QUANTIZATION EFFECT ON PHASE RESPONSE AND ITS APPLICATION TO MULTIPLIERLESS ANC

Yunhua Wang, Linda S. DeBrunner, Victor E. DeBrunner, Dayong Zhou
School of Electrical & Computer Engineering;
University of Oklahoma
Norman, OK 73019, USA
{xiao9, ldebrunnner, vdebrunn, dayong}@ou.edu

ABSTRACT

Adaptive filtering is a widely used technique in active noise control (ANC). In order to make the adaptive filter in an FXLMS ANC system stable, the reference signal must pass through an estimation filter whose phase response is within ±90° of the phase of the secondary path. In this paper, we study the quantization effects on the filter phase response and the relationship between the phase response and the location of the zeros and poles. In addition, we propose a filter structure and nonuniform quantization method in which we quantize the filter coefficients so they each contain a small number of nonzero bits based on the distance of the zeros/poles to the unit circle to guarantee that the ±90° allowable phase deviation is met — greatly reducing the implementation cost. We combine these ideas with that of multiplierless implementations of adaptive FIR filters to realize an efficient active noise control using field programmable gate arrays or other digital hardware.

1. INTRODUCTION

Two types of acoustic noise exist in the environment: broadband noise and narrowband noise. Active noise control (ANC) techniques can successfully be used to reduce both the narrowband and broadband noise that exists in many applications. The filtered-x LMS (FXLMS) and adjoint LMS are two widely applied techniques [1].

The FXLMS algorithm is illustrated in Fig. 1, where

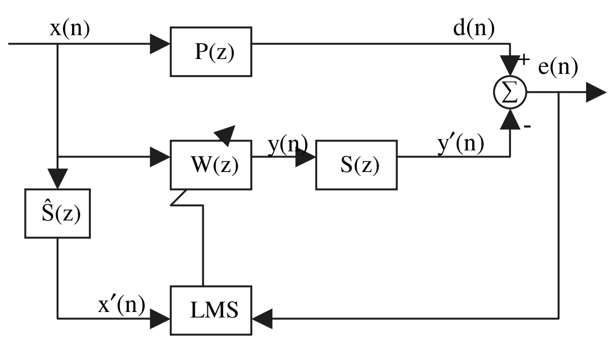

Fig. 1. FXLMS active noise canceller

the output $y(n)$ is:

$$y(n) = \sum_{i=0}^{N} w_i(n) \cdot x(n-i) \tag{1}$$

where $w_i(n)$ is the i[th] coefficient of the adaptive FIR filter $W(z)$ at time *n*, and $x(n)$ is the reference signal vector at time *n*. The error signal is

$$e(n) = d(n) - y'(n) = d(n) - s(n) * y(n) \tag{2}$$

where $s(n)$ is the impulse response of the secondary-path $S(z)$ at time *n*. Then the FXLMS update algorithm is

$$w(n+1) = w(n) - \mu * e(n) * x'(n) \tag{3}$$

where μ is the step size of the algorithm. Therefore, the input vector $x(n)$ is filtered by $S(z)$ before updating the weight vector. However, in practical applications, $S(z)$ is unknown and must be estimated by the filter $\hat{S}(z)$

$$x'(n) = \hat{s}(n) * x(n) \tag{4}$$

The main idea behind the filtered-x LMS algorithm is to keep the reference signal $x(n)$ and the error signal $e(n)$ aligned in time. In order to do that, the reference signal must be filtered by a filter $\hat{S}(z)$ that should have the same phase response as the secondary path $S(z)$. According to the research in [2], the filtered-x LMS algorithm is robust in the sense that it will converge when there is no more than 90° of phase error between $\hat{S}(z)$ and the secondary path $S(z)$ at all frequencies. Based on this knowledge, we could relax the precision constraints on the adaptive filter. Which means we could use only limited bits to quantize $\hat{S}(z)$, and still have it converge with sufficient precision to work in the ANC problem. Thus, the method's robustness can be exploited to produce a low complexity implementation.

In Section 2, we describe the quantization effects on the phase response of the filter and their relationship to the pole and zero locations. Then, we propose a quantization architecture in Section 3. Multiplierless implementation of ANC is discussed in Section 4. Simulations are given in Section 5. Conclusions are presented in Section 6.

2. QUANTIZATION EFFECTS ON PHASE RESPONSE

Quantization is one of the most important issues when implementing a filter using VLSI hardware. The definition of the quantization function affects not only the hardware requirements, but also the performance of the filter. By identifying a good quantization approach, we can significantly reduce hardware complexity.

For our ANC application, we need to find a quantization technique that affects phase response as little as possible. Consequently, we must determine the relationship between the phase errors and the quantization functions. For example, consider an FIR filter and its coefficients quantized version

$$H(z) = a_0 + a_1 z^{-1} + a_2 z^{-2} \cdots + a_N z^{-N} \tag{5}$$

$$\hat{H}(z) = \hat{a}_0 + \hat{a}_1 z^{-1} + \hat{a}_2 z^{-2} \cdots + \hat{a}_N z^{-N} \tag{6}$$

where $H(z)$ is the infinite precision transfer function. Quantization of the filter coefficients results in a new transfer function $\hat{H}(z)$, related to $H(z)$ via the quantization function $Q(z)$

$$H(z) = \hat{H}(z) \cdot Q(z) \tag{7}$$

$$H(e^{j\omega}) = \hat{H}(e^{j\omega}) \cdot Q(e^{j\omega}) \tag{8}$$

so,

$$\begin{aligned} Q(e^{j\omega}) &= \frac{H(e^{j\omega})}{\hat{H}(e^{j\omega})} \\ &= \frac{(e^{j\omega} - k_1)(e^{j\omega} - k_2)\cdots\cdots(e^{j\omega} - k_N)}{(e^{j\omega} - \hat{k}_1)(e^{j\omega} - \hat{k}_2)\cdots\cdots(e^{j\omega} - \hat{k}_N)} \\ &= \left|Q(e^{j\omega})\right| \cdot e^{j\phi(\omega)} \end{aligned} \tag{9}$$

where $\phi(\omega) = \arg\{Q(e^{j\omega})\}$ is the phase response of $Q(z)$, and $\left|Q(e^{j\omega})\right|$ is the magnitude response of $Q(z)$; $k_1, k_2, \cdots k_N$ are zeros of FIR filters.

$$\phi(\omega) = \sum_{i=1}^{N} \arg\{\frac{e^{j\omega} - k_i}{e^{j\omega} - \hat{k}_i}\} = \sum_{i=1}^{N} \theta_i \tag{10}$$

Fig. 2 shows that the phase error for a filter is related to zero locations. The quantization phase error is equal to the sum of the θ_i. Similar results hold for IIR filters [3], [4]. The closer the poles and zeros are to the unit circle, the bigger the θ_i, and thus the phase error. This relationship leads us to propose a cascade form filter structure with each section characterized by either a first-order or a second-order transfer function, which is less sensitive to the quantization:

$$H(z) = a_0 \prod_{m=1}^{M} (1 + \beta_{1m} z^{-1} + \beta_{2m} z^{-2}) \prod_{n=1}^{N} (1 + \alpha_n z^{-1}) \tag{11}$$

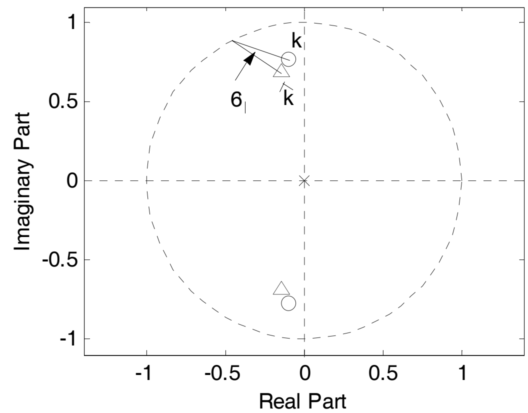

Fig. 2. The relations between quantization error θ_i and the poles and zeros locations

For the first-order part, zeros are just equal to $-\alpha_n$; for the second-order part, zeros are equal to $\frac{-\beta_{1m} \pm \sqrt{\beta_{1m}^2 - 4\beta_{2m}}}{2} = \sqrt{\beta_{2m}}$. When $\alpha_n = 1$ or $\frac{\beta_{1m}}{2} = 1$ and $\sqrt{\beta_{2m}} = 1$, the zeros will be on the unit circle.

Consequently, our quantization rule is to use more bits to represent the coefficients when their zeros or poles are near unit circle; otherwise, when the zeros and poles are far from the unit circle, we use less bits for the coefficient to reduce the hardware complexity.

3. QUANTIZATION ARCHITECTURES

From our previous discussion, we know that the closer the poles and zeros are to the unit circle, the bigger the phase error is. Based on this understanding, we propose a quantization function that quantizes each coefficient value to limited combinations of powers-of-two, such as the sum or difference of two or three powers-of-two, which leads to an efficient multiplierless approach. Fig. 3 shows the corresponding implementation.

In binary arithmetic, multiplication by a power-of-two is implemented in hardware structures simply by a shift operation, which can be implemented by properly "wiring" the circuit. Thus, we can realize the filter coefficients using a few adders/subtractors along with the routing required for shifting [5], [6], [7]. In our case, based on the distance to the unit circle, we separate the coefficient to four intervals:

Interval I:

$$-\frac{1}{2} \le \alpha_n \le \frac{1}{2}, \qquad -\frac{1}{4} \le \beta_{2m} \le \frac{1}{4} \quad \text{and} \quad -1 \le \beta_{1m} \le 1$$

For this interval the zeros are relatively far from the unit circle, so we use only two power-of-two bits to represent

them. This implementation is shown in Fig. 3(a).
Interval II:

$$\frac{1}{2} < |\alpha_n| \le 1\frac{1}{2}, \qquad \frac{1}{4} < |\beta_{2m}| \le 2\frac{1}{4} \qquad \text{and} \qquad 1 < |\beta_{1m}| \le 3 .$$

Phase error is more sensitive to coefficients in this interval, so we use three power-of-two bits to represent them. This implementation is shown in Fig. 3(b).
Interval III:

$$1\frac{1}{2} < |\alpha_n| \le 3, \qquad 2\frac{1}{4} < |\beta_{2m}| \le 9 \qquad \text{and} \qquad 3 < |\beta_{1m}| \le 6 .$$

As in Interval I, the zeros are relatively far from the unit circle, so we use two power-of-two bits to implement these coefficients.
Interval IV:
The larger coefficients outside the ranges of Intervals I, II, and III, are far from the unit circle, and so the effect of quantization of phase is small enough that a single bit can be used to represent the coefficient!

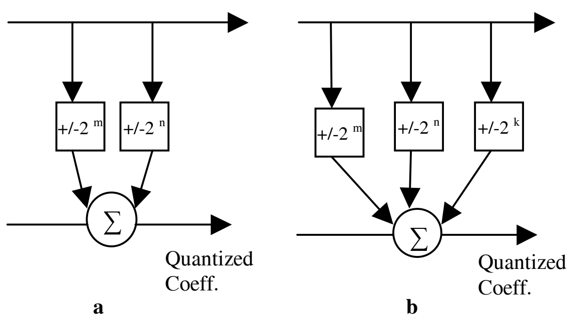

Fig.3.a.Quantized coefficients with two power-of-two bits
b.Quantized coefficients with three power-of-two bits

Here we quantize a second order sub-system $1+\beta_{1m}z^{-1}+\beta_{2m}z^{-2}$ based on above method. The absolute value of the maximum phase error histogram is given in Fig. 4. Note that in the calculation, we ignore the points very close to the unit circle, because for an FIR system this means that the magnitude is very small and for an IIR system these points yield unstable systems. We can see that most of the large phase errors are within 10º. Thus, the chances are very slight that we can accumulate phase errors larger than 90º for a high order adaptive filter using the cascade form implementation.

4. MULTIPLIERLESS IMPLEMENTATION OF FX-LMS ANC

To implement FX-LMS for ANC, we first need to implement the Least Mean Square (LMS) adaptive FIR filter. Since LMS requires adaptation of the filter coefficients, implementations typically require multiplier circuits that can multiply different values. In earlier work [8], we developed a multiplierless adaptive FIR filter by using a limited number of nonzero canonical signed digit (CSD) bits to create a multiplier circuit using shifts and adds. We also modified the adaptation algorithm so that it could be performed using a single shift. To implement the FXLMS ANC, we need two adaptation processes. One is $w(z)$, which is required for system identification, and the other is $\hat{S}(z)$, which is used to estimate secondary path $S(z)$. A detailed diagram showing the implementation of FXLMS ANC is shown in Fig. 5.

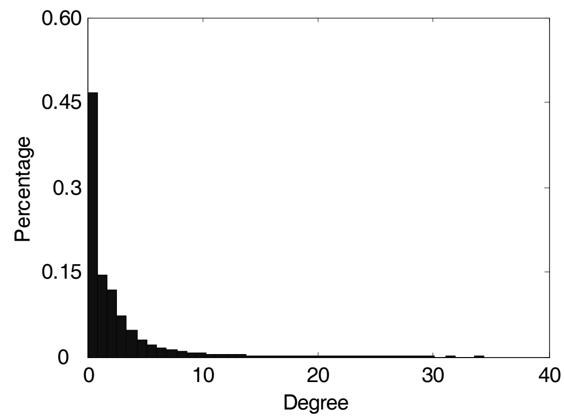

Fig. 4. The histogram of the maximum absolute phase error

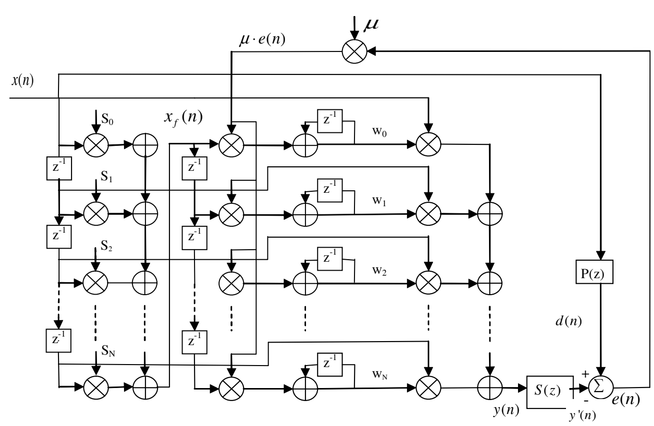

Fig. 5. Block diagram for Nth -order FX-LMS

5. SIMULATION RESULT

Consider the 12th order IIR filter $\hat{S}(z)$ that is used to estimate the secondary path $S(z)$ in an ANC system with coefficients:

$$\frac{1-1.0347z^{-2}+2.5244z^{-4}-1.5362z^{-6}}{1+0.8254z^{-2}+2.6216z^{-4}+1.3456z^{-6}}$$
$$\frac{+1.8239z^{-8}-0.5401z^{-10}+0.3771z^{-12}}{+2.1235z^{-8}+0.5401z^{-10}+0.5314z^{-12}}$$

Fig. 6 shows the results of implementing this filter without quantization and also using three different quantization methods: quantization to the 5 most significant digits, quantization of direct form to 3 nonzero CSD digits, and our quantization algorithm. As shown in Fig. 6, our method works quite well. Next, we apply these quantized filters in the ANC FX-LMS algorithm shown in Fig. 5. The filter resulting from our quantization method is stable and removes significant amounts of noise, while the other two quantization methods actually diverge, (see Fig. 7).

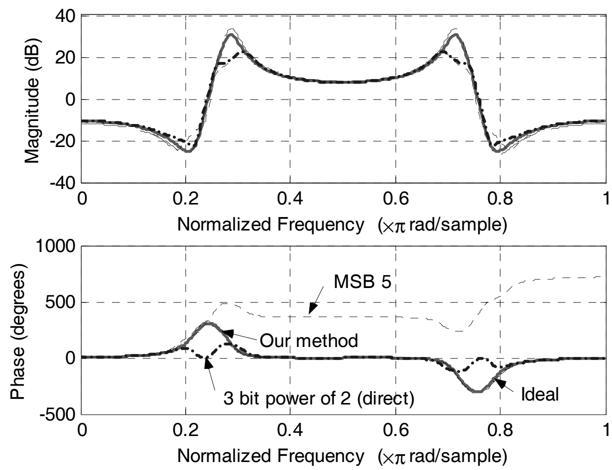

Fig. 6. Comparison of filter magnitude and phase response of three different quantization methods.

6. CONCLUSION

For a long time, many researchers have considered the effect of quantization on magnitude, with little consideration given to phase errors. We described the effect of quantization on phase response and its application to ANC implementation using FX-LMS. By choosing a quantization function that varies depending on the location of the poles and zeros, we can reduce the hardware required. Our quantization method provides more bits for those coefficients that have a greater effect on the phase error (i.e. poles and zeros that are close to the unit circle). Our method allows us to reduce the phase error while reducing the hardware complexity. We incorporate this approach into the implementation of a multiplierless FX-LMS active noise controller. The reduced complexity makes FX-LMS more desirable for ANC implementations.

7. REFERENCES

[1] Sen M. Kuo, Dennis R. Morgan, *Active noise control systems: algorithms and DSP implementations,* John Wiley & Sons, 1996.

[2] S. D. Snyder and C. H. Hansen, "Effect of transfer function estimation errors on the filtered-X LMS algorithm," *IEEE Transactions on Signal Processing*, 42, (4), pp. 950-953. 1994.

[3] A. V. Oppenheim and R. W. Schafer, *Digital Signal Processing*, Prentice-Hall, 1975.

[4] Victor DeBrunner, *Digital Filter Sensitivity*, M.S. Thesis, Virginia Polytechnic Institute and State University, 1986.

[5] Henry Samueli, "The design of multiplierless digital data transmission filters with powers-of-two coefficients," *Telecommunications Symposium, SBT/IEEE International,* pp. 425 -429, 3-6 Sept. 1990.

[6] H. Samueli, "An improved search algorithm for the design of multiplierless FIR filters with power of two coefficients," *IEEE Trans. Circuits and Systems*, vol. 36, no. 7, pp. 1044-1047, Jul. 1989.

[7] X. Hu, L. S. DeBrunner, and V. DeBrunner, "An Efficient Design for FIR Filters with Variable Precision," *Proc. of the IEEE ISCAS'02,* Phoenix, AZ, May 2002.

[8] L. S. DeBrunner, Y. Wang, V. DeBrunner, and M. Tull, "Multiplierless implementations of adaptive FIR filters," *accepted for publication in the Asilomar Conference on Signals, Systems, and Computers, to be held November* 2003.

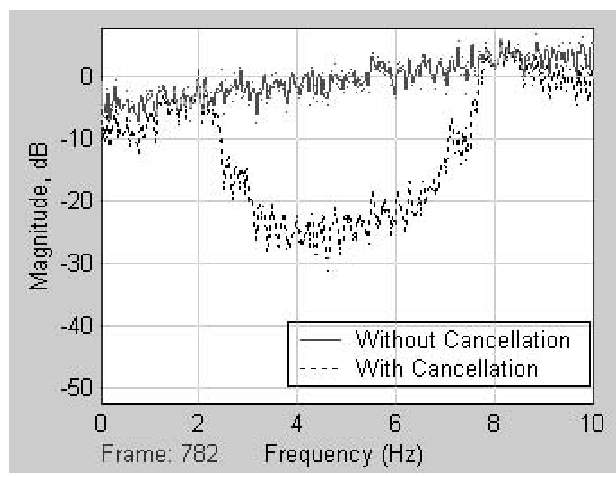

Fig.7. PSD result of the ANC system using our quantization method