ADAPTIVE VOLTERRA FILTERS FOR NONLINEAR ACOUSTIC ECHO CANCELLATION

A. Stenger and R. Rabenstein

University of Erlangen-Nürnberg, Telecommunications Laboratory
Cauerstr. 7, D-91058 Erlangen, Germany
stenger@nt.e-technik.uni-erlangen.de

ABSTRACT

More than 20 years ago it was shown that adaptive Volterra filters can be used for telephone echo cancellation, but the inherent need for a large number of coefficients has inhibited their practical use. Recently, first attempts have been made to combat nonlinearities in the acoustic echo path of hands-free telephones, which are caused by low-cost audio components. As these attempts are based on a memoryless nonlinear model, they are constrained to saturation effects, which can not describe the nonlinear behavior of small loudspeakers accurately enough. In this contribution, we report real time measurements with an adaptive Volterra filter. At first, we verify that loudspeaker nonlinearities encountered in the echo path of hands-free telephones have to be modelled with memory. Based on these results, we propose an adaptive Volterra filter structure with a reduced number of coefficients. The new structure is experimentally compared to echo cancellers with a memoryless nonlinear model using a real acoustic echo cancelling setup with a small loudspeaker. With the new technique, an echo reduction improvement of 7 dB over conventional linear adaptive filters is achieved. Furthermore it is shown that the proposed adaptive Volterra filter structure outperforms the memoryless approaches.

1. INTRODUCTION

The typical setup for acoustic echo cancellation is shown in Fig. 1. The microphone signal $y[k]$ contains the linear echo $d[k] = x[k] * h[k]$ and additional signal components $n[k]$. Given a linear echo path, the echo can be cancelled by a linear adaptive filter $\hat{h}[k]$, and only the local signal $n[k]$ is transmitted. In low-cost consumer products the loudspeaker signal may contain a certain level of nonlinear distortions, e.g. 20 dB below the signal level, without annoying the local telephone subscriber. This is because the nonlinear signal components are masked by the linear ones, and small distortions in speech signals are commonly accepted by telephone customers.

Now consider the case where $n[k]$ also contains nonlinear echo components, and the linear adaptive filter $\hat{h}[k]$

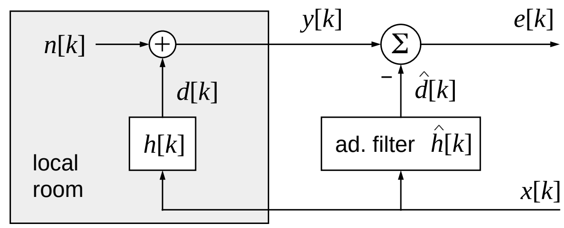

Figure 1: linear adaptive filter for acoustic echo cancellation

(Fig. 1) has converged. As it can not cancel the nonlinear echo components, the far end subscriber will hear a highly distorted echo. Consequently the required echo reduction of 30 - 45 dB [1] can not be reached in most hands-free telephone configurations. Therefore, additional adaptation measures have to cancel these echo components. These measures depend on the kind of nonlinearities encountered in the transmission chain shown in Fig. 2.

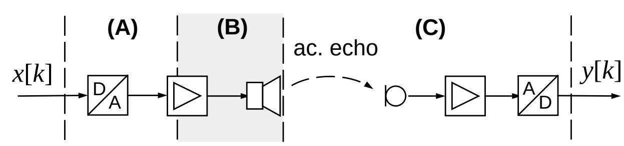

Figure 2: nonlinear echo path

The main source of nonlinearities is found in part (B), since the loudspeaker and the power amplifier are operated at the highest signal level of the transmission chain. This part of the system is assumed to be weakly time-variant, e.g. due to temperature drift. The acoustic echo path (C) is known to be linear and time-variant, while the microphone and the amplifier (C) can be modeled as LTI systems because of their low signal amplitudes. Also the nonlinear quantization of the A/D and D/A converters can be neglected in this context.

If nonlinear distortions are mainly caused by an overdriven amplifier, they are approximately memoryless and can be modeled by a saturation curve [2, 3]. Two existing approaches, which are specialized on this type of nonlinearity, are discussed in Sec. 2.

Another kind of nonlinearity is caused by the loudspeaker [4], especially when it is operated at its power limit. Due to the long time constants of the electro-mechanical system, the memory of this nonlinear behaviour cannot be neglected, as our measurements in Sec. 3 confirm. To combat this type of nonlinearity, adaptive systems with memory are required. A time-delay neural network, being such a system, is proposed in [5]. With a cascade of a time-delay neural network and an adaptive FIR filter, considerable improvement of nonlinear echo reduction is achieved. A disadvantage is the need for a second reference microphone to provide an error signal for the adaptive neural network. In [6] adaptive Volterra filters have been proposed for line echo cancelling. However, due to their high numerical complexity they have not been used in practical systems yet. In this paper, we develop an acoustic echo canceller with a second order adaptive Volterra filter and propose a method that keeps the computational complexity modest.

Before we propose a new echo canceller and point out its computational complexity in Sec. 3, Volterra filters are reviewed and some aspects for their application to acoustic echo cancellation are discussed. All mentioned echo cancellers are compared in Sec. 4 using a real echo path with a small loudspeaker operated at its power limit.

2. MEMORYLESS NONLINEAR MODEL

Acoustic echo cancellers consisting of a cascade of linear systems and memoryless nonlinear systems have already been proposed. In [3], parts (A) and (C) of Fig. 2 are modelled with adaptive FIR filters with N_A and N_C coefficients, respectively, and part (B) is realized by a saturation curve with one adaptive parameter. The adaptation of part (A) costs $O(N_A \cdot N_C)$ multiplications, which for typical values of $N_A = 20$ is computationally very demanding. Convergence problems due to local minima of the error surface can be circumvented by a special inititalization procedure, if the impulse response of part (A) has a dominant peak. For this case as much as 8 dB ERLE improvement over a linear adaptive filter are reported.

A system with non-adaptive nonlinearity [2] models part (A) in Fig. 1 as a delay, part (B) by a 7-th order polynomial, and part (C) as a classical NLMS adaptive filter. With neglectable additional effort (14 multiplications per sample) an ERLE improvement of 4 dB is obtained, without affecting convergence properties of the adaptive filter.

Experiments in Sec. 4.1 show, that both systems obtain their good results only if the major cause of nonlinearities is a clipping amplifier. In many non-portable applications, like PC telephones or videophones, the power amplifier is not necessarily overdriven, but it is still desirabe to operate a small, cheap speaker at its power limit. With such an echo path the systems [3] and [2] do not achieve remarkable ERLE improvements (see Sec. 4.2). This shows the need to develop another kind of nonlinear echo canceller which is appropriate for systems with loudspeaker nonlinearities. We will do this in the next section.

3. ADAPTIVE VOLTERRA FILTERS FOR ACOUSTIC ECHO CANCELLING

The huge number of coefficients of Volterra filters with long memory inhibits their practical use for acoustic echo cancellation, unless special precautions are taken. This section discusses some of the pitfalls to watch out for in the practical application of adaptive Volterra filters to nonlinear acoustic echo cancellation. After reviewing adaptive Volterra Filters in Sec. 3.1, we discuss two approaches for adaptation stepsize normalization. Then in Sec. 3.3 the required memory length is examined, resulting in a considerable complexity reduction. Finally in Sec. 3.4 an echo cancelling structure with a second order Volterra filter and modest computational complexity is proposed.

3.1. Basics

An N-th order discrete Volterra filter with input $x[k]$, output $y[k]$ and memory length M can be described as

$$\begin{aligned} y[k] &= h_0 + \sum_{r=1}^{N} \sum_{\kappa_1=0}^{M} \cdots \sum_{\kappa_r=\kappa_{r-1}}^{M} h_r[\kappa_1, \cdots, \kappa_r] \cdot \\ &\quad \cdot x[k-\kappa_1] \cdots x[k-\kappa_r]\,, \end{aligned} \tag{1}$$

where h_r are the r-th order Volterra kernels [9]. Such a filter can be adapted by the same algorithms as linear filters [6]. Because of its good tracking behaviour in an acoustic echo cancelling context [10] we employ the NLMS algorithm for our experiments. [8] shows that Volterra kernels are symmetric, which is exploited in (1) by considering only coefficients with non-decreasing indices κ_r, i.e. $\kappa_r \geq \kappa_{r-1}$. With the vectors

$$\mathbf{x}_1[k] = (x[k], x[k-1], \cdots, x[k-M+1])$$

$$\hat{\mathbf{h}}_1 = \left(\hat{h}_1[0], \hat{h}_1[1], \cdots, \hat{h}_1[M-1]\right)$$

for the first order Volterra kernel, and

$$\begin{aligned} \mathbf{x}_2[k] &= \big(x^2[k], x[k]x[k-1], \cdots, x[k]x[k-M+1], \\ &\quad x[k-1]x[k-1], \cdots, \\ &\quad x[k-M+1]x[k-M+1]\big) \end{aligned}$$

$$\begin{aligned} \hat{\mathbf{h}}_2 &= \big(\hat{h}_2[0,0], \hat{h}_2[0,1], \cdots, \hat{h}_2[0,M-1], \\ &\quad \hat{h}_2[1,1], \cdots, \hat{h}_2[M-1,M-1]\big) \end{aligned}$$

for the second order Volterra kernel, a second order LMS adaptive Volterra filter can be formulated as

$$e[k] = y[k] - \hat{h}_0[k] - \hat{\mathbf{h}}_1[k]{\mathbf{x}_1}^T[k] - \hat{\mathbf{h}}_2[k]{\mathbf{x}_2}^T[k] \tag{2}$$

$$\hat{h}_0[k+1] = \hat{h}_0[k] + \mu_0\, e[k] \quad (3)$$

$$\hat{\mathbf{h}}_1[k+1] = \hat{\mathbf{h}}_1[k] + \mu_1\, e[k]\mathbf{x}_1[k] \quad (4)$$

$$\hat{\mathbf{h}}_2[k+1] = \hat{\mathbf{h}}_2[k] + \mu_2\, e[k]\mathbf{x}_2[k] \quad (5)$$

3.2. Normalization of adaptation stepsize

Consider a linear LMS adaptive FIR filter, consisting only of $\mathbf{h}_1$ and $\mathbf{x}_1$. It has been proven that this algorithm converges for $\mu_1 < \frac{2}{||\mathbf{x}_1||_2^2}$ [7]. Therefore it is straightforward to normalize the LMS adaptive Volterra filter using $\mu_1 = \mu_2 < \frac{\alpha}{||\mathbf{x}_1||_2^2 + ||\mathbf{x}_2||_2^2}$, where convergence can be proven for $0 < \alpha < 2$ in a similay way.

A disadvantage of this kind of normalization becomes obvious, if e.g. for a second order Volterra filter $||\mathbf{x}_1||_2^2 \gg ||\mathbf{x}_2||_2^2$. Then the coefficients $\mathbf{h}_2[k]$ are updated in very small steps which severely slows down the second order kernel convergence. However, in [9] the LMS adaptive Volterra filter is formulated with different stepsize for each kernel, which motivates us to suggest a separate normalization for first and second order kernel:

$$\mu_1 = \frac{\alpha_1}{||\mathbf{x}_1||_2^2} \quad (6)$$

$$\mu_2 = \frac{\alpha_2}{||\mathbf{x}_2||_2^2} \quad (7)$$

Experiments showed safe convergence with this normalization only for second order systems, which can be explained by the orthogonality of first and second order excitations.

3.3. Memory length of higher order kernels

The memory length of the Volterra kernels is determined by the whole transmission chain in Fig. 2, which it is typically several hundred taps for acoustic echo cancellers. Due to the complexity of $O(M^N)$, the memory length of the higher order Volterra kernels must be much less in a practical application. For analog audio systems the envelope of the higher order kernels is determined by the envelope of the impulse response [11], which typically has a dead zone followed by a peak, and then exponentially decays. We verified this assumption for an acoustic echo path with a highly excited small loudspeaker in a room with short reverbaration time. Fig. 3 shows the first 50 taps of the first order kernel, i.e. the impulse response, and Fig. 4 shows the second order kernel, both obtained by an NLMS adaptive second order Volterra filter with $M = 50$. The zero elements in the upper half of Fig. 4 have not been used for symmetry reasons. As its coefficients have relatively smaller values than the linear impulse response, the second order kernel may be truncated to shorter memory length than the linear kernel causing the same error power in the output signal.

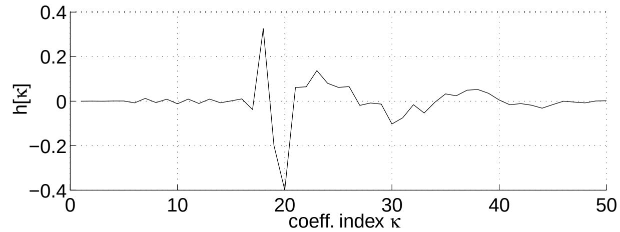

Figure 3: Linear FIR system with memory length 50

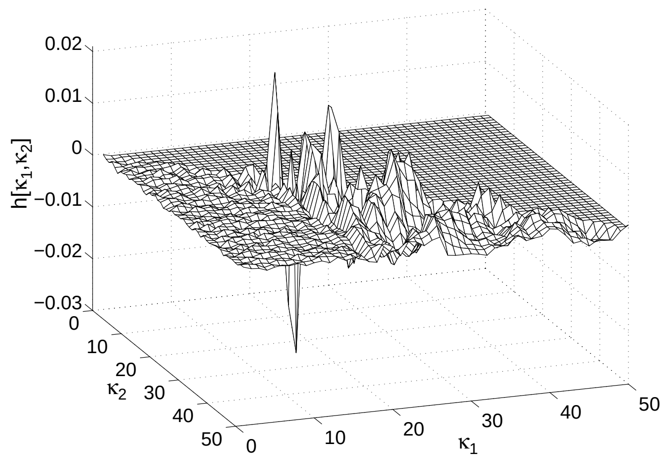

Figure 4: Second order Volterra kernel adapted from a small loudspeaker

3.4. Proposed structure

As experiments with higher orders Volterra filters without orthogonalization yielded only marginal improvements in terms of echo reduction, we propose a second order Volterra filter. Fig. 4 suggests to truncate the first elements of the second order Volterra kernel until near the first peak in both dimensions, i.e. $\kappa_1 \leq \Delta, \kappa_2 \leq \Delta$. In the example of Fig. 3 and Fig. 4 a good choice would be $\Delta = 17$. This leads to a modified second order Volterra representation with different memory length M_1 and M_2:

$$y[k] = h_0 + \sum_{\kappa=0}^{M_1-1} h_1[\kappa]x[k-\kappa] + \sum_{\kappa_1=\Delta}^{M_2+\Delta-1} \sum_{\kappa_2=\kappa_1}^{M_2+\Delta-1} h_2[\kappa_1,\kappa_2]x[k-\kappa_1]x[k-\kappa_2]\,. \quad (8)$$

An adaptive Volterra filter of the above kind is shown in Fig. 5. The NLMS coefficient update for the linear filter is performed according to (4) using the normalization (6). The DC coefficient adaptation is given in (3), and the second order kernel is adapted using (5), where the vector $\hat{\mathbf{h}}_2^{(tr)}[k]$ is the truncated second order Volterra kernel.

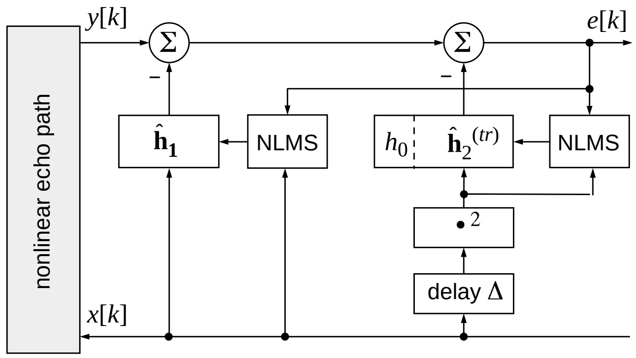

Figure 5: New nonlinear acoustic echo canceller

Its elements are equivalent to those of $\hat{\mathbf{h}}_2[k]$ with indices $\Delta \leq \kappa_1, \kappa_2 < \Delta + M_2$. As the DC component of the input signal is correlated with the second order excitation, normalization for these two components is performed in common, i.e.

$$\mu_0 = \mu_2 = \frac{\alpha_2}{||\mathbf{x}_2[k]||_2^2 + 1} \tag{9}$$

where the 1 in the denominator accounts for the DC component and can be omitted, if no DC coefficient is implemented.

4. EXPERIMENTAL RESULTS

The new system has been tested with different low cost speakers between 0.1 and 0.4 Watts and a one-chip amplifier placed in an enclosure with low reverberation. As excitation $x[k]$ white Gaussian noise was used. The amplitude was chosen so that the amplifier is not highly overdriven and loudspeaker nonlinearities dominate over saturation effects. We compare different nonlinear echo cancelling approaches with a linear adaptive filter in terms of the Echo Return Loss Enhancement

$$\text{ERLE} = \frac{\mathcal{E}\{y^2[k]\}}{\mathcal{E}\{e^2[k]\}}\,.$$

The steady-state ERLE of the linear filter is a coarse measure of the nonlinear distortion ratio within the loudspeaker signal, which is percieved by the near-end user in the local room.

4.1. Systems with memoryless nonlinearity

The behaviour of the systems described in Sec. 2 is compared for two situations: in Fig. 6 the nonlinearity in the echo path contains mainly saturation effects due to a lowered amplifier supply voltage. As the linear adaptive filter (1) can reduce the echo only by 15 dB, the nonlinear distortions contained in the loudspeaker signal are at about -15 dB, i.e. they are well audible, but might be just tolerable for low cost applications. For this type of nonlinearity both schemes allow a further echo reduction by 4-5 dB, see curves (2) and (3).

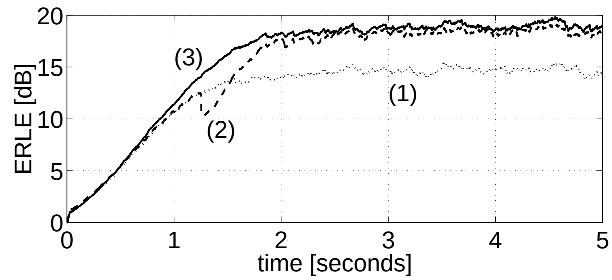

Figure 6: Saturation effect: Memoryless nonlinearity is appropriate model

In Fig. 7 the amplifier was not overdriven, and therefore loudspeaker nonlinearities dominate. In this case the memoryless nonlinear echo cancellers (2) and (3) do not improve the echo reduction compared to a linear adaptive filter (1).

4.2. Adaptive Volterra Filter

Fig. 7 shows a comparision between a conventional linear acoustic echo canceller with NLMS algorithm (1) with $M_1 = 250$ coefficients, both memoryless nonlinear echo cancellers (2) and (3), and the new system (4) with $\Delta = 17$, $M_2 = 25$. The additional computational load of the second order Volterra kernel is $\frac{3}{2}{M_2}^2 + \frac{3}{2}M_2 + 2$, if symmetry is exploited, while the linear NLMS adaptive filter costs $2M_1+2$ multiplications per sample. In our example we have 977 additional multiplications/sample due to the 2nd order kernel, while the linear filter costs 502 multiplications/sample.

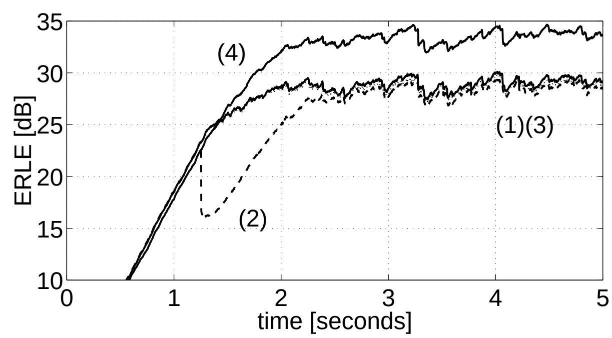

Figure 7: Loudspeaker nonlinearity: 2nd order Volterra filter is suitable

Comparing curves (1) and (4) shows, that convergence speed is not lowered through the additional second order kernel, and steady state ERLE is 5 dB higher. With larger memory $M_2 = 35$ up to 7 dB can be gained.

4.3. Effect of a DC offset in the microphone signal

Often $y[k]$ or $x[k]$ have a small DC offset due to A/D conversion. Here we discuss the need for an adaptive DC coefficient h_0, and claim that for a fair comparision between linear and nonlinear adaptive filters the DC components of $x[k]$ and $y[k]$ have to be removed.

The following experiment has been performed with the same setup as above, but a small DC offset, which originally occured in $y[k]$, has not been removed. Fig. 8 shows that the performance of the linear filter is degraded through the DC offset, and only 25 dB ERLE can be reached. However, if an adaptive DC coefficient is addded to the system, i.e. we adapt h_0 and $\mathbf{h}_1$ in parallel, this problem can be circumvented, and the expected 28 dB ERLE are reached, as with the linear filter in Fig. 7. The upper two curves in Fig. 8 result with a 2nd order Volterra filter. They show that the use of an adaptive DC coefficient h_0 does not significantly improve the model acuracy. From this we conclude, that a second order Volterra filter can model a DC component to a certain extent, and therefore an additional adaptive DC coefficient h_0 is not required in the nonlinear echo path model.

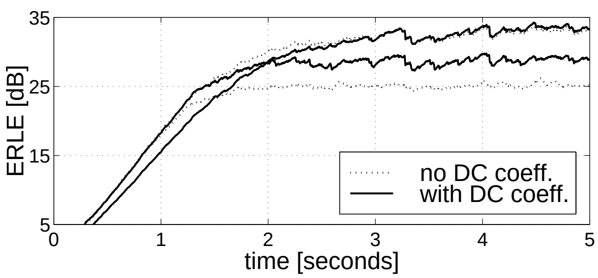

Figure 8: Apative echo cancellation with and without DC adaptive coefficient

Further we can see, that it is important to exclude the effects of a possible DC offset in $x[k]$ and $y[k]$, if a second order adaptive Volterra filter is compared to a linear adaptive filter in the context of second order loudspeaker nonlinearities. Otherwise we would measure a higher ERLE difference than could be gained by modelling second order nonlinearities of a loudspeaker. To ensure a fair comparision in the experiments of Sec. 4.1 and 4.2, we therefore removed any DC offset in the signals $x[k]$ and $y[k]$.

5. SUMMARY

The performance of linear acoustic echo cancellers is limited by nonlinear components in the echo path. Existing acoustic echo cancellers for memoryless nonlinearities are specialized for saturation effects and cannot model loudspeaker nonlinearities accurately enough. We showed that a second order adaptive Volterra filter is suitable for echo cancellation in such cases, and a DC coefficient needs not to be incorporated into the model, even if there is a small DC offset in the microphone signal. Two normalization schemes for the LMS algorithm have been discussed, and a design rule for the choice of the relevant coefficients was proposed to reduce the computational load. Practical implementations of systems with different small loudspeakers operated at their power limit have been investigated. For situations, where memoryless nonlinear models could not improve echo reduction, the new system achieves an echo reduction gain up to 7 dB over linear adaptive filters. With only modest computational load, moer than 5 dB are gained.

6. REFERENCES

[1] International Telecommunication Union, “General characteristics of international telephones connections and international telephone circuits – acoustic echo controllers,” *ITU-T Recommendation G.167*, 1993

[2] A. Stenger, R. Rabenstein, “An Acoustic Echo Canceller With Compensation of Nonlinearities,” *Proc. EUSIPCO 98*, Isle of Rhodes, Greece, September 1998, pp. 969-972

[3] Bryan S. Nollet, Douglas L. Jones “Nonlinear Echo Cancellation For Hands-Free Speakerphones,” *Proc. NSIP'97*, September 8-10, 1997, Michigan USA

[4] Franklin X.Y. Gao, W. Martin Snelgrove, “Adaptive Linearization of a Loudspeaker,” *Proc. ICASSP 91*, pp. 3589-3592.

[5] A. N. Birkett, R. A. Goubran, “Acoustic Echo Cancellation Using NLMS-Neural Network Structures”, *Proc. ICASSP 1995, Detroit, MI*, pp. 2035-3038

[6] E.J. Thomas, “Some considerations on the application of the Volterra representation of nonlinear networks to adaptive echo canceller,” *Bell Systems Technical Journal*, Vol 50, No. 8, October 1971, pp 2797-2905.

[7] Samuel D. Widrow, Bernard Stearns: “Adaptive Signal Processing,” Prentice Hall, Englewood Cliffs, New Jersey, 1985

[8] Martin Schetzen, “The Volterra and Wiener Theories of Nonlinear Systems,” *Wiley and Sons*, New York, 1980

[9] V. John Mathews, “Adaptive Polynomial Filters,” *IEEE Signal Processing Magazine*, 8(3), pp.10-26, July 1991

[10] E. Hänsler, “From Algorithms to Systems – It's a Rocky Road,” *Proc. IWAENC'97*, London, Sept. 11-12, 1997, pp. K.1 - K.4

[11] Martin J. Reed, Malcom O. Hawksford, “Practical Modelling of Nonlinear Audio Systems Using the Volterra Series,” *100th covention of AES*, Copenhagen, May 11-14, 1996, Preprint 4264 (R-2)