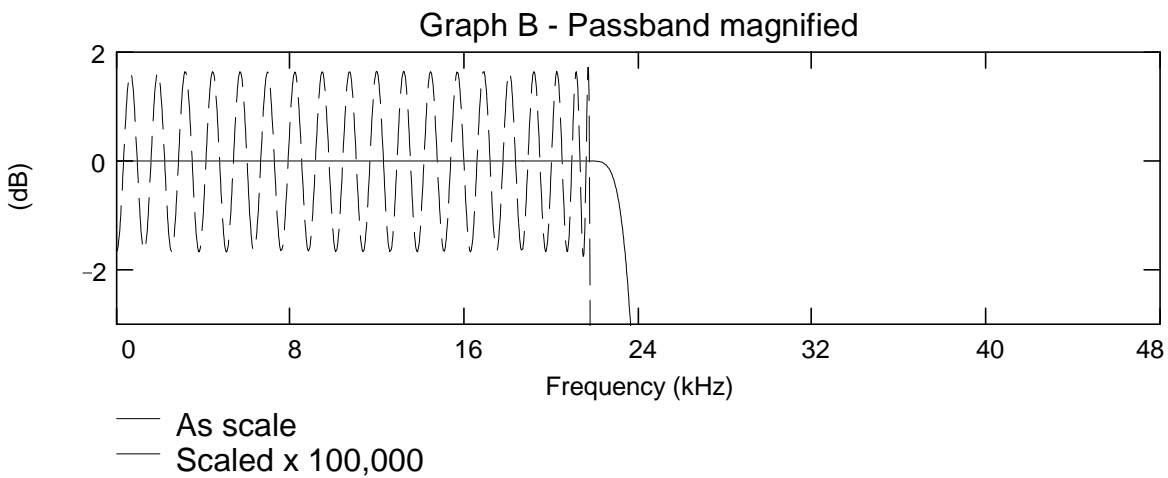
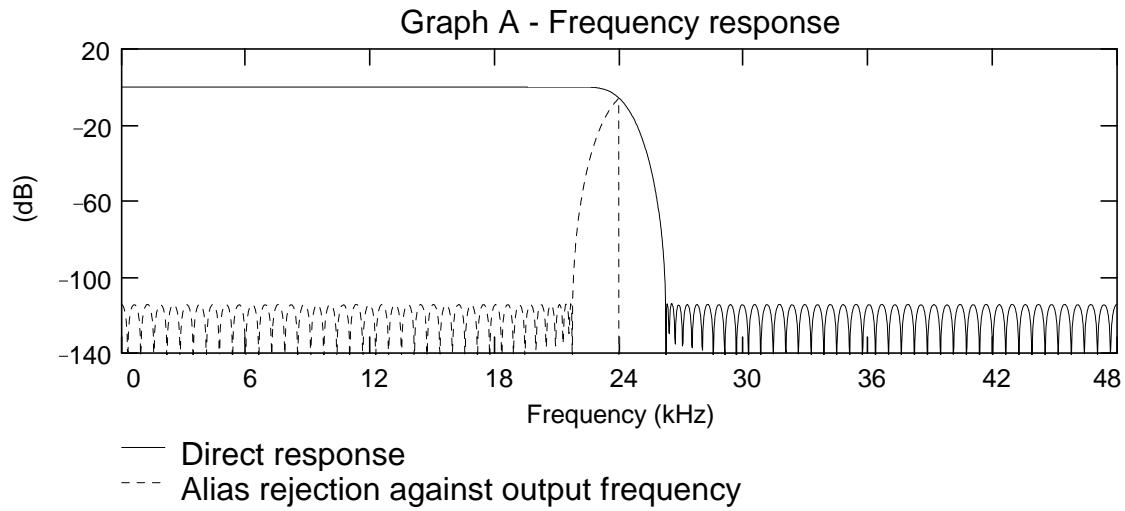
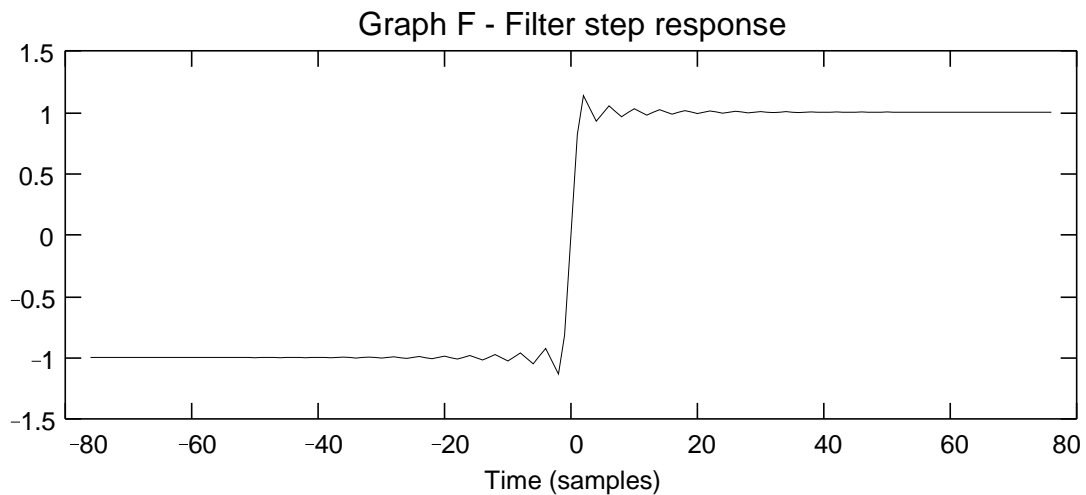
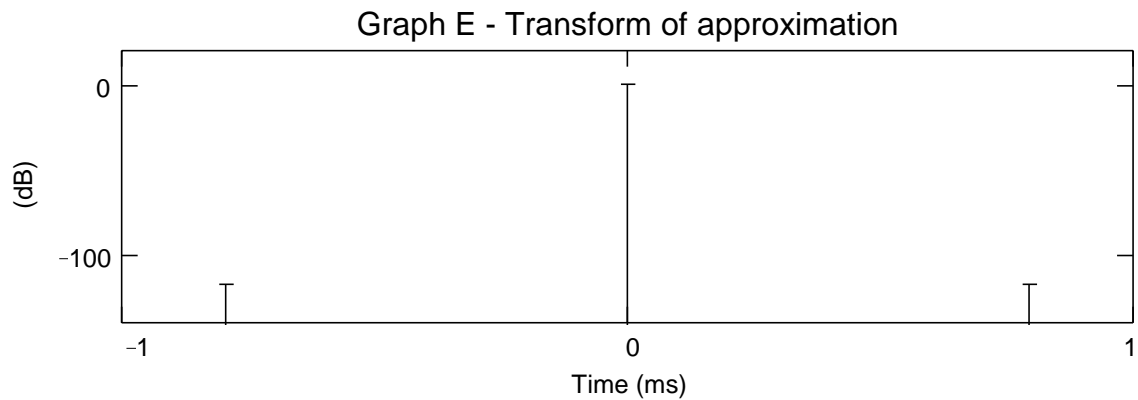
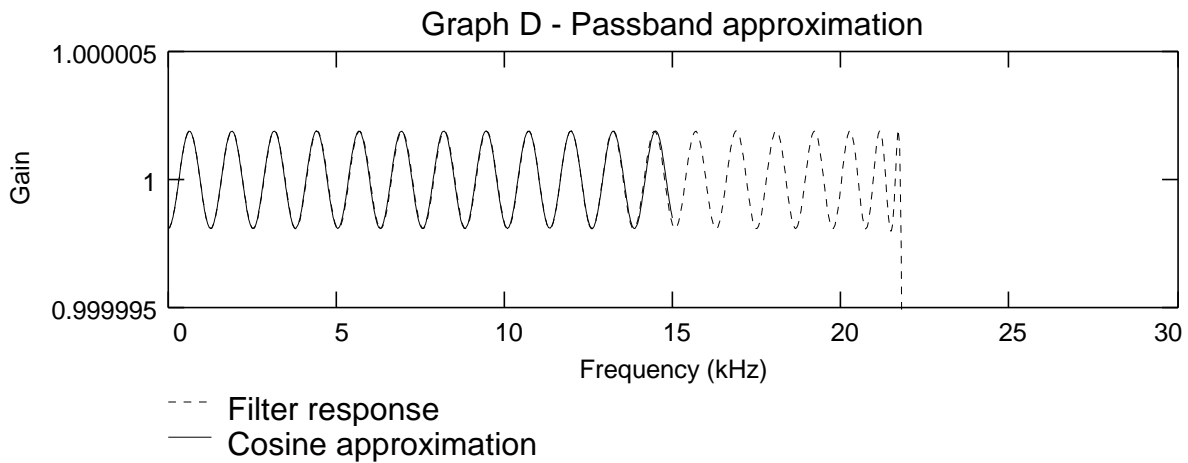


Figure 2
 © Julian Dunn 1998



Step_overshoot = 1.108 · dB

Fsample = 96 · kHz

Maximum_overshoot = 5.683 · dB

31 tap decimation/interpolation filter
 Higher fs = 192kHz, Lower fs = 96kHz
 REMEZ design parameters:

	BAND 1	BAND 2
LOWER BAND EDGE	0.00000000	0.25000000
UPPER BAND EDGE	0.08000000	0.50000000
DESIRED VALUE	1.00000000	0.00000000
WEIGHTING	1.00000000	40.00000000

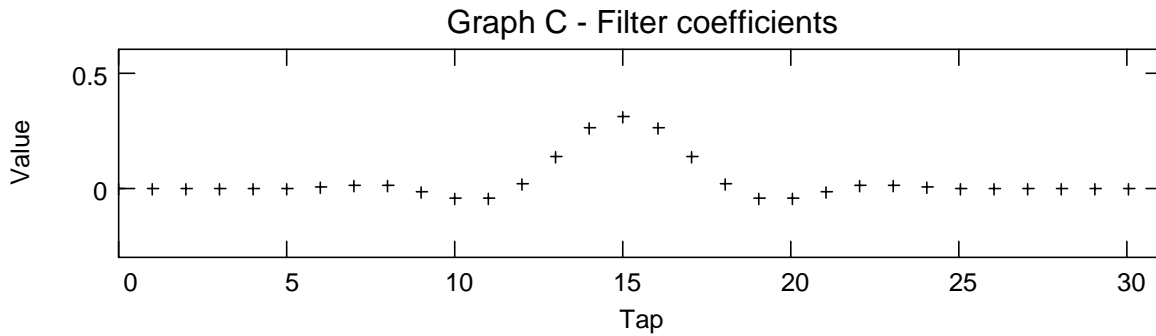
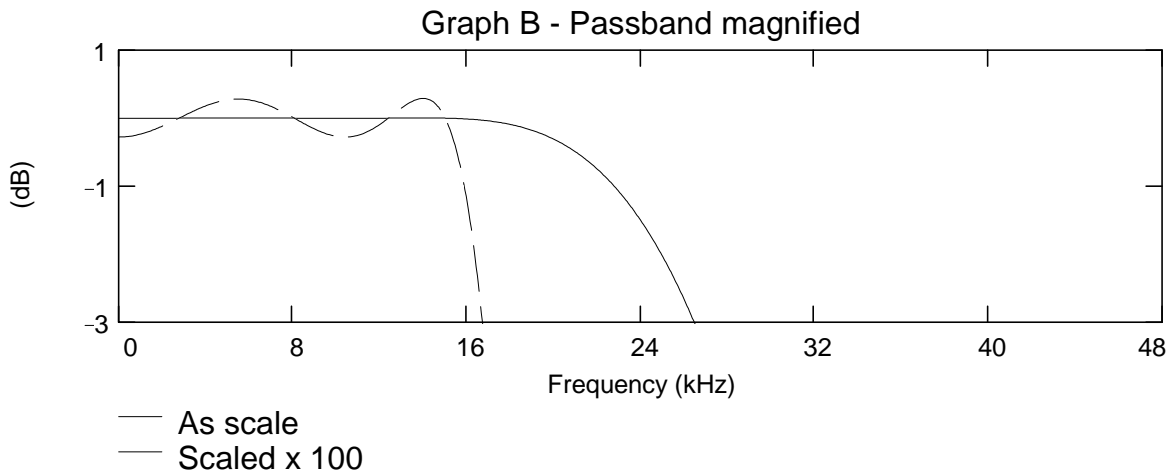
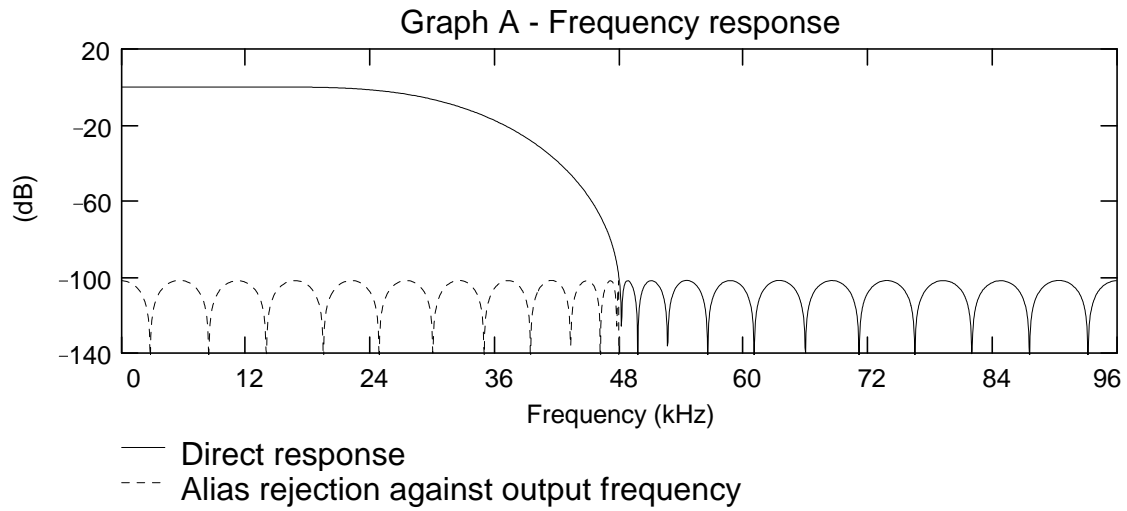
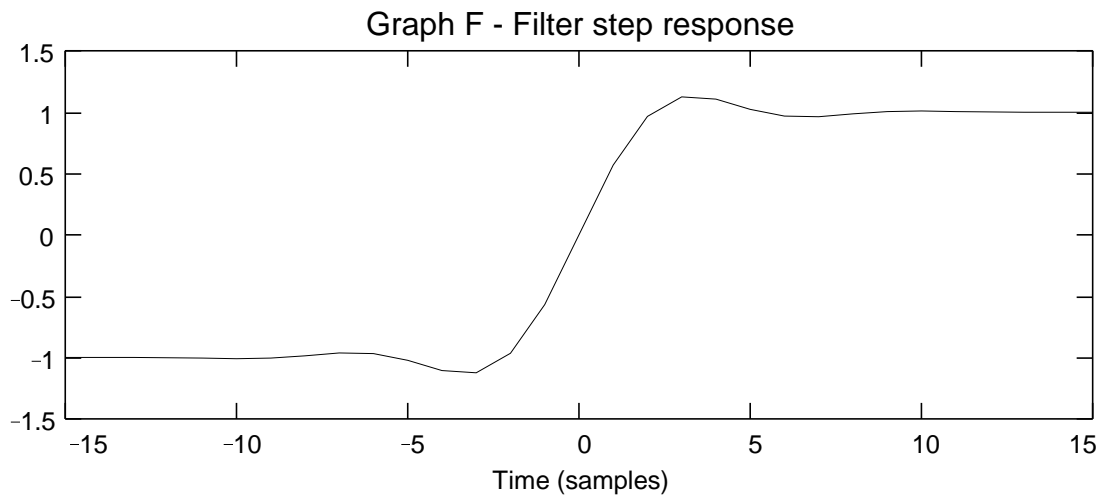
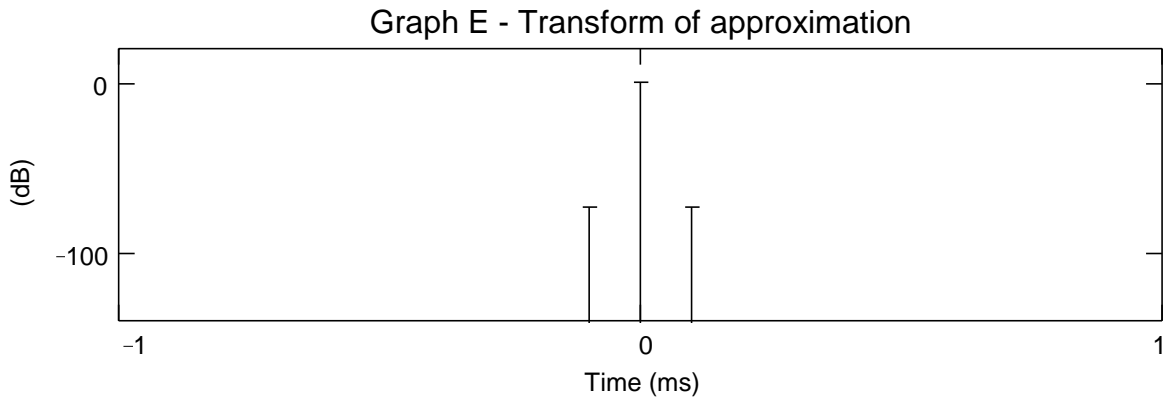
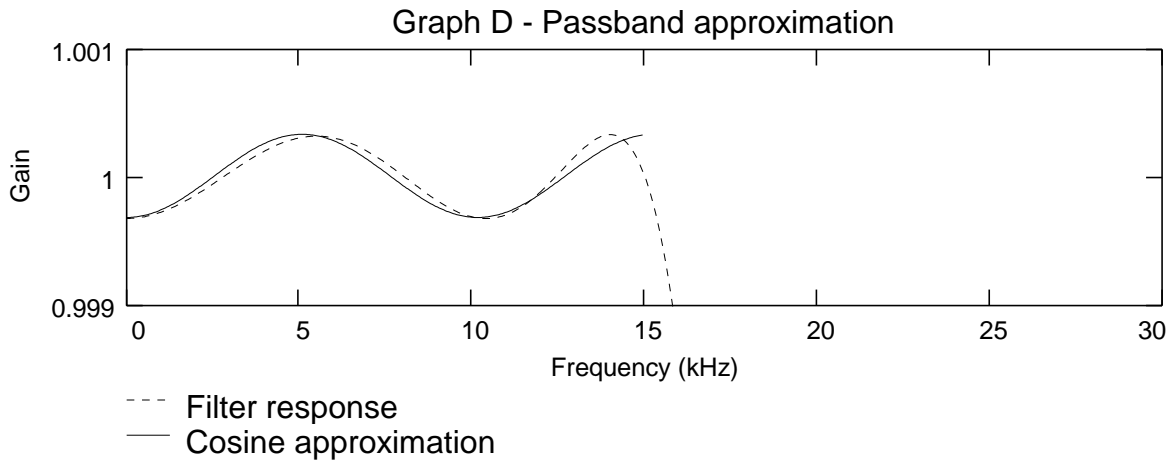


Figure 3
 © Julian Dunn 1998



Step_overshoot = 1.027 · dB

Fsample = 192 · kHz

Maximum_overshoot = 2.974 · dB

63 tap decimation/interpolation filter
 Higher fs = 192kHz, Lower fs = 96kHz
 REMEZ design parameters:

	BAND 1	BAND 2
LOWER BAND EDGE	0.00000000	0.25000000
UPPER BAND EDGE	0.13020000	0.50000000
DESIRED VALUE	1.00000000	0.00000000
WEIGHTING	1.00000000	1.00000000

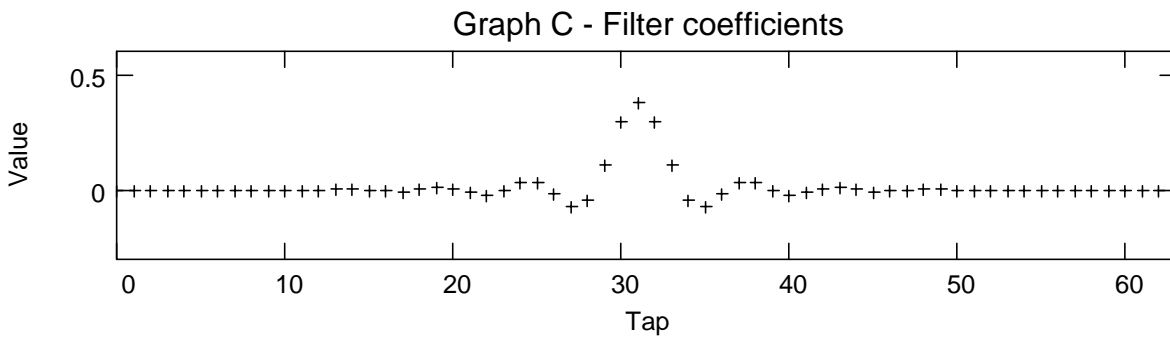
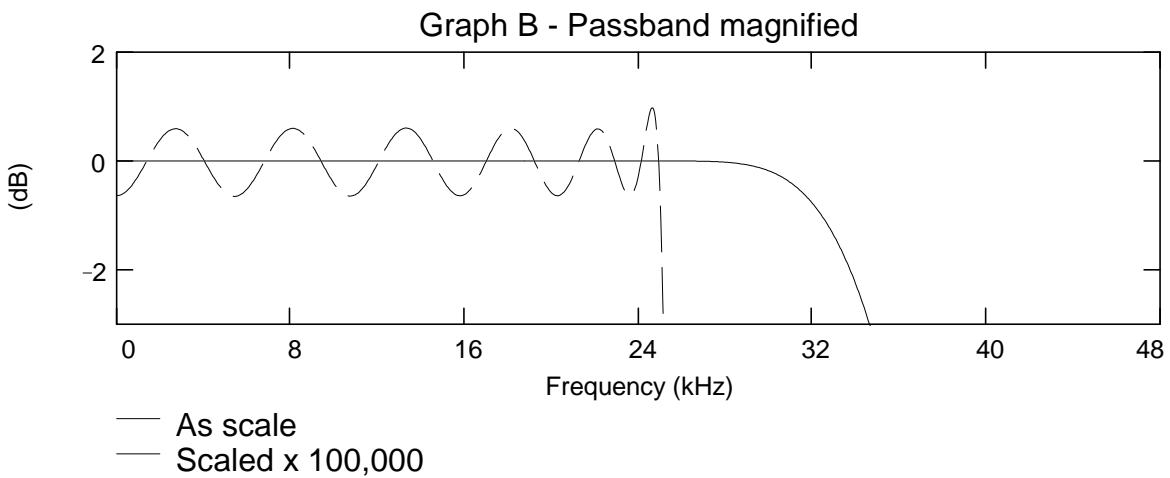
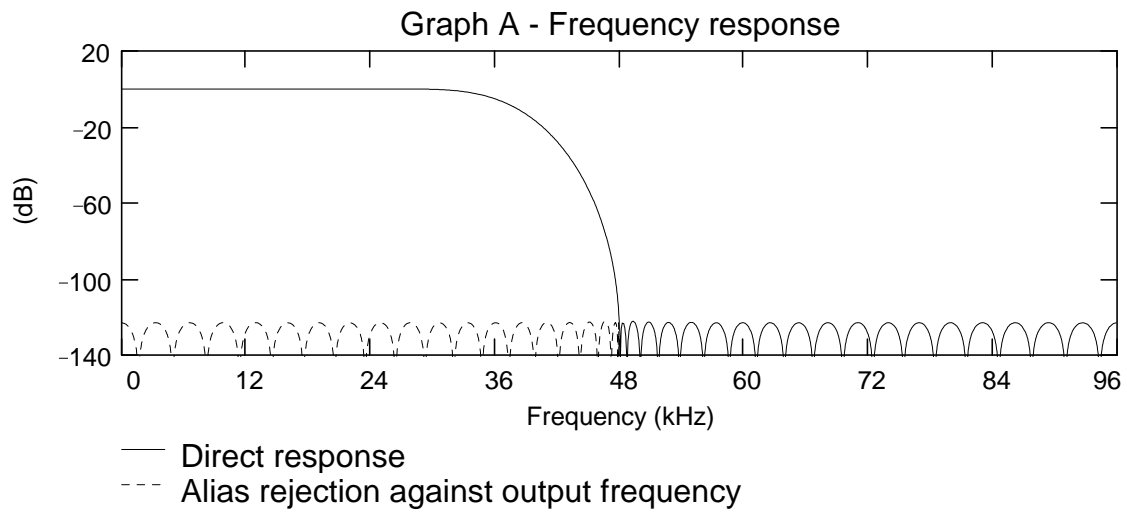
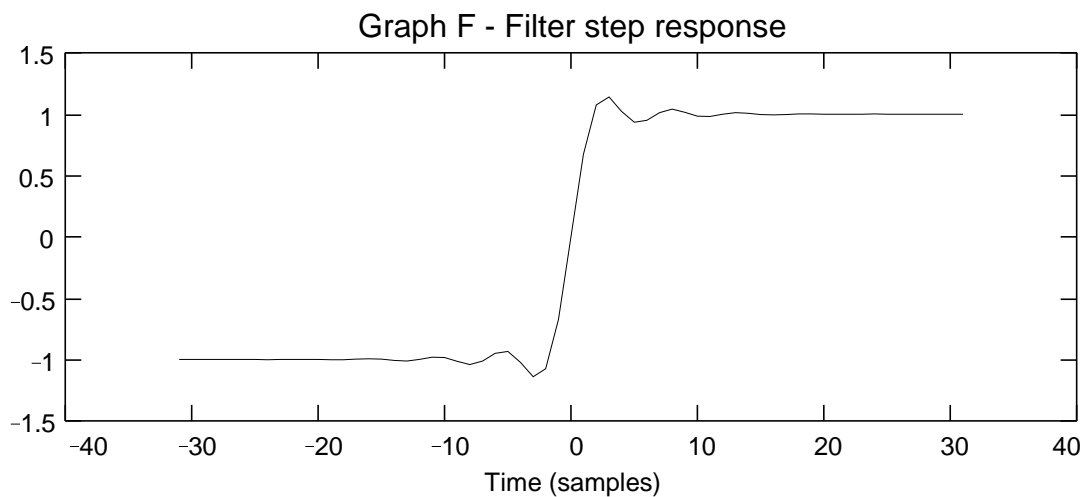
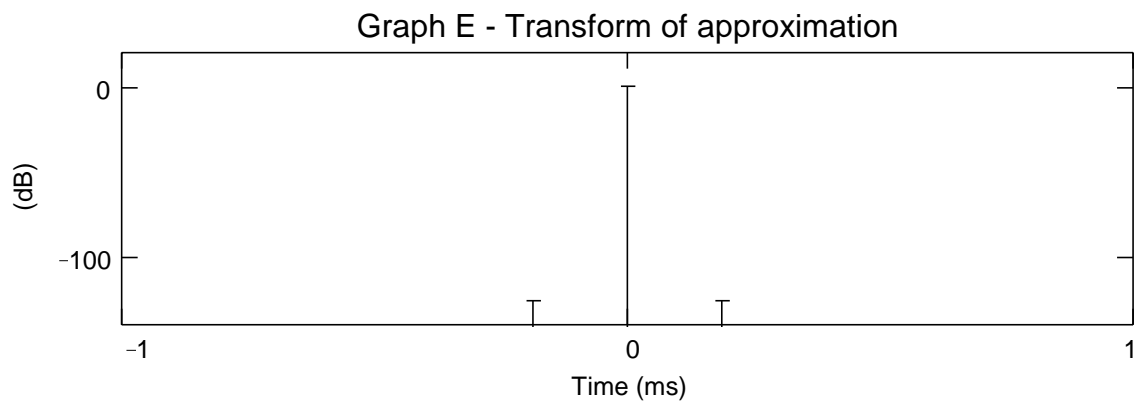
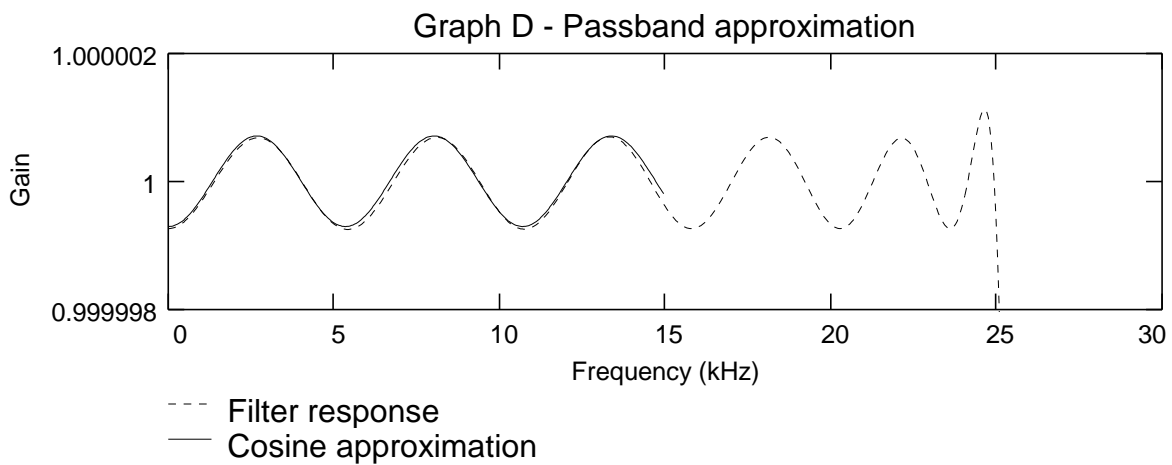


Figure 4
 © Julian Dunn 1998



Step_overshoot = 1.131 · dB

Fsample = 192 · kHz

Maximum_overshoot = 4.761 · dB