A NOVEL VOLTERRA-WIENER MODEL FOR EQUALIZATION OF LOUDSPEAKER DISTORTIONS

Khosrow Lashkari

DoCoMo Communications Laboratories USA Inc.
San Jose, CA, USA

ABSTRACT

Future mobile multimedia devices such as videophones will require high playback levels where devices are used at an arm's length from the listener. When small loudspeakers are driven at high playback levels the nonlinear characteristics of these speakers become a major source of sound degradation and can drop the MOS score by as much as one point. Conventional approaches to loudspeaker compensation based on the Volterra model improve the sound quality only at low playback levels and may introduce more distortion at high playback levels. This paper presents a new Volterra-Wiener model that is a better match to the loudspeaker response and lends itself to having an exact nonlinear inverse. Simulations and real measurements show that compensation based on the new model greatly reduces the linear and nonlinear distortions of small loudspeakers mounted in the cell phones. Objective and subjective evaluations show drastic improvements (0.8 PESQ) in the sound quality.

1. INTRODUCTION

With advances in coding, transport and radio access technologies, there is a growing trend toward providing multimedia services on wireless devices. Enhanced multimedia services such as mobile entertainment and videophones require delivery of higher quality speech and audio over small devices at *high sound levels*. In voice applications, the phone is often very close to the ear and a low sound level from the loudspeaker is acceptable (except in the speaker phone scenario). In multimedia applications, the terminal (videophone, PDA or screen of a laptop computer) is at about an arm's length from the user requiring higher sound levels. To get high sound volumes from small loudspeakers, loudspeakers are driven close to their maximum rated drive signals.

Loudspeakers in general have both linear and nonlinear distortions. When small electrodynamic loudspeakers are driven close to their maximum rated power, they exhibit significant nonlinearity. The nonlinear distortions introduced into the sound result in noticeable audible distortions and degradation in the perceived quality.

The nonlinearity is inherent in the electroacoustic conversion process and is due to both the electrical circuit and the mechanical suspension system. The electrical nonlinearity is due to the variable inductance caused by the motion of the voice coil. The mechanical nonlinearity arises from the nonlinear compliance of the suspension system.

Research in mobile applications has primarily focused on the codec part to improve the sound quality. State-of-the-art speech and audio codecs such as wideband AMR and high efficiency AAC (HE-AAC or aacPlus) provide sufficient quality in mobile applications before analog playback. Consequently, in the source to destination path of the mobile audio shown in figure 1, small loudspeakers constitute the bottleneck in the overall sound quality. Compared to headphones, small loudspeakers may reduce the mean opinion score (MOS) by up to one point depending on the codec [6]. Compensating for the distortions of the loudspeaker is therefore very important.

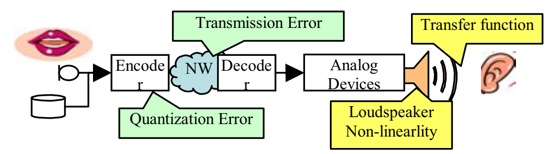

Figure 1: Source-to-destination path for mobile audio

There have been many attempts in the past to compensate for the distortions of electrodynamic loudspeakers. To compensate for the effects of the electroacoustic conversion, the signal is passed through a predistortion filter placed between the audio signal source and the loudspeaker as shown in figure 2. A model H of the loudspeaker is derived from physical measurements of the loudspeaker. Using this model, a predistortion filter G also known as a precompensator or a pre-inverse is then computed.

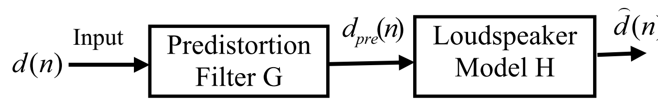

Figure 2: Predistortion Filter for Loudspeaker Compensation

The desired input speech or audio signal (for example from the decoder output) $d(n)$ is fed into the predistortion filter. The loudspeaker's moving coil is driven by the predistorted signal $d_{pre}(n)$; the filtering is designed to ideally be inverse of the loudspeaker's distortion, so that the actual displacement of the moving coil accurately matches the ideal motion prescribed by the original signal $d(n)$.

Both physical and mathematical models of the loudspeaker have been attempted. Adaptive nonlinear filters for loudspeaker compensation have been studied for some time, and applied in other fields [1-3]. Most approaches model the loudspeaker by a truncated Volterra series as described in section 2. The predistortion filter is then taken to be the p-th order inverse of the Volterra model as described in [4].

Predistortion filters based on the p-th order inverse of the Volterra model completely compensate for the linear distortion and some of the nonlinear distortions [2]. However, they can introduce higher order distortions as a by-product. At low playback levels, the higher order distortions are small and the conventional p-th order compensation scheme improves the sound quality.
However, at higher playback levels necessary for multimedia applications, the extra nonlinear distortions introduced by this method could be more than the distortions of the no-compensation case. Two factors contribute to this problem: 1) the p-th order inverse is only an approximation to the exact nonlinear inverse of the Volterra model, and 2) the Volterra model of the loudspeaker does not lend itself to having an exact nonlinear inverse.
This paper presents a novel modified Volterra-Wiener model for characterizing a loudspeaker's response. The new model provides a better match to the loudspeaker's response and also lends itself to having an *exact nonlinear inverse* in the sense described in [6].
The organization of the paper is as follows. Section 2 briefly describes the Volterra model. Sections 3 and 4 introduce the new Volterra-Wiener model. Section 5 discusses the exact invertibility of the model. Section 6 presents the results. Finally section 7 contains some concluding remarks.

2. VOLTERRA MODEL OF THE LOUDSPEAKER

In this section we briefly present the Volterra model. Volterra expansion is a general method for modeling nonlinear systems with soft or weak nonlinearities. This includes saturation-type nonlinearities observed in power amplifiers and loudspeakers. A truncated p-th order Volterra expansion is given as:

$$y(n) = \sum_{k=0}^{p} H_k[x(n)] \quad (1)$$

In this representation, $H_k = h_k(m_1, m_2, m_3, ...m_k)$ is the k-th order operator and $h_k(.,.,...)$ is called the k-th order Volterra kernel or generalized impulse response. The k-th order operator $H_k[x(n)]$ is given as:

$$H_k[x(n)] = \sum_{m_1=0}^{M_k}\sum_{m_2=0}^{M_k}..\sum_{m_k=0}^{M_k} x(n-m_1)x(n-m_2)....x(n-m_k) \quad (2)$$

Here M_k is the memory of the k-th kernel. Volterra series can be regarded as a Taylor series with memory or as the extension of FIR filters to nonlinear systems. Small loudspeakers can be sufficiently modeled by a 2nd or 3rd order Volterra system. The 2nd order model is given as:

$$y(n) = h_0 + \sum_{i=0}^{M_1} h_1(i)x(n-i) + \sum_{m_1=0}^{M_2}\sum_{m_2=0}^{M_2} h_2(m_1,m_2)x(n-m_1)x(n-m_2) \quad (3)$$

The first term is a constant and is generally assumed to be zero, the second term is the linear response (H_1), and the third term is the quadratic response (H_2). Figure 3 shows the 3rd order Volterra model of the loudspeaker based on (1). The model usually consists of several thousand coefficients. These coefficients are found by minimizing the weighted mean square error (WMSE) between the model output and the desired signal.

$$E = \sum_{n=0}^{N} \lambda^{N-n}(d(n) - y(n))^2 \quad (4)$$

Here, λ is the weight factor, N is the adaptation length and $d(n)$ is the desired loudspeaker output. This minimization is accomplished by using the LMS or the RLS algorithms [5].

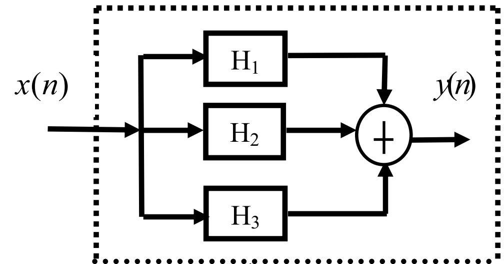

Figure 3: 3rd Order Volterra Model

3. THE NEW VOLTERRA-WIENER MODEL

In modeling the characteristics of small loudspeakers, we observed several shortcomings of the Volterra model. Firstly, the low frequency response of the Volterra model is not a good match to the loudspeaker's response. The predicted output does not match the output of the real loudspeaker at lower frequencies. Secondly, the distortions predicted at high sound levels are higher than the actual distortions of the loudspeaker. As seen in figure 3, the Volterra model consists of a parallel combination of linear and nonlinear subsystems. Figure 4 shows the linear transfer function (H_1 in figure 3) of the loudspeaker. We used 150 linear taps, a memory of 40 samples for the quadratic part and a memory of 20 samples for the cubic component. As seen here low frequencies cannot be reproduced by a small loudspeaker.

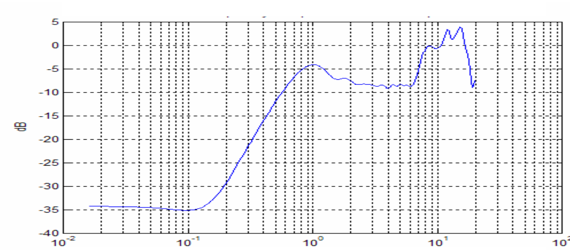

Figure 4: Transfer function of the linear part of the Volterra model (dB vs. frequency in Hz)

Although a low frequency tone may not produce any response from the loudspeaker, due to the parallel structure of the Volterra model, its 2nd or 3rd harmonics can still be large especially at high sound levels. Finally, a major shortcoming of the Volterra model is that it does not lend itself to having an exact nonlinear inverse in the sense described in [6]. The primary reason is that the linear transfer function H_1 is usually non-minimum phase and does not have a *zero-delay* stable inverse. Lack of a zero-delay linear inverse causes the exact inverse of the Volterra model to be unstable. The proposed modified Volterra-Wiener model shown in figure 5 overcomes all of these shortcomings.
The modified Volterra-Wiener (MVW) model is a *cascade* combination of a linear system and a *modified* Volterra system.

Here the input signal $x(n)$ first passes through a linear system H'_1 producing the output $y(n)$ which is then fed to a *modified* Volterra system where the linear subsystem H_1 is forced to be unity ($H_1 = 1$).

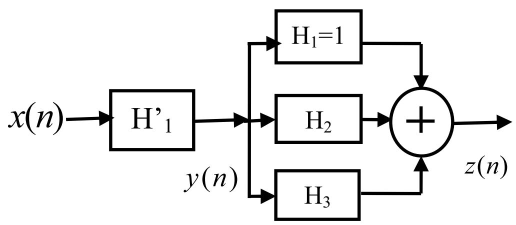

Figure 5: The modified Volterra-Wiener model

The linear system H'_1 filters the input signal first ensuring that a tone with a small fundamental does not produce large harmonics. The modified Volterra system is a p-th order nonlinearity with memory that provides the degrees of freedom necessary to model the loudspeaker. The cascade structure of the new model in conjunction with a modified stable linear component in the Volterra part of the model (i.e. $H_1 = 1$) causes the new model to have a stable nonlinear inverse. The input-output relationship of the MVW model is given as:

$$z(n) = \sum_{k=0}^{P} H_k[y(n)] \tag{5}$$

The output of the linear system $y(n)$ is given as:

$$y(n) = \sum_{i=0}^{M} a(i)x(n-i) \tag{6}$$

Here, M is the memory and $\{a(i),..i = 0,1,....M\}$ are the coefficients of the linear filter.

The Volterra-Wiener model combines the desirable feature of the Volterra model (memory for nonlinear kernels) and the Wiener model (existence of an exact nonlinear inverse) to overcome the shortcomings of the Volterra model.

4. COMPUTING THE MODEL PARAMETERS

There are two methods to compute the parameters of the linear system and the modified Volterra model. The first method is a two-step process that is fast but provides an approximate solution. Here, we first fit a Volterra model to the loudspeaker data derived from its input/output measurements and compute the kernels $\{H_k, k = 1,2,3..\}$. Now we set the linear filter H'_1 equal to the linear kernel. Referring to figure 5, we compute the output of the filter $y(n)$ and find a new Volterra model with kernels $\{H'_k, k = 1,2,3..\}$ with input $y(n)$ and the desired output $d(n)$.

The second method uses the gradient descent algorithm to simultaneously compute parameters of the linear filter and the Volterra model. For a third order Volterra model, the r-th parameter of the i-th kernel at the j-th iteration can be written as:

$$h_i^{(j+1)}(r) = h_i^{(j)}(r) + \mu \frac{\partial z(n)}{\partial h_i(r)} e(n) \tag{7a}$$

Here, $\frac{\partial z(n)}{\partial h'_i(r)}$ is the partial derivative of the model output $z(n)$ relative to the r-th parameter of the i-th kernel and $e(n)$ is the modeling error given as:

$$e(n) = d(n) - z(n) \tag{7b}$$

The partial derivatives are given as:

$$\frac{\partial z(n)}{\partial h'_1(r)} = x(n-r) + 2\sum_{m_1=0}^{M_2}\sum_{m_2=0}^{M_2} h_2(m_1,m_2)x(n-m_1-r)y(n-m_2) \quad (8a)$$

$$+\sum_{m_1=0}^{M_2}\sum_{m_2=0}^{M_3}\sum_{m_3=0}^{M_3} h_3(m_1,m_2,m_3)[x(n-m_1-r)y(n-m_2)y(n-m_3)+$$

$$x(n-m_2-r)y(n-m_1)y(n-m_3)+x(n-m_3-r)y(n-m_1)y(n-m_2)]$$

Similarly:

$$\frac{\partial z(n)}{\partial h_2(r_1,r_2)} = y(n-r_1)y(n-r_2) \tag{8b}$$

and:

$$\frac{\partial z(n)}{\partial h_3(r_1,r_2,r_3)} = y(n-r_1)y(n-r_2)y(n-r_3) \tag{8c}$$

5. THE EXACT INVERSE PROPERTY

The conventional p-th order inverse of a p-th order Volterra model is shown in figure 6.

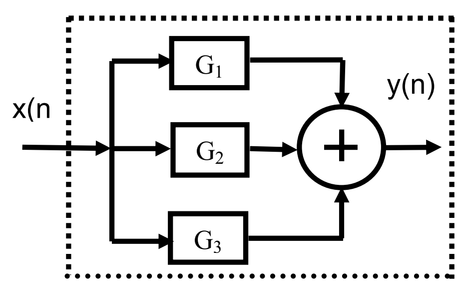

Figure 6: 3rd Order Inverse of a 3rd order Volterra Model

Here the kernels G_1, G_2 and G_3 are given as [4]:

$$G_1 = H_1^{-1} \tag{9a}$$

(9b)

$$G_{k+1} = -G_k H_2 H_1^{-1} \qquad k=1,2 \qquad (9b)$$

There are two major problems with the conventional inverse. First, the inverse of a p-th order Volterra system is in theory an infinite order Volterra system given by equation (9). This means that high order inverses must be used to find a good approximation to the true inverse. This is computationally very intensive because higher order inverses require millions of computations per sample. The second problem is that the p-th order inverse *tries* to cancel out up to the p-th order nonlinear distortions at the expense of introducing higher order distortions. This is why the quality of the sound after compensation might be worse than the uncompensated loudspeaker as discussed in [6]. The technique described in [6] gives the exact inverse of a nonlinear Volterra system by minimizing the instantaneous error between the input and the compensated loudspeaker output. The exact inverse of a p-th order Volterra model turns out to be a p-th order time-varying nonlinear filter. The predistorted signal is the root of a p-th order polynomial with time-varying coefficients. Computationally, the exact inverse is much faster than the conventional p-th order Volterra inverse as described in [6]. Furthermore, the exact inverse of the Volterra-Wiener model shown in figure 7 is a cascade of the exact inverse of the modified Volterra model shown as H^{-1} followed by the inverse of the linear system H'^{-1}_1 . The modified Volterra model with $H_1 = 1$ lends itself to having an exact nonlinear inverse.

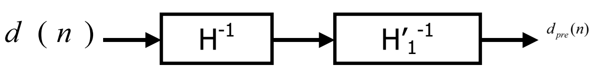

Figure 7: 3rd Exact Inverse of the Volterra-Wiener Model

Precompensation based on this model removes both the linear and the nonlinear distortions and results in drastic improvement in the sound quality.

6. RESULTS

Input/output noise measurements (0 to $20kHz$) from a small loudspeaker in a Nokia cell phone were used to find the parameters of the Volterra-Wiener model according to the methods described in section 4. The sampling frequency of the measurements was 65536 Hz to capture up to the third harmonics of the loudspeaker distortions. The linear part consisted of a 150-tap FIR filter and the nonlinear part was a 3rd order modified Volterra model with quadratic memory of 40 samples and cubic memory of 20 samples. The gradient algorithm in section 4 typically converges within a few hundred iterations. The model parameters were then used to find the exact inverse of the modified Volterra-Wiener model as described in [6]. Male and female speech materials as well as various music samples such as pop music and orchestra were used in the evaluation. On the average, 0.8 PESQ [7] improvement was measured for speech samples. Listening tests showed significant improvement in the sound quality at high playback levels. Figure 8 shows the efficacy of the new model and the exact inverse in completely removing both the linear and the nonlinear distortions when two tones are input into the loudspeaker. As seen in the middle plot, the loudspeaker produces harmonic and intermodulation distortions in response to the two input tones. These distortions are completely removed after precompensation.

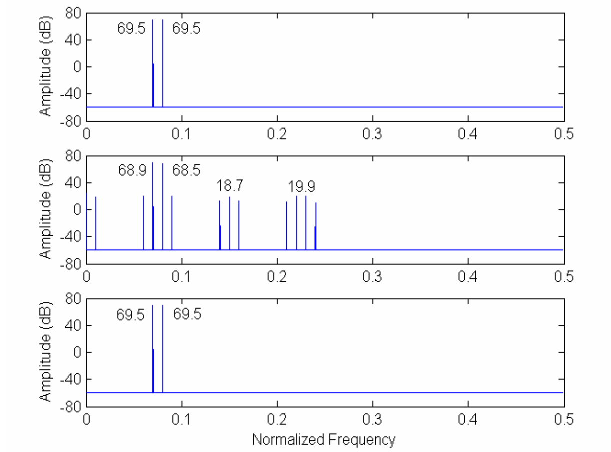

Figure 8 Spectra for input (tones with normalized frequencies of 0.07 and 0.08), uncompensated and compensated outputs of the new model

7. CONCLUSION

In this paper we discussed the limitations of loudspeaker compensation based on the Volterra model and presented a new model based on combining the Volterra and Wiener types of nonlinearities. The new model provides a better match to the loudspeaker response and lends itself to having an exact nonlinear inverse. Exact precompensation based on this model results in significant improvement in the quality of sound from small loudspeakers.

8. REFERENCES

[1] X. Y. Gao, W.M. Snelgrove, "Adaptive Linearization of a Loudspeaker," *Proc. IEEE Intl. Conf. Acoust., Speech, Signal Processing, pp. 3589-3592, 1991.*

[2] W. Frank, R. Reger, and U. Appel, "Loudspeaker nonlinearities-Analysis and compensation", Conf. Record 26th, Asilomar Conference on Signals, Systems and Computers, Pacific Grove, CA 756-760, Oct. 1992.

[3] R. de Vries, A.P., Berkhoff, C.H. Slump and O.E. Herrmann, "Digital Compensation of Nonlinear Distortion in Loudspeakers," ICASSP 92, I-165-I-168

[4] M. Schetzen, "Theory of pth-order Inverses of Nonlinear Systems," IEEE Trans. On Circuits and Systems, CAS-23, No. 5, May 1976, pp. 285-291

[5] V. J. Matthews, "Adaptive Polynomial Filters," IEEE *SP* magazine, vol. 8, no. 3, pp. 10-26, July 1991.

[6] K. Lashkari "High quality sound from small loudspeakers using the exact inverse, " Conf. Record 38th, Asilomar Conference on Signals, Systems and Computers, Pacific Grove, CA 430-434 November 2004

[7] Perceptual Evaluation of speech quality (PESQ), an objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs. ITU-T Recommendation P.862, February 2001.