Frequency Warping in the Design and Implementation of Fixed-Point Audio Equalizers

Chalee Asavathiratham, Paul E. Beckmann, Alan V. Oppenheim

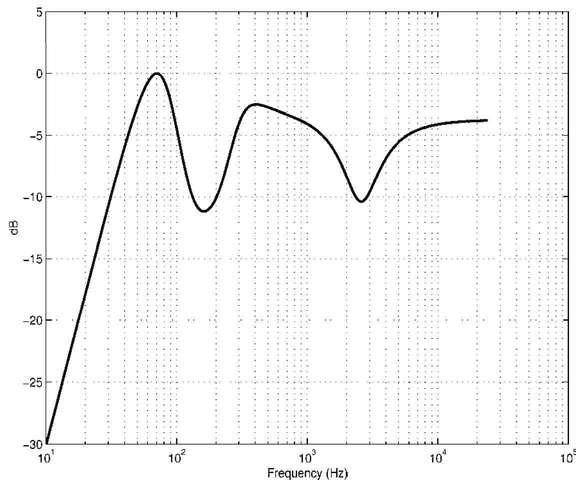

Fig. 1. Example of a typical audio equalizer's frequency response, plotted on a logarithmic frequency scale.

Abstract

This paper discusses frequency warping, a technique for designing and implementing discrete-time filters using allpass filters. This technique is particularly useful for audio filters because specifications are often given on a logarithmic frequency scale. It is shown that frequency warping allows a class of recursive filters to be designed using standard FIR techniques, and naturally leads to a structure for implementing the filters. The fixed-point behavior of this filter structure is analyzed and is shown to be relatively insensitive to coefficient quantization and round-off noise.

I. Introduction

In this paper, we apply the technique of frequency warping to the problem of designing filters for speech and audio applications. In these applications, logarithmically spaced frequency specifications often arise due to the natural frequency sensitivity of the human auditory system. Consider the problem of designing a filter to approximate the magnitude specification shown in Figure 1. The specifications are given using a logarithmic frequency axis and are typical of the requirements placed on audio equalizers used to compensate for room modes and loudspeaker responses. When viewed on a linear frequency scale, the majority of the detail in the frequency response occurs below 5 kHz (or below 0.23π if a sampling rate of 44 kHz is used) and this causes difficulty for most filter design techniques since they try to minimize error on a linear frequency scale.

The dense low frequency details cause problems for FIR design algorithms. Regardless of whether the "error" is actually measured in the least-square or Chebyshev sense, an FIR filter which fits the specifications shown in Figure 1 reasonably well requires at least a few hundred points. Weighted least-square, for example, could be used with an error weighting function which is heavily biased towards the low-frequency region. However, it tends to suffer from numerical problems caused by the narrow spacing of the low-frequency region. This eventually limits the order of the filter which can be designed. Likewise, the Parks-McClellan algorithm suffers numerical problems and fails to converge for these high lengths. Our experience shows that the Parks-McClellan algorithms fails to converge for *arbitrary* (not piecewise constant) responses when the filter order is greater than 35, assuming that double precision arithmetic is used in the design.

Alternatively, an IIR filter would offer a significant reduction in filter order. However, an IIR filter requires very high precision in both its design and actual operation. Furthermore, most IIR techniques do not directly address the design a filter of arbitrary magnitude response, and thus convergence to an acceptable solution is not guaranteed.

The allpass warping technique can be viewed as a hybrid of both design methods: it offers the flexibility of an FIR design algorithm, yet requires a low order similar to an IIR filter. This increased robustness is achieved by recognizing that it is much simpler to design a filter to fit a desired specification $D(\omega)$ on the log-scale than on linear scale. Thus by seeking a method of transforming the specification from $D(\omega)$ to $D(\Theta^{-1}(\omega))$, where $\Theta(\omega) \approx \log(\omega)$, the resulting filter order should be much smaller, and the design technique much more robust.

Frequency warping has been recognized for some time [1]-[4], and its primary use has been in filter design [5]. Recently, the ability of frequency transformations to approximate logarithmic frequency responses has been recognized. This property has been independently observed by [6]-[8].

II. Frequency Warping

Let $h[n]$ be an arbitrary real impulse response with z-transform $H(z)$. A function $G(z)$ is allpass-warped version of $H(z)$ if $G(z)$ can be obtained from $H(z)$ by replacing every delay with an allpass filter, i.e.,

$$G(z) = H(\Theta_a(z)) \quad (1)$$

where

$$\Theta_a^{-1}(z) = \frac{z^{-1} - a}{1 - az^{-1}} \text{ where } |\,a\,| < 1.$$

The subscript a is used to emphasize the dependence of the frequency response on the warping factor a.

Let $G(e^{j\omega})$ and $H(e^{j\theta})$ denote the Fourier transforms of $g[n]$ and $h[n]$ respectively. Note that different frequency variables are used for each transform. $G(e^{j\omega})$ and $H(e^{j\theta})$ are related by the substitution

$$\begin{aligned} G(e^{j\omega}) &= H(\Theta_a(e^{j\omega})) \\ &= H(e^{j\theta})\Big|_{\theta=\theta_a(\omega)} \end{aligned}$$

where the second equality follows from the fact that an allpass function maps the unit circle to itself. Thus, the phase response of the allpass $\theta_a(\omega)$ describes the mapping between the two frequency variables. It can then be shown that

$$\theta_a(\omega) = \arctan \frac{(1-a^2)\sin\omega}{-2a + (1+a^2)\cos\omega}. \tag{2}$$

The parameter a in $\theta_a(z)$ is called the *warping factor*. It is a free parameter that controls the warping from $H(z)$ to $G(z)$. Examples of the function for a few values of a are plotted in Figure 2. The function $\theta_a(\omega)$ is one-to-one on the interval $[0, \pi]$ and interestingly, its inverse is simply another allpass filter whose warping factor is $-a$. That is,

$$\theta_a^{-1}(\omega) = \theta_{-a}(\omega).$$

The warped response $G(e^{j\omega})$ can be visualized by starting with $H(e^{j\theta})$ and then manipulating the frequency axis. For $-1 < a < 0$, frequency warping stretches the low frequencies and compresses the high frequencies. Similarly, for $0 < a < 1$, the effect is the opposite — low frequencies are compressed and the high frequencies stretched. This stretching and compressing is the key benefit of frequency warping and yields a substantial reduction in filter order.

A filter with a log-scale frequency specification can be warped with $-1 < a < 0$ to approximately linearize its frequency response. A standard design technique is then applied to match this linear frequency response. Finally, the filter is warped again using $-a$ to return to the desired frequency scale.

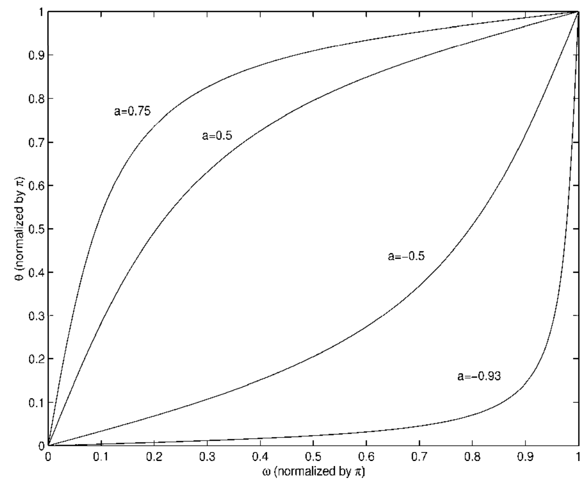

Fig. 2. Relationship between the warped (ω) and the prototype (θ) frequency variables for several warping factors.

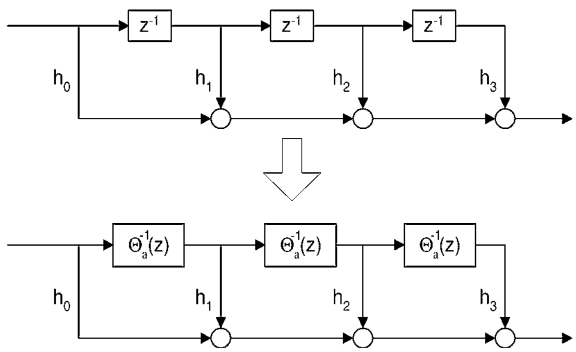

Fig. 3. Direct substitution of delays with allpass filters.

III. Implementation

Thus far, we have focused upon using frequency warping as a design tool. In this section, we show that it also leads to numerically robust filter structures, and analyze in detail the results of coefficient quantization and round-off noise.

Frequency warping is accomplished by the substitution of variables

$$z^{-1} \rightarrow \Theta_a^{-1}(z) = \frac{z^{-1} - a}{1 - az^{-1}} \text{ where } |\, a \,| < 1 \tag{3}$$

into the system function of a prototype filter $G(z)$. This substitution can also be performed directly on the system *structure* as shown in Figure 3.

That is, every delay in the system structure is replaced by an allpass filter. No further simplification is performed for two main reasons:

1. It is well-known that structures based on a cascade or sum of low-order sections are more robust than high-order structures.

2. First-order allpass filters remain allpass regardless of the quantization of the warping factor a, as long as the coefficients in the numerator and the denominator are quantized in the same manner. Thus, a true frequency warping is always realized. The quantization of a is of no real concern because a can be limited to realizable values during the design procedure.

Furthermore, there is no penalty in memory over higher order implementations.

Direct replacement, however, limits our consideration to structures obtained from FIR prototypes. This is because direct replacement in IIR filters leads to delay-free loops. This problem has been recognized and discussed in [8] and [9]. Therefore, while the paper has thus far considered frequency warping applied to arbitrary filters, both IIR and FIR, the rest of this section is devoted to warped FIR filters.

A. Fixed-Point Implementation

This section presents a detailed analysis of the warped FIR filter in the presence of fixed-point arithmetic. It is shown that by selecting a suitable allpass filter topology, the warped FIR structure yields a numerically robust scheme for implementing audio equalizers.

We assume that a modern digital signal processor is the target platform for the warped FIR filter. A DSP contains fixed bit-width memory and registers, and typically has a large bit-width accumulator for storing intermediate results. Quantization and truncation generally occur only when data is transferred from the accumulator back into a register or memory location, or output from the system. This quantization model differs from the traditional method of introducing a quantization error following every multiplication [5]. In terms of analyzing a signal flow graph, quantization occurs only when:

1. A multiplication is not uniquely and immediately followed by an addition, or
2. An addition is not uniquely and immediately followed by another addition.

We further assume that B-bit signed fractional arithmetic is used. Under this assumption, there are 2^B allowable values ranging from -1 to $1-\Delta$ where $\Delta = 2^{-B+1}$ represents a change in the least significant bit.

The errors resulting from fixed-point arithmetic can be placed into 3 main categories: coefficient quantization, signal overflow, and numerical noise. We analyze each type of error separately.

The warping factor a and the FIR coefficients $h[n]$ are typically designed using double precision arithmetic. These must then be quantized to the nearest representable value. Since there is some flexibility in the choice of warping factor, we assume that the warping factor is selected *a priori* from the set of allowable numerical values. Thus, the warping factor will not undergo any quantization. Once the warping factor is selected, the FIR coefficients $h[n]$ are designed and then quantized. It is well-known that the frequency response of FIR filters is relatively insensitive to the effects of coefficient quantization [5], and by extension, warped FIR filters are relatively insensitive as well.

Signal overflow occurs when the output of the filter, or a result stored in an intermediate delay element, exceeds the allowable range of $-1 \leq x \leq 1 - \Delta$. Although it is difficult to fully characterize the specific input conditions under which an overflow occurs, it is possible to develop some useful heuristics. A good starting point is to assume that the gain from the input of the system to all intermediate delay elements is less than or equal to 1 for all frequencies. If this condition is satisfied, then overflows will be completely avoided for sine wave inputs (in the steady state). If the gain exceeds 1, then the input must be scaled down accordingly, and this will reduce the overall signal-to-noise ratio (SNR) of the filter. We have found this to be a useful method of comparing the performance of different filter topologies. The best filter topology will require the least gain reduction of the input signal.

There are 4 possible topologies for the first-order allpass filter: direct form I, direct form II, and their transposes. Using our assumptions regarding the roundoff noise, and taking into account the possibility of numerical overflow, we conclude that the Direct Form I structure is the overall best choice.

The main apparent drawback of the Direct Form I structure is that it requires an additional delay element between every allpass stage. However, if we realize the filter $z^{-1}H(\Theta(z))$ rather than $H(\Theta(z))$, delay elements of neighboring stages may be combined. Thus, a cascade of N allpass filters can be implemented using only $N{+}1$ delay elements — one more than the canonical version. The final filter topology, including all quantization errors, is shown in Figure 4.

We make three assumptions regarding the quantization noise, and these parallel the presentation in [5]:

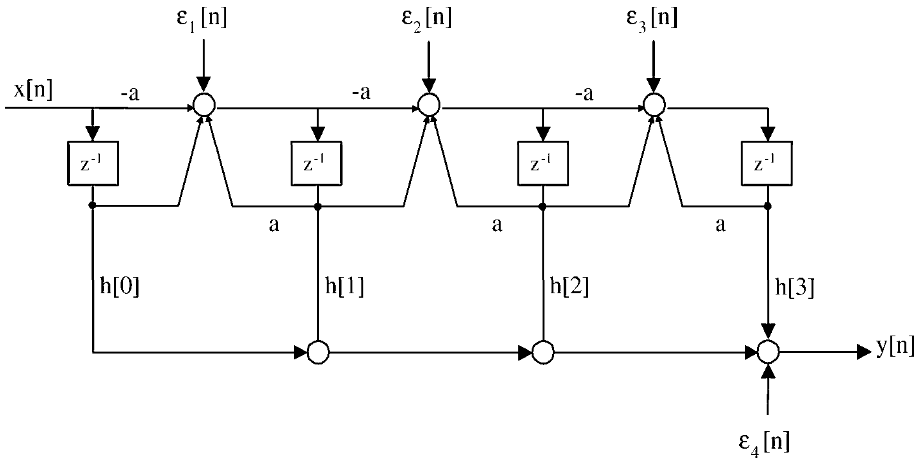

Fig. 4. Warped FIR structure with additive noise shown. $\epsilon_1[n]$, $\epsilon_2[n]$, $\epsilon_3[n]$ are due to the allpasses, while $\epsilon_4[n]$ is due to the FIR coefficients.

1. Each quantization noise $\epsilon_i[n]$ is a wide-sense stationary white noise process.

2. Each noise source has a uniform distribution over the range $\pm\frac{1}{2}\Delta$. Thus the variance of each noise source is $\sigma_\epsilon^2 = \frac{2^{-2B+2}}{12}$

3. Each quantization noise source is uncorrelated with the input to the quantizer, all other quantization noise sources, and the input to the system.

Since we assume that all of the noise sources are uncorrelated, the overall SNR of the filter is derived by summing all of the noise contributions at the output of the filter. Define $H_i(z)$ as the partial transfer function $H_i(z) \triangleq \sum_{k=i}^{M-1} h[k]z^{-(k-i)}$, where $h[0], \ldots h[M-1]$ are the FIR coefficients. Then the transfer function from the i^{th} noise source to the output would be $\frac{H_i(\Theta_a(z))}{1-az^{-1}}$. The noise spectrum at the filter's output can be expressed as

$$\begin{aligned} & \sigma_\epsilon^2 \sum_{i=1}^{M-1} \left| \frac{H_i(\Theta_a(e^{j\omega}))}{1-ae^{-j\omega}} \right|^2 + \sigma_\epsilon^2 \\ = & \frac{\sigma_\epsilon^2}{|\,1-ae^{-j\omega}\,|^2} \left(\sum_{i=1}^{M-1} |\, H_i(\Theta_a(e^{j\omega}))\,|^2 + 1 \right). \end{aligned} \tag{4}$$

It is difficult to draw general conclusions about the output noise spectrum because it is a function of the specific filter coefficients used. However, all of the terms in (4) have a common denominator of $1/\,|\,1-ae^{-j\omega}\,|^2$. The common term represents the noise shaping due to the pole in the allpass filter, and we can think of the output noise as being shaped by this first order low pass filter. In practice, we have found that the output noise spectrum is heavily influenced by this low pass characteristic. This is due to the choice of a which is typically close to 1. This is ideal for audio applications due to the decreased sensitivity of the human ear to low frequencies. Also, audio generally has more energy at low frequencies, and this tends to further mask the quantization noise.

B. Design Example

This section contains a complete example including design, implementation, and analysis of the resulting filter to the specification in Figure 1. The procedure consists of warping the specification with parameter a, designing an FIR filter, and implementing the warped-FIR filter with parameter $-a$.

The optimal warping factor is defined as the value of a which minimizes the length of the FIR filter. We determined the optimal warping factor by iteratively repeating the design procedure and searching for the smallest filter length which still met the specifications. The optimal a is usually close to the warping parameter which best linearizes the original log-frequency specifications, and we used this as a starting point. For the filter in Figure 1, the optimal warping factor was found to be $a = 0.963$.

Furthermore, additional saving in filter length may be achieved by designing it to be minimum-phase. We used a standard least-square design to obtain a linear-phase filter, then performed a spectral factorization on it. This yielded a 21-point prototype filter. The resulting frequency response is shown in Figure 5. As can be seen, warped filters can be used to match the frequency response even at relatively low order.

Figure 6 illustrates the noise at the output of the system for the case of 16-bit arithmetic. The solid curve represents the quantization noise measured at the output of the system (defined as the difference between the output of a 16-bit filter and a double precision filter, using white noise as the input). The dashed curve represents the noise predicted by Equation 4. Note that there is good agreement between the theoretically predicted and actually observed noise. Also, the overall shape of the noise spectrum is dominated by the lowpass term $1/\,|\,1 - ae^{-j\omega}\,|$ as mentioned earlier.

IV. Conclusion

In this paper, the theory and practical aspects of frequency warping based on allpass filters have been discussed. Allpass warping allows one to derive a filter from a prototype by

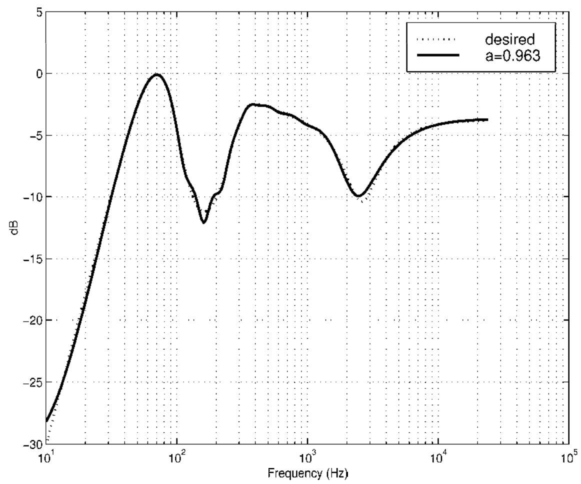

Fig. 5. Comparison of the desired (solid) and resulting (dashed) frequency responses.

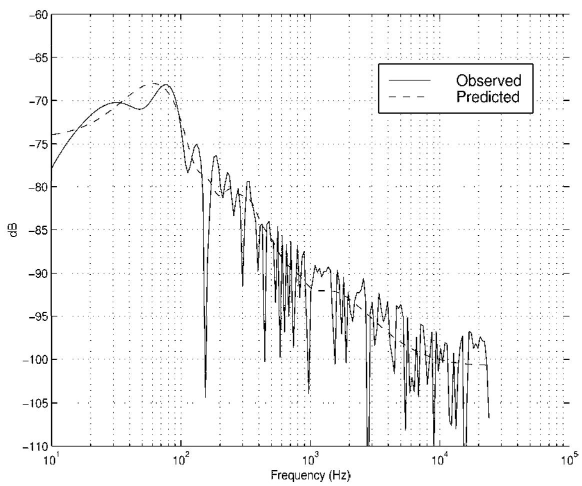

Fig. 6. Comparison of measured (solid) and predicted (dashed) quantization noise power spectral densities when 16-bit arithmetic is used. Length FIR prototype=21, $a = 0.963$.

replacing every delay with an allpass filter. One of the main benefits of frequency warping is its ability to yield closed form solutions for difficult filter design problems. This is especially true for filters whose specifications are highly nonuniform such as audio filters specified on a log-frequency scale. We also presented a structure for implementing filters using fixed point arithmetic. This was based upon a direct substitution of allpass filters directly into the system structure of a filter. Simulation results confirm that the quantization noise output is relatively low and has a low-pass characteristic.

References

[1] A. G. Constantinides, "Spectral Transformation for Digital Filters," *Proc. of the IEEE*, vol. 117, no. 8, pp. 1585-1590, Aug. 1970

[2] A. V. Oppenheim, D. H. Johnson, "Discrete Representation of Signals," *Proc. of the IEEE*, vol. 60, no. 6, pp. 681-691, 1972.

[3] C. Braccini, A. V. Oppenheim, "Unequal Bandwidth Spectral Analysis using Digital Frequency Warping," *IEEE Trans. on Acoustics Speech and Signal Processing*, vol. 22, no. 4, Aug. 1974.

[4] H. W. Strube, "Linear Prediction on a Warped Frequency Scale," *J. Acoust. Soc. Am.*, vol. 68, no. 4, Oct. 1980.

[5] A. V. Oppenheim, R. W. Schafer, *Discrete-Time Signal Processing*, McGraw-Hill, 1989.

[6] C. Asavathiratham, *Digital Audio Filter Design Using Frequency Transformations*, M. Eng. Thesis, Elec. Eng. and Comp. Sci , MIT, 1996.

[7] J. O. Smith, J. S. Abel, "The Bark Bilinear Transform," *Proc. IEEE Workshop on Appl. Signal Processing to Audio and Acoustics, New Paltz, NY*, Oct. 1995, IEEE Press. Available online at http://www-ccrma.stanford.edu/~jos.

[8] M. Karjalainen, A. Harma, U. K. Laine, "Realizable Warped IIR Filters and Their Properties", *IEEE Int. Conf. on Acoustics, Speech, and Signal Processing*, vol. 3, pp. 2205-8, Apr. 1997.

[9] K. Steiglitz, "A Note on Variable Recursive Digital Filters," *Proc. ASSP*, vol. 28, no. 1, pp. 111-112, Feb. 1980.