ONLINE ESTIMATION OF THE OPTIMUM QUADRATIC KERNEL SIZE OF SECOND-ORDER VOLTERRA FILTERS USING A CONVEX COMBINATION SCHEME

Marcus Zeller[1], Luis A. Azpicueta-Ruiz[2] and Walter Kellermann[1]

[1] Multimedia Communications
and Signal Processing*
University of Erlangen-Nuremberg
Cauerstr. 7, 91058 Erlangen, Germany
{zeller,wk}@LNT.de

[2] Department of Signal Theory
and Communications†
Universidad Carlos III de Madrid
28911 Leganés-Madrid, Spain
lazpicueta@tsc.uc3m.es

ABSTRACT

This paper presents a method for estimating the optimum memory size for identification of an unknown second-order Volterra kernel. As these structures may imply considerable computational demands, it is highly desirable to design adaptive realizations with a minimum number of coefficients. Therefore, we propose a combination scheme comprising two Volterra filters with time-variant sizes of the actually used quadratic kernels. By following some simple rules, the number of diagonals in the quadratic kernels is increased or decreased in order to find the optimum memory configuration in parallel to the coefficient adaptation. Thus, the arbitrary choice of the nonlinear system size is overcome by a dynamically growing/shrinking system. Experimental results for various signals and nonlinear scenarios demonstrate the effectiveness of the proposed method.

Index Terms— Adaptive Volterra Filter, Nonlinear System Identification, Convex Combination, Structure Selection

1. INTRODUCTION

It is widely acknowledged that in numerous of today's telecommunication scenarios, the assumption of purely linear systems is not fulfilled. Therefore, nonlinear models play an increasing role in system identification and compensation of nonlinear distortions. Volterra filters are very popular since they represent a rather general class of nonlinear filters with memory. On the other hand, the computational complexity of these structures can be quite demanding and hence it is desirable to design adaptive filters with a minimum number of coefficients while maintaining the optimum performance. However, this is a challenging task due to the lack of a-priori knowledge about the dimensions of the unknown nonlinear system.

In this contribution, we propose a new method which can determine the necessary number of diagonals in the second-order kernel by means of an adaptive convex combination of two different nonlinear filters. The methodology of filter combinations has been successfully applied to various linear adaptive filtering problems [1, 2, 3] and is based on a relatively simple concept. In our context, we use the combination in order to determine the memory requirements of the second-order Volterra kernel online, instead of setting this to

a fixed value. This is a decisive difference to the commonly used approach of genetic algorithms, as the memory configuration of only one (combined) adaptive system is performed in parallel to the actual adaptation of the filter coefficients.

The rest of the paper is organized as follows: In Sec. 2, we briefly summarize the second-order Volterra filtering technique in the diagonal structure which will be exploited for scaling the kernel size. Sec. 3 presents the idea of combining adaptive filters before the actual estimation algorithm is described in Sec. 4. Finally, Sec. 5 illustrates several experimental results for different real-world Volterra systems as well as for noise and speech signals. The main results are then summarized in Sec. 6.

2. SECOND-ORDER VOLTERRA FILTERING IN DIAGONAL COORDINATES

The output of a second-order Volterra filter is given by

$$y(k) = y_1(k) + y_2(k) \tag{1}$$

which represents the superposition of the kernel outputs denoted by $y_1(k)$ and $y_2(k)$. The output of the linear kernel

$$y_1(k) = \sum_{n=0}^{N_1-1} h_{1,n} \cdot x(k-n) := h_{1,k} * x(k) \tag{2}$$

is given by a convolution of the impulse response $h_{1,k}$ with the input signal $x(k)$, whereas the nonlinear processing in the second-order or *quadratic* kernel reads

$$y_2(k) = \sum_{n_1=0}^{N_2-1} \sum_{n_2=n_1}^{N_2-1} h_{2,n_1,n_2} \cdot x(k-n_1) \cdot x(k-n_2). \tag{3}$$

Note that the redundancy due to index permutations has been taken into account here and thus the coefficients h_{2,n_1,n_2} correspond to the so-called triangular form of the 2D kernel plane [4].

Although (3) implies the full summation over all unique elements of the quadratic kernel, it has been shown that for many applications with cascaded elements (as e.g. in nonlinear echo cancellation) it is a valid assumption to find non-zero coefficients mainly around the main diagonal [5, 6]. This fact can be exploited by adopting the *diagonal coordinate representation* which yields the equivalent form of the second-order kernel

$$y_2(k) = \sum_{w=0}^{W-1} \sum_{n=0}^{N_2-w-1} \underbrace{h_{2,n,w+n}}_{:=h_{2,w,n}} \cdot \underbrace{x(k-n) \cdot x(k-w-n)}_{:=x_w(k-n)} \tag{4}$$
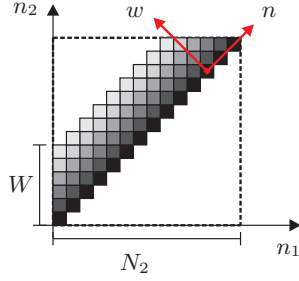
Fig. 1. Diagonal coordinate representation of a second-order Volterra kernel ($N_2 = 16$, $W = 6$)

and corresponds to (3) for $W = N_2$. However, as the summations in (4) are carried out along diagonals w, the computations may be restricted to a certain *width* $W < N_2$ in the 2D kernel plane, thus reducing the algorithmic complexity [6]. An illustration of the diagonal coordinates is given in Fig. 1. In this contribution, we propose an algorithm which yields an estimate for such a restriction, i.e. which determines the number of actually needed kernel diagonals.

3. PROPOSED COMBINATION SCHEME

In order to determine the optimum kernel width W_{opt} for an unknown system to be identified, we present a system of two Volterra filters following a combination scheme as depicted in Fig. 2. In contrast to the usual idea of combining two or more adaptive linear filters with different convergence properties [1, 2] the primary goal of this scheme is to estimate the memory width W_{opt} of the unknown quadratic kernel. As we seek to perform both the system identification task and the estimation of the number of necessary diagonals in the quadratic kernel at the same time, we employ two Volterra filters $\widehat{h}_{\mathrm{A}}(k)$ and $\widehat{h}_{\mathrm{B}}(k)$ which share the same kernel dimensions N_1, N_2 but differ in their number of diagonals, i.e. $W_{\mathrm{A}} \neq W_{\mathrm{B}}$. Therefore, the obtained modelling errors of both components will be different.

The global output $\widehat{y}(k)$ is given by the convex combination

$$\widehat{y}(k) = \eta(k) \cdot \widehat{y}_{\mathrm{A}}(k) + (1 - \eta(k)) \cdot \widehat{y}_{\mathrm{B}}(k) \qquad (5)$$

where the corresponding outputs after (1), (2) and (4) are mixed with $0 < \eta(k) < 1$ for every time instant k. Although both filters are driven with the same excitation $x(k)$ and seek to match the same reference signal

$$d(k) = y(k) + n(k), \qquad (6)$$

they are nevertheless operated independently. Therefore, the residual errors for both components c $\in \{\mathrm{A}, \mathrm{B}\}$ are calculated according to

$$e_{\mathrm{c}}(k) = d(k) - \widehat{y}_{\mathrm{c}}(k) \qquad (7)$$

and are used for the individual adaptation of the Volterra filters. The adjustment of the kernel coefficients is performed following the LMS-type updates

$$\widehat{h}_{\mathrm{c},1,n}(k+1) = \widehat{h}_{\mathrm{c},1,n}(k) + \frac{\alpha_{\mathrm{c},1}}{P_1} \cdot e_{\mathrm{c}}(k) \cdot x(k-n) \qquad (8)$$

$$\widehat{h}_{\mathrm{c},2,w,n}(k+1) = \widehat{h}_{\mathrm{c},2,w,n}(k) + \frac{\alpha_{\mathrm{c},2}}{P_2} \cdot e_{\mathrm{c}}(k) \cdot x_w(k-n) \quad (9)$$

with the definitions from (4) for the quadratic kernel. As can be seen, (8) and (9) require a separate normalization of both kernels

to their corresponding input power estimates P_1, P_2 (SNLMS) [6]. Analogously to (5), the resulting global error reads

$$e(k) = \eta(k) \cdot e_{\mathrm{A}}(k) + (1 - \eta(k)) \cdot e_{\mathrm{B}}(k). \qquad (10)$$

Note that both the calculation of the inividual errors (7) as well as the adaptation of the component filters $\widehat{h}_{\mathrm{A}}(k), \widehat{h}_{\mathrm{B}}(k)$ are not depicted in Fig. 2 for the sake of readability.

As can be seen from (5) and (10), the mixing parameter $\eta(k)$ itself is also time-variant in order to obtain an effective and flexible combination. Instead of changing $\eta(k)$ directly, the mixing is moreover designed via a sigmoid activation function [2]

$$\eta(k) := \frac{1}{1 + e^{-a(k)}} \qquad (11)$$

which ensures the boundedness of $\eta(k)$ to the unit interval and reduces the gradient noise near the limit values [1]. At each time instant, the control parameter $a(k)$ is updated in order to minimize the squared global error (10), i.e., it is adjusted in direction of the negative gradient $-\frac{\partial e^2(k)}{\partial a(k)}$. This can be understood as a second adaptation layer whose input is given by the difference $\Delta e(k) := e_{\mathrm{B}}(k) - e_{\mathrm{A}}(k)$ of the component errors. Recently, it has been shown that it is advantageous to implement the corresponding updates of $a(k)$ in the form of [3]

$$a(k+1) = a(k) + \mu_a \cdot \frac{\eta(k) \cdot (1 - \eta(k))}{\widetilde{\Delta e^2}(k)} \cdot e(k) \cdot \Delta e(k), \quad (12)$$

which thus denotes the corresponding update equation for the *normalized combination control (NCC)* block in Fig. 2. In (12), a smoothed version of the error difference energy

$$\widetilde{\Delta e^2}(k) := \beta \cdot \widetilde{\Delta e^2}(k-1) + (1 - \beta) \cdot \Delta e^2(k) \qquad (13)$$

is used and μ_a is a step size which can be used to essentially adjust the rate of change for the mixing parameter. In order not to stall the mixing adaptation for extreme values of $\eta(k)$, the control $a(k)$ itself is bounded to the interval $[-4, +4]$ (see [1]).

4. KERNEL WIDTH ESTIMATION ALGORITHM

The effect of the combination mechanism and its improvement on the overall performance is well understood and has already been investigated thoroughly in [1, 2, 3]. Specifically, it has been shown that
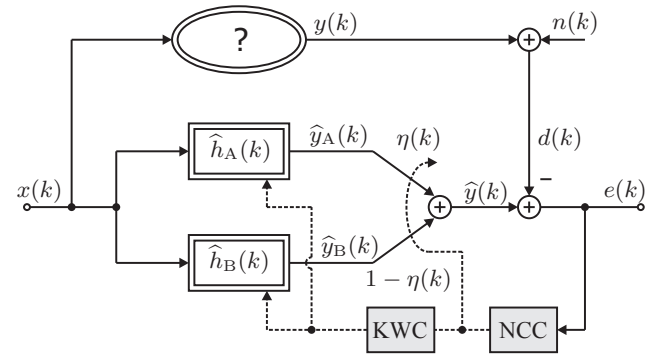

Fig. 2. Convex combination scheme (NCC) and control algorithm (KWC) for the estimation of the optimum width of the unknown quadratic Volterra kernel

the value of $\eta(k)$ may also be interpreted as an indicator towards the (temporarily) superior one of the two competing filters which can be directly related to the mean-squared errors of the individual components [2]. Because the mixing parameter is governed by the smooth curve given by (11), this scheme provides a *soft decision* on the adaptive filter performance which will be exploited in order to determine the optimum width W_{opt} for modelling the unknown second-order kernel in the plant.

First of all, as the widths of both component filters are realized adaptively, the W in (4) are replaced by their time-variant versions $W_c(k)$ for $c = \{A, B\}$. Note that this implies potential reconfigurations of the memory size during the convergence phase of the kernel coefficients. Hence, it is clear that modifications of the width by means of the *kernel width control (KWC)* depicted in Fig. 2 should be based on a robust indication for the performance in the Volterra filter combination. In order to obtain such a measure, we set the update parameter in (12) to a reasonably small value, e.g. $\mu_a = 0.01$, as to slow down the changes in $\eta(k)$. Nevertheless, it is still necessary to smooth the slope of the mixing further and thus

$$\widetilde{\eta}(k) := \lambda_{\text{KWC}} \cdot \widetilde{\eta}(k) + (1 - \lambda_{\text{KWC}}) \cdot \eta(k) \qquad (14)$$

denotes a suitable indicator for the modification of $W_c(k)$. In all of the conducted experiments, the forgetting factor has been chosen to $\lambda_{\text{KWC}} = 0.9999$ which proved to be a reliable parameter for various systems and signal constellations.

The estimation is initialized by employing a set of filters with $W_A(0) = 0$ (i.e. linear filter) and $W_B(0) = \Delta W$ which denotes the fixed distance in terms of kernel diagonals between A and B. Since the quadratic kernel of an unknown, significantly nonlinear system will in general exhibit a width different from $W_B(0)$, the quadratic kernels must be enabled to increase and/or decrease to this size as well. A suitable updating mechanism for the $W_c(k)$ is given by the following set of modification rules:

$$W_c(k+1) = \begin{cases} W_c(k) + 1, & \text{if } \widetilde{\eta}(k) \leq \eta_{\min} + \varepsilon_{\text{inc}} \\ W_c(k) - 1, & \text{if } \widetilde{\eta}(k) \geq \eta_{\max} - \varepsilon_{\text{dec}} \\ W_c(k), & \text{else} \end{cases} . \qquad (15)$$

Regarding (15), the number of diagonals for both filters are obviously only changed if given thresholds of $\widetilde{\eta}(k)$ are surpassed or dropped below, respectively. These thresholds are defined by the limit values $\eta_{\min} = 0.018$ and $\eta_{\max} = 0.982$ which are a direct consequence of the restriction of $a(k)$ and by the controls ε_{inc} and ε_{dec}. In order to guarantee for significant decisions, these parameters should be selected quite small (e.g. $\varepsilon_{\text{inc}} = 0.1$). Moreover, it is imperative to avoid interference of the memory configuration adaptation with the convergence of the filter coefficients. Therefore, after changes in $W_c(k)$ according to (15) have been invoked, the modification rules remain idle for a *standby time* of K samples. This ensures proper reconvergence of the filters that is mainly important for newly added diagonals in case of an increase in width.

The mechanism is further illustrated by Fig. 3 which provides an instructive example with a changing optimum width W_{opt}. As can be seen in the initial phase ①, the widths of both filters are below the nonlinear memory of the unknown system. However, as $W_B(k)$ is always larger than $W_A(k)$, $\widehat{h}_B(k)$ is capable of better modelling the true quadratic kernel. Consequently, the soft switch $\widetilde{\eta}(k)$ will clearly point to B and invoking (15) will yield an enlargement of the number of diagonals for both filters. This situation persists until A approaches the true width W_{opt} in phase ② as then the performance gain of component B is lost and the decision will remain halfway between both filters. If the unknown nonlinear system changes to a
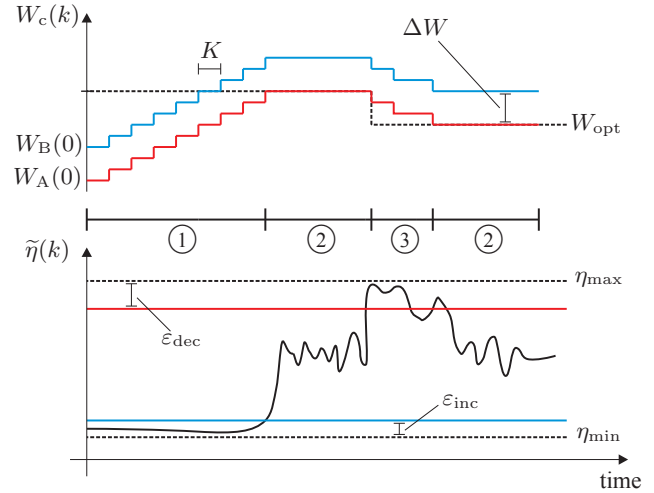

Fig. 3. Ideal tracking behaviour of $W_A(k), W_B(k)$ and evolution of the smoothed mixing parameter $\widetilde{\eta}(k)$

higher W_{opt}, the combination will track these changes and increase the filter widths to comprise a higher number of diagonals by repeating the evolution in ①. In contrast, in case of a decreasing nonlinear memory both filters exhibit an over-representation of the true Volterra system. However, as $W_A(k)$ has less kernel diagonals and thus re-adapts faster in order to match the new situation, an indication of $\widetilde{\eta}(k)$ towards A can be expected. This behaviour is illustrated in phase ③ and will stop only if $W_A(k)$ again approaches the true quadratic kernel width. We can therefore conclude that in steady-states of the mixing parameter evolution, system A matches the quadratic memory size of the nonlinear system to be identified, i.e. $W_A(k) \approx W_{\text{opt}}$ holds and can be used for the estimation of the optimum kernel width.

5. EXPERIMENTAL RESULTS

In order to demonstrate the robustness of the proposed method, we provide several experimental results in the following. All of the experiments have been performed in a scenario with a noisy reference $d(k)$ at an SNR of 30 dB. Moreover, the strength of the nonlinear distortions has been set such that the linear-to-nonlinear ratio (LNLR) of signal powers was 10 dB. The step size parameters for the adaptation have been chosen to $\alpha_{c,1} = 0.3$ and $\alpha_{c,2} = 0.1$ for both filters $c \in \{A, B\}$. For the proposed combination scheme, we have employed $\mu_a = 0.01$, $\beta = 0.9$ and a smoothing of the mixing parameter by $\lambda_{\text{KWC}} = 0.9999$ throughout all simulations. As it has been found that the distance in terms of diagonals between the filters A and B has rarely an impact on the performance of the algorithm, this has been selected to $\Delta W = 2$. Likewise, the same standby time $K = 16000$ (which corresponds to 2 seconds duration) has been applied for all experiments, as it represents a good compromise between robustness and response time.

A first experiment is illustrated in the left plot of Fig. 4 which shows the results for white Gaussian noise excitation on a second-order Volterra system with $N_1 = 320$, $N_2 = 64$ and width $W_{\text{opt}} = 16$ that has been obtained by measurements from a small low-cost loudspeaker. Different versions of the quadratic kernels have been created by truncating the width to a total of $W_{\text{opt}} = 24/16/8$ diagonals. As can be seen, the true memory width is closely matched

within in tolerance of one or two diagonals. To the right, the ERLE[1] performance of both an SNLMS [6] with $W = 16$ diagonals and the dynamically growing system with the KWC approach are displayed for the unknown nonlinear system with $W_{\mathrm{opt}} = 16$.
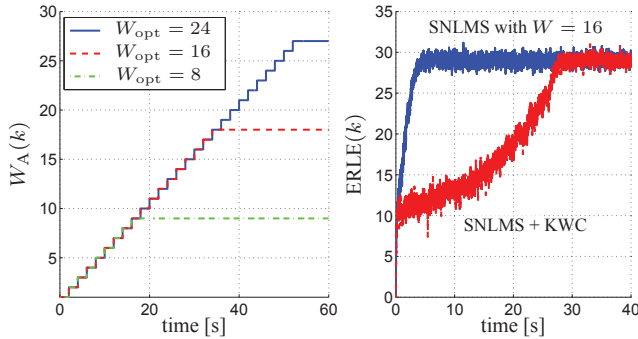

Fig. 4. Optimum width estimation for second-order Volterra kernels ($N_1 = 320$, $N_2 = 64$, loudspeaker) by white Gaussian noise

Fig. 5 shows a second experiment where a speech signal was used as input $x(k)$. Here, the unknown nonlinear system was represented by a Volterra filter with sizes $N_1 = 256$, $N_2 = 48$ that has been extracted from an up-to-date mobile phone. For comparison, the results for different choices of the modification tolerances $\varepsilon_{\mathrm{inc}}, \varepsilon_{\mathrm{dec}}$ are presented. Although an approach of $W_{\mathrm{A}}(k)$ towards the true quadratic kernel width can be seen, it should be mentioned that it is in general more difficult to estimate the nonlinear memory size from this signal which is due to the non-persistent excitation of the quadratic kernel and the highly dynamic nature of speech.
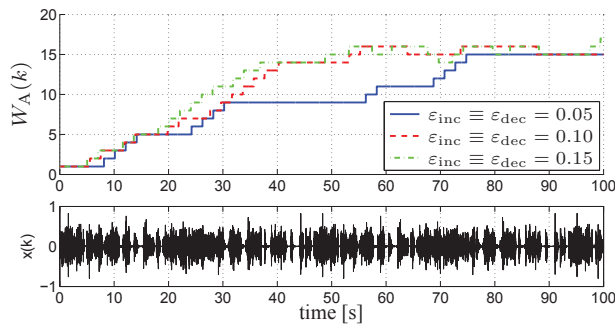

Fig. 5. Width estimation for excitation by a male speaker and various tolerance thresholds ($N_1 = 256$, $N_2 = 48$, mobile phone)

Finally, Fig.6 presents some results on the tracking behaviour of the KWC mechanism from Sec. 4. Therefore, the second-order Volterra filter from Fig. 4 is used once again where the initial number of diagonals $W_{\mathrm{opt}} = 16$ is changed twice. First, the width is reduced to $W_{\mathrm{opt}} = 8$ diagonals after 100 seconds and then raised to $W_{\mathrm{opt}} = 24$ after 200 seconds (see reference line in Fig. 6). Clearly, it can be seen that changes of the kernel width can be followed quite well by the KWC algorithm for both white Gaussian (using $\varepsilon_{\mathrm{inc}} \equiv \varepsilon_{\mathrm{dec}} = 0.05$) and speech-like coloured Laplacian noise ($\varepsilon_{\mathrm{inc}} \equiv \varepsilon_{\mathrm{dec}} = 0.15$). Moreover, the evolution of the soft switch $\widetilde{\eta}(k)$ is depicted in the lower plot for the coloured noise.
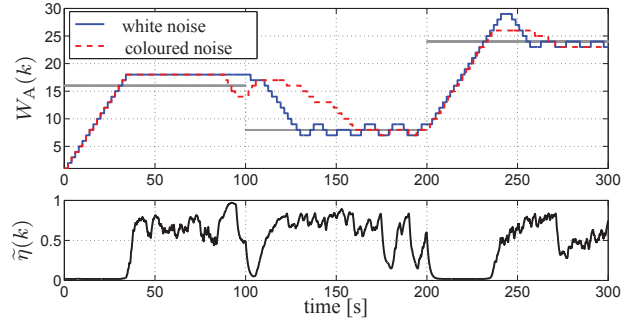
---

[1]echo return loss enhancement [5]

Fig. 6. Tracking of optimum kernel width ($W_{\mathrm{opt}} = 16/8/24$) for noise excitation and soft decision $\widetilde{\eta}(k)$ for coloured noise

6. CONCLUSIONS

A novel method for an online estimation of the optimum memory size for the identification of an unknown second-order Volterra kernel has been proposed. As the estimation of the number of involved kernel diagonals is performed in parallel to the adaptation of the filter coefficients, a robust indication on the filter performance was found by a combination of two Volterra filters with different widths of nonlinear memory. Following some basic rules, a control mechanism has been outlined and was shown to be effective for identifying systems with both fixed or time-varying quadratic kernels. Since the realized system is dynamically growing or shrinking, the critical task of defining the nonlinear system size a-priori is therefore overcome. Experimental results have confirmed the theoretically expected behaviour for both noise and speech signals and various systems. Future research will focus on modifications for enhanced estimation results as well as on extensions to higher-order and DFT-domain Volterra filtering.

7. REFERENCES

[1] J. Arenas-García, M. Martínez-Ramón, A. Navia-Vázquez, and A. R. Figueiras-Vidal, "Plant identification via adaptive combination of transversal filters," *Signal Processing*, vol. 86, no. 9, pp. 2430–2438, Sept. 2006.

[2] J. Arenas-García, A. R. Figueiras-Vidal, and A. H. Sayed, "Mean square performance of a convex combination of two adaptive filters," *IEEE Trans. on Signal Processing*, vol. 54, no. 3, pp. 1078–1090, March 2006.

[3] L. A. Azpicueta-Ruiz, A. R. Figueiras-Vidal, and J. Arenas-García, "A normalized adaptation scheme for the convex combination of two adaptive filters," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, Las Vegas, April 2008.

[4] V. J. Mathews and G. L. Sicuranza, *Polynomial Signal Processing*, New York: John Wiley and Sons, 2000.

[5] F. Küch and W. Kellermann, "Nonlinear line echo cancellation using a simplified second-order Volterra filter," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, Orlando, May 2002.

[6] M. Zeller and W. Kellermann, "Coefficient pruning for higher-order diagonals of Volterra filters representing Wiener-Hammerstein models," in *Proc. Int. Workshop on Acoustic Echo and Noise Control (IWAENC)*, Seattle, Sept. 2008.