LOW COMPUTATIONAL COST METHOD FOR SOLVING ROUND-OFF ERRROR AT LOW FREQUENCY IN DIGITAL AUDIO EQUALIZATION.

PACS 43.60 c

Germán Ramos; José J. López
Departamento de Comunicaciones, Universidad Politécnica de Valencia
Camino de Vera s/n
46022 Valencia
Spain
Tel. +34 96 2849348
Fax +34 96 2849309
Email: gramosp@eln.upv.es, jjlopez@dcom.upv.es

ABSTRACT

The *quantization* and *round-off* errors in IIR second order sections of digital filters are well known and resolved by different ways increasing the computational cost and not taking advantage of the special architecture of modern DSP. These undesired effects increase when the resonance frequency of the filter lowers respect to the Nyquist frequency. In this work, a combination of serial and parallel filter decomposition is developed with application to low frequencies where round-off errors are especially problematic. Noise-Shaping techniques are also used in order to move the round-off noise to high frequencies, obtaining nearly perfect behavior. The proposed structure is easy to implement in a DSP with very low additional computational cost compared to standard IIR Canonical forms, and also much less than other solutions. Finally this algorithm has been implemented and tested on a 32 bit floating-point DSP.

1. INTRODUCTION

The implementation of digital IIR filters with limited word-length processors (fixed or floating-point architectures) causes two different types of quantization errors that in general grow as the resonance frequency of the filter is lower (as usual in digital audio equalization) :

- The *quantization of the filter coefficients* results in linear distortion and generates a deviation from the ideal frequency response due to the discrete pole-zero locations.
- The *quantization of the mathematical operations and stored values at the delay line* generates round-off noise that limits the dynamic range of the filter and defines its noise behavior. Secondary effects may occur like limit cycles [Zöl97] and internal overflow.

To solve these problems there are two possibilities: increase the word-length of the processors or/and use different filter structures (like Gold and Rader, Kingsbury and Zölzer [Zöl97], [Opp89]) that are less sensible to quantization and pole-zero locations than the conventional and *academic direct forms*. All of these structures are more complex than the *direct form*, increasing strongly the computational cost, and also not allowing taking advantage of the special internal architecture of modern DSP for these applications, like circular addressing and parallel operations. With the actual DSP of 24 bit fixed-point and 32 bit floating-point, the coefficient quantization is no so important because in general, there is enough resolution to define the filter's frequency response and its effects are not annoying in audio applications. However, the effects of *operations quantization and word-length reduction* generate noise and non-linear distortion that could be audible and is undesirable for audio applications. Under several assumptions, with fixed-point arithmetic, this noise could be considered white and uncorrelated with the signal. However, using floating-point arithmetic, this noise depends on the signal magnitude due to the exponent value changes, and is somewhat correlated with the input signal and complicate to determine and evaluate a priori. In general, if the mantissa bit size for floating-point representation is equal to the bit size for a fixed-point representation, the floating-point one generates lower noise levels [Opp89].

FIR filtering solutions are much less sensible to these quantization effects because the error in the operations is not feedbacked inside the filter structure. Besides FIR filters have several disadvantages that make them not suitable for real-time and live audio applications due to the high delay introduced when equalizing low frequencies.

There are also IIR solutions to avoid the problems of low frequency IIR equalization, like double precision arithmetic or *warpped filters* [Kar99], that solve this problem, but always with an increase in design complexity and computational cost by a factor of three or more.

2.- PROBLEM ESTABLISHMENT

Let's examine the problem of low frequency IIR audio equalization analyzing what happens with a conventional IIR second order section (SOS) digital filter.

A conventional IIR SOS filter could be described by this transfer function *H(z)* :

$$H(z) = \frac{N(z)}{D(z)} = \frac{b_0 + b_1 \cdot z^{-1} + b_2 \cdot z^{-2}}{1 + a_1 \cdot z^{-1} + a_2 \cdot z^{-2}}$$

that is completely described by its five coefficients b_0, b_1, b_2, a_1 and a_2. The Direct Form II or *Canonical* filter structure is shown on *Figure 1,* where *x[n]* is the input of the filter, *y[n]* is the output and $w_0[n]$, $w_1[n]$ and $w_2[n]$ are the internal nodes that should be stored with a word length normally lower than the internal processor registers.

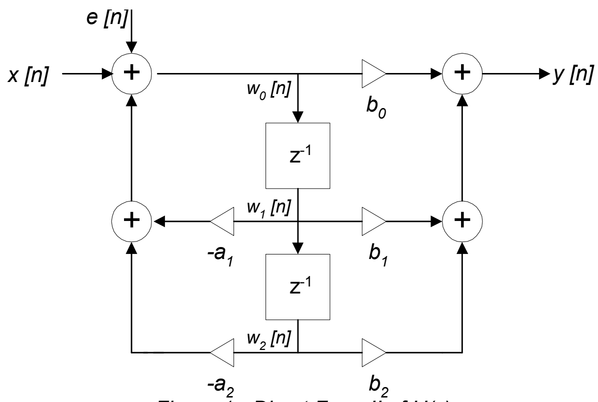


Figure 1 - Direct Form II of H(z)

We are going to focus the analysis on the recursive part of the filter *D(z)* since it is responsible of the mayor part of the *round-off* noise generated, due to the feedback effect. The quantization effect appears after each multiplication by *a1* and *a2* generating a noise, *e[n]*=*e1[n]*+*e2[n],* that is added to *x[n]* at the input of the filter. Lets *G(z)* be the transfer function of this error, and assuming that this error is not correlated with the signal, the L_2 norm of this function $||G||_2$ is [Jur64] :

$$G(z) = \frac{1}{1 + a_1 \cdot z^{-1} + a_2 \cdot z^{-2}} \qquad \|G\|_2 = \frac{1 + a_2}{1 - a_2} \cdot \frac{1}{(1 + a_2)^2 - a_1^2}$$

If a2 tends to 1, the poles are more close to the unit circle, ||*G*|| tends to infinite and the power of the quantization error increases rapidly.

Commonly, when working with IIR filters at low frequencies for audio equalization, the poles are near z=1. For clarifying this, let's take a typical audio example with a *parametric equalizer* at 50 Hz with a gain of 6 dB and a Q of 2. It is used for bass enhancement or to extend the low frequency band of the speaker. The normalized *H(s)* of a parametric filter is:

$$H(s) = \frac{s^2 + s\cdot(A/Q)+1}{s^2 + s/(A\cdot Q)+1} \quad , \quad A = \sqrt{10^{\frac{GaindB}{20}}}$$

Applying conventional bilinear transformation with a sampling frequency of 48kHz, the coefficients of the designed filter *H(z)* are : b_0=1.00115154 , b_1=-1.99764316, b_2=0.99653440, a_1=-1.99764316, a_2=0.99768594. Notice how a_2 is really close to 1. The *Figure 2* shows the magnitude frequency response of the digital filter designed *H(z)* and the quantization error transfer function *G(z)*. *At 50 Hz the magnitude of G(z) reaches 96 dB* so any error generated will be amplified too much degrading the dynamic range and reducing the quality of the audio. If we observe the evolution of the internal nodes of *H(z)* with *x[n]* being an impulse of amplitude 1, $w_{0,1,2}[n]$ arrive up to values of 31400 needing 16 bits more of resolution than the input data to represent this values. This demonstrates that conventional Direct Form II SOS structure has problems to filter correctly at low audio frequencies, generating too much noise even for 24 bit fixed point or 32 bit floating-point DSPs.

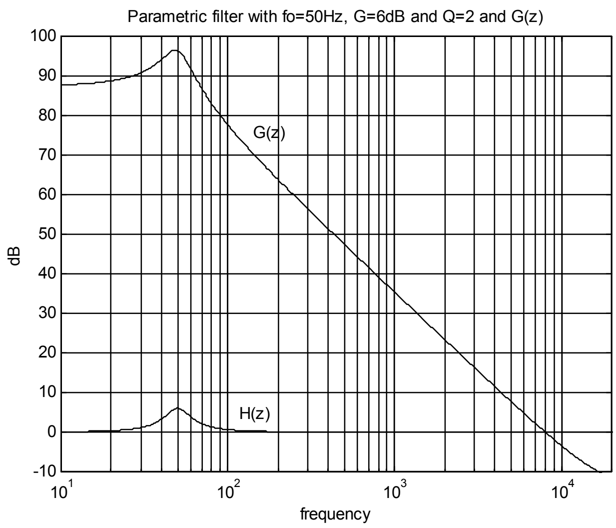


Figure 2

3. SOLUTION METHODS. SERIAL-PARALLEL DECOMPOSITION PLUS NOISE SHAPING

Other filter structures like Gold and Rader, Kingsbury or Zölzer [Zöl96], [Opp89] have been developed to solve the round-off quantization noise and improve the pole-zero discrete location for low frequencies. However, these filter structures require more computational cost (up to three times or more) than the Canonical one. The solution proposed here looks for a low computational cost solution with improved performance for low frequency audio equalization. This requires two steps:

- Use of First or Second order Noise-Shaping techniques to move the quantization noise at high frequencies.
- Parallel decomposition of the filter *H(z)* in *H(z)=1+Hp(z)* only when *Hp(z)* have *lowpass characteristic* in order to filter and eliminate the noise moved at the high frequencies by the previous Noise-Shaping stage.

3.1. Parallel Decomposition

This easy parallel decomposition will be done only when de parallel filter *Hp(z)* has *lowpass characteristic*. The generic structure for this filter is shown *at Figure 3*. The new coefficient *p* will be 1 if the filter *Hp(z)=H(z)-1* has *lowpass characteristic.* In this case, the new coefficients of the numerator will be $b_0'=b_0-1$, $b_1'=b_1-a_1$ and $b_2'=b_2-a_2$. If *Hp(z)* has no *lowpass characteristic* then *p=0* and the filter remains in Canonical form. In this case the filter will have *lowpass characteristic* intrinsically. *Table 1* shows parallel decomposition for common filters used in audio applications. S-domain has been used for simplicity:

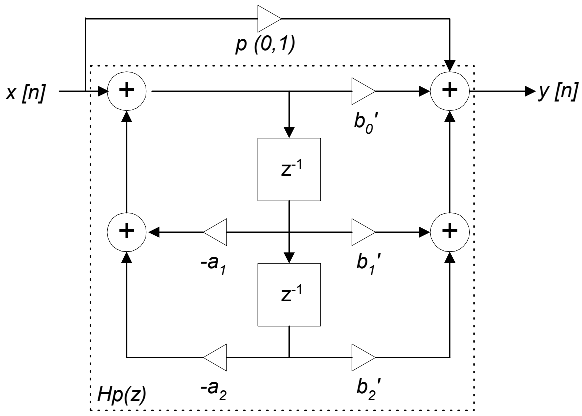


Figure 3 - Parallel Decomposition
*H(z)=p*1+Hp(z)*

Filter	*H(s)*	*Hp(s)=H(s)-1*	*LowPass ?*	*Coefficient p*
Parametric	$\frac{s^2+s\cdot(A/Q)+1}{s^2+s/(A\cdot Q)+1}$	$\frac{s\cdot(A^2-1)/(A\cdot Q)}{s^2+s/(A\cdot Q)+1}$	*Bandpass Filter YES*	*1*
Highpass	$\frac{s^2}{s^2+s/Q+1}$	$\frac{-s/Q-1}{s^2+s/Q+1}$	*Lowpass Filter YES*	*1*
Lowpass	$\frac{1}{s^2+s/Q+1}$	$\frac{-s^2-s/Q}{s^2+s/Q+1}$	*Highpass Filter NO*	*0*
Allpass	$\frac{s^2-s/Q+1}{s^2+s/Q+1}$	$\frac{-2\cdot s/Q}{s^2+s/Q+1}$	*Bandpass Filter YES*	*1*

Table 1 - Parallel Filter Decomposition H(s)=1+Hp(s)

3.2. Noise-Shaping

The use of Noise-Shaping is well known and used in oversampling converters for moving all the quantization noise out of the audible spectrum. This technique is also used in audio applications to distribute this round-off noise inside the audio band in a manner that will be less perceptible to the human ears, normally bring it to the high frequencies, [Hic95]. In this work we will use conventional noise shaping to bring all this noise at high frequencies and then eliminating it using the *lowpass characteristic* of filter *Hp(z)*.

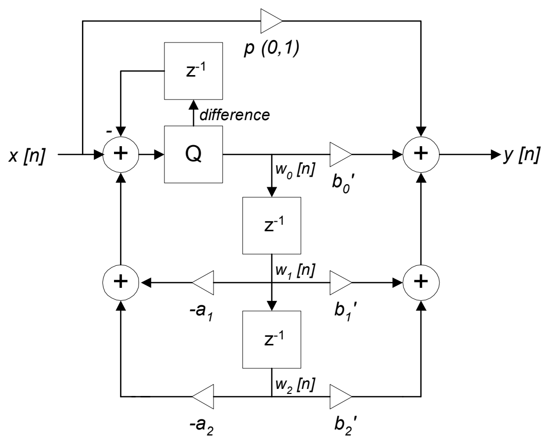


Figure 4 - H(z) with 1st order Noise-Shaping

To perform *First Order Noise-Shaping*, the quantization error should be feedbacked. In DSP systems the main quantization error is produced when the resulting operation at internal node $w_0[n]$ is stored in memory for implementing the delay line. Generally the word size of the memory is shorter than the internal processor's registers. This is represented at *Figure 4*. The transfer function of the error $G_{NS1}(z)$ is :

$$G_{NS1}(z)=\frac{1-z^{-1}}{1+a_1\cdot z^{-1}+a_2\cdot z^{-2}}$$

obtained *adding a zero to G(z)* due to the feedback. This new zero is only added to the transfer function of the error, not to the filter *Hp(z)*.

Second Order Noise-Shaping adds another zero to *G(z)* resulting in the filter structure of *Figure 5*. The error transfer function $G_{NS2}(z)$ is:

$$G_{NS2}(z) = \frac{1 - 2 \cdot z^{-1} + z^{-2}}{1 + a_1 \cdot z^{-1} + a_2 \cdot z^{-2}}$$

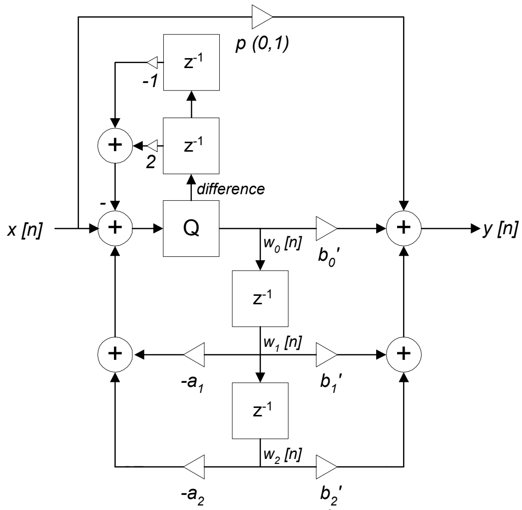


Figure 5 - H(z) with 2nd order Noise-Shaping

This new zero or zeros added to the transfer functions attenuate the effect of the poles at low frequencies. *Figure 6* represents the magnitude of *G(z)*, $G_{NS1}(z)$ and $G_{NS2}(z)$. With only one zero added, $G_{NS1}(z)$ improves more than 40 dB with low computational costs. With a double zero at $G_{NS2}(z)$, almost ideal performance is obtained, needing one multiplication and one memory more. It is also possible to arrive at $G_{NS2}(z)=1$ if the coefficients of the Noise-Shaping are the same than a_1 and a_2, placing the complex zeros at the same location that the poles.

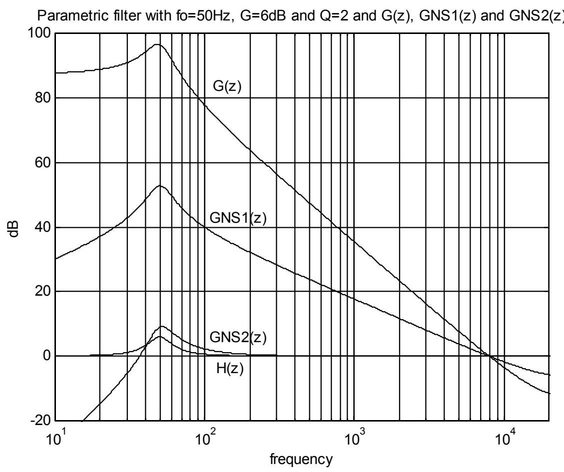


Figure 6 – Effect of Noise-Shaping

Figure 7 shows the effects of applying Noise-Shaping to the filter moving the round-off noise to the high frequencies, and how the parallel decomposition *H(z)=1+Hp(z)*, filters the high frequency noise by itself, providing *quasy-ideal* performance with very low increment in the computational costs.

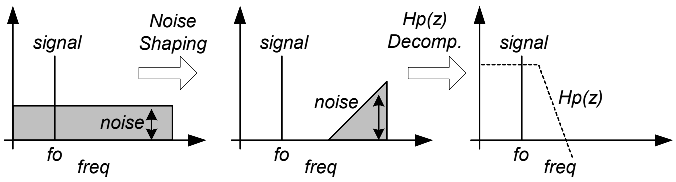


Figure 7

4.- FLOATING-POINT DSP IMPLEMENTATION AND MEASURES

This modified Canonical filter implementation has been tested on a TMS320C32 32bit floating-point DSP with 40 bit internal registers with the hardware implementation of *Figure 8*. This example shows an implementation of a second order allpass filter at 20 Hz with a Q of 3. The analysis is made with a FFT size of 65536 points with Hanning windowing. *Figures 9* and *10* explain algorithm goodness.

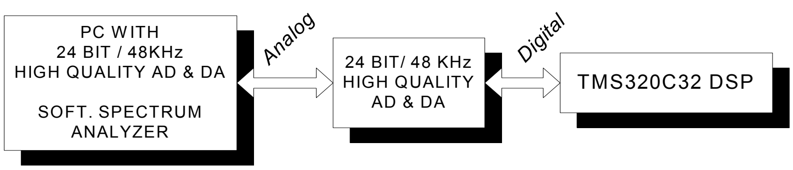
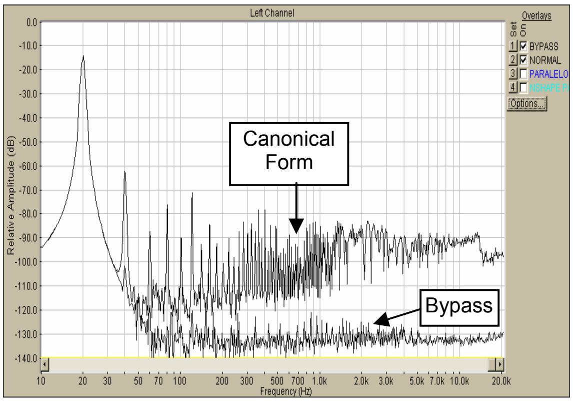


Figure 8 - Implementation



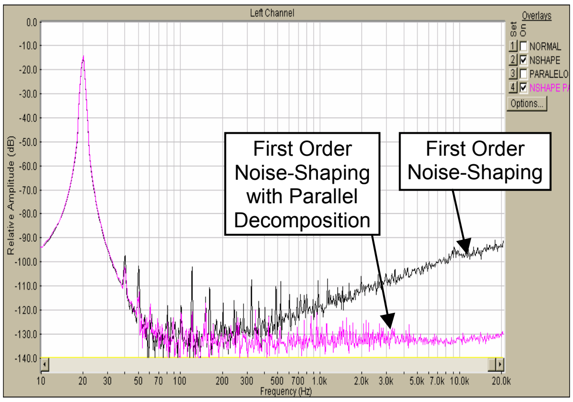


Figure 9 – Bypass of the system with 20 Hz input signal and Canonical implementation of the allpass filter degrading the signal more than 40 dB with noise and distortion.

Figure 10 – Canonical implementation with First Order Noise-Shaping where the round-off noise is moved to the high frequencies, and with Parallel Decomposition, where Hp(z) (bandpass) filters this noise, obtaining the same performance than the bypass measure.

Table 2 shows computational cost in clock cycles per SOS filter for the algorithms implemented. With a very low increase in computational cost compared to the Canonical Form it is possible to improve the behavior of the filter at low frequencies obtaining *quasy-ideal* performance.

Canonical Structure	5
Parallel Filter H(z)=1+Hp(z)	6
Canonical + First Order Noise-Shaping	9
Parallel Filter (allways) + First Order Noise-Shaping	9
Generic Parallel Filter + First Order Noise-Shaping	10

Table 2

5. CONCLUSIONS

A new structure for recursive filters that reduces the round-off noise at low frequencies with low computational cost has been proposed, analyzed and implemented. The algorithm consists in using noise-shaping within the audio band together with a parallel decomposition of the digital filter. The resulting system improves the quality of the SOS filter arriving to *quasy-ideal* performance with very low increase in the computational cost. The filter structure remains in canonical form, so it is efficient to implement on DSPs taking advantage of their special internal architecture (parallel instructions, addressing modes, etc). The system has been tested and evaluated on a real audio application using a floating point DSP.

This work was supported by Generalitat Valenciana, grant GV00-094-14

REFERENCES

[Gre95] Greenfield, R.G. and Hawksford, M.O.J.: On The Dither performance of High-Order Digital Equalization for Loudspeaker *JAES*, vol. 43, no. 11, pp 908-915, Nov. 1995

[Hic95] Christofer Hicks : *The Application of Dither and Noise-Shaping to the Nyquist-Rate Digital Audio : an Introduction*. Cambridge University Engineering Department.

[Jur64] E.I. Jury. *Theory and Applications of the Z-Transform Method*. Wiley 1964

[Kar99] Karjalinen M., Piirila E., Järvinen A.: *Comparision of Loudspeakers Equalization Methods Based on DSP Techniques. JAES Vol 47, pp 14-31 Jan 1999*

[Opp89] Alan V.Oppenheim, Ronal W.Schafe*r: Discrete-Time Signal Processing*. Prentice-Hall

[Orf96] Sopholcles J. Orfanidis: *Introduction to Signal Processing*. Prentice-Hall

[Poh95] Ken C. Pohlmann: *Principles of Digital Audio*. Mc Graw Hill

[Rez98] Rezso Dunay,István Kollár,Bernart Widrow:*Dithering for Floating-Point Number Representation*.1st Int. On-Line WorkShop on Dithering in Measurement. Mar 98

[Rob96] J. Robert Stuart and Rhonda J. Wilson: Dynamic Range Enhancement Using Noise-Shaped Dither at 44.1, 48, and 96 kHz.. AES 100th Convention.

[Wan97] Robert A. Wannamaker: *The Theory of Ditehred Quantization*. PhD. U.of Waterloo.

[Zöl94] Udo Zölzer: *Round-off Error Analysis of Digital Filters*. Vol 42,Nº 4 pp.232 JAES 1994.

[Zöl97] Udo Zölzer: *Digital Audio Signal Processing*. John Wiley & Sons