COEFFICIENT PRUNING FOR HIGHER-ORDER DIAGONALS OF VOLTERRA FILTERS REPRESENTING WIENER-HAMMERSTEIN MODELS

Marcus Zeller and Walter Kellermann

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

ABSTRACT

Block-based nonlinear structures as, e.g., the Wiener-Hammerstein model, are popular for analyzing a broad class of nonlinear distortions. On the other hand, these models can also be related to Volterra filters in diagonal coordinates which represent very general nonlinear filters with memory. In this work, we propose an approach for estimating the significant coefficients of the nonlinear Volterra kernels during adaptation of such block-based filters by relating the linear impulse response and the diagonals of higher-order kernels. As the number of coefficients for the higher-order kernels is generally very large, this information can then be used to prune the number of necessary coefficients and thus lower the computational complexity. Experimental results for several test signals and nonlinear systems demonstrate the effectiveness of such an approach in a nonlinear acoustic echo cancellation scenario.

Index Terms— Nonlinear Acoustic Echo Cancellation, Adaptive Volterra Filters, Model Selection

1. INTRODUCTION

Acoustic echo cancellation (AEC) is a key technology for a variety of modern telecommunication systems and improves the quality of speech dialogues noticeably. While AEC is well-developed for linear echo paths, small-size and low-cost audio components suffer from nonlinear distortions which severely hamper the convergence of conventional linear echo cancellers [1].

This problem can be overcome by using nonlinear adaptive filters that adequately model the nonlinear echo path. For example, Volterra filters are a popular and very general concept of such nonlinear filtering techniques [2]. Nevertheless, the number of filter coefficients to be identified is exponentially increasing with higher-order filter kernels which makes the adaptation of all unknowns a cumbersome, if not impossible, task.

In this paper, we propose a comparably simple method to infer the significant coefficients in the nonlinear filter kernels from the linear filtering branch and to selectively prune the coefficients of the higher-order parts of the model. Therefore, we will first review both the block-based description of nonlinear models in Sec. 2 and the Volterra filtering concept in Sec. 3 before summarizing the scenario of nonlinear AEC in Sec. 4. The proposed estimation of significant coefficients and the subsequent pruning method are explained in Sec. 5 and results of various experiments are documented in Sec. 6. Finally, Sec. 7 concludes this work and provides some outlook on future research.

This project was supported by the Deutsche Forschungsgemeinschaft (DFG) under contract number KE 890/3.

2. BLOCK-BASED NONLINEAR MODELS FOR AEC

In order to motivate the proposed method in this contribution, we will first consider the input/output relations of a so-called Wiener-Hammerstein model as it is depicted in Fig. 1. Since this model consists of a series of linear and nonlinear operations, it is also referred to as *LNL (linear-nonlinear-linear) model*.

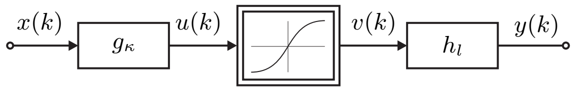

Fig. 1. Cascaded LNL (Wiener-Hammerstein) block model

According to Fig. 1, the output $u(k)$ of the first linear stage reads

$$u(k) = \sum_{\kappa=0}^{N_g-1} g_\kappa \cdot x(k-\kappa) \tag{1}$$

with the impulse response g_κ. Furthermore, $v(k)$ is given by

$$v(k) = \sum_{p=1}^{P} a_p \cdot u^p(k) \tag{2}$$

based on the assumption that the nonlinearity is memoryless and can be modelled sufficiently well by a truncated Taylor series of P-th order [3]. Inserting (1) into (2) and raising the power yields

$$\begin{aligned} v(k) &= \sum_{p=1}^{P} a_p \cdot \left(\sum_{\kappa=0}^{N_g-1} g_\kappa \cdot x(k-\kappa) \right)^p = \\ &= \sum_{p=1}^{P} a_p \cdot \left(\sum_{\kappa_1=0}^{N_g-1} \cdots \sum_{\kappa_p=0}^{N_g-1} \underbrace{g_{\kappa_1} \cdot \ldots \cdot g_{\kappa_p}}_{:= g_{\kappa_1,\ldots,\kappa_p}} \cdot \prod_{q=1}^{p} x(k-\kappa_q) \right) \end{aligned} \tag{3}$$

and reveals that the nonlinear input/output relation generally exhibits some memory, unless the filter g_κ equals a unit impulse. Finally, the output after the second filter is again given by

$$y(k) = \sum_{l=0}^{N_h-1} h_l \cdot v(k-l). \tag{4}$$

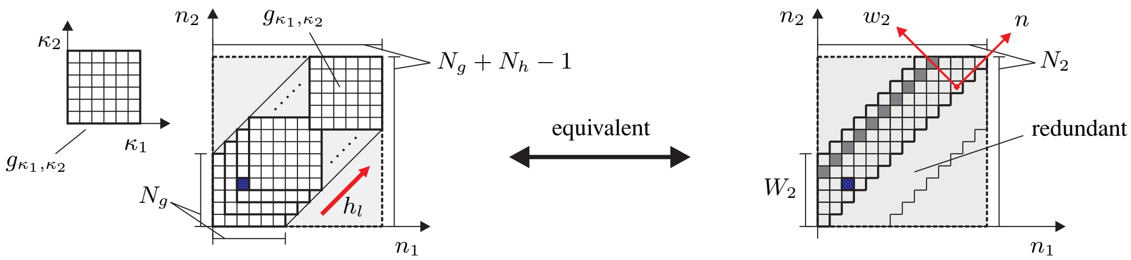

Fig. 2. Illustration of the construction of a second-order kernel by superposition of basic kernels g_{κ_1,κ_2} weighted by the linear filter h_l (left), comparison with a second-order Volterra kernel in diagonal coordinate representation (right)

Inserting (3) into (4) and re-arranging all of the signal components, the complete output of the LNL system is obtained by

$$y(k) = \sum_{p=1}^{P} a_p \cdot \left(\sum_{l=0}^{N_h-1} h_l \cdot \left[\sum_{\kappa_1=0}^{N_g-1} \cdots \sum_{\kappa_p=0}^{N_g-1} g_{\kappa_1,\ldots,\kappa_p} \times \prod_{q=1}^{p} x(k-l-\kappa_q) \right] \right), \quad (5)$$

with the p-dimensional *basic kernels* $g_{\kappa_1,\ldots,\kappa_p}$ as defined in (3). From (5) it can be seen that the filter coefficients which affect products of input samples are essentially given by a superposition of these basic kernels. As illustrated in Fig. 2 for the second-order case ($p = 2$), the $g_{\kappa_1,\ldots,\kappa_p}$ are thereby weighted by the linear filter coefficients h_l and shifted along the main diagonal.

3. VOLTERRA FILTERS IN DIAGONAL COORDINATES

Despite its advantage in the analysis of signal properties after the independent building blocks, the above cascade is, however, not well-suited for adaptive realizations, as the convergence to an optimum solution cannot be guaranteed [4]. Therefore, Volterra filters (VF) provide a more reliable approach to nonlinear filtering and can be understood as a generalization of the linear filtering concept [3]. The output of such a filter is given by the superposition

$$y(k) = \sum_{p=1}^{P} y_p(k) \quad (6)$$

where the individual kernel outputs $y_p(k)$ are obtained by

$$y_p(k) = \sum_{n_1=0}^{N_p-1} \sum_{n_2=n_1}^{N_p-1} \cdots \sum_{n_p=n_{p-1}}^{N_p-1} h^{(\mathrm{VF})}_{p,n_1,n_2,\ldots,n_p} \cdot \prod_{q=1}^{p} x(k-n_q), \quad (7)$$

which is closely related to multidimensional convolution and is based on the Cartesian coordinates $n_1, n_2, \ldots, n_p$.

Besides this, VFs may also be implemented by adopting the so-called diagonal coordinates which are formed by re-interpreting the coordinate indices as $n_1 := n$, $n_2 := w_2 + n$, $\ldots$, $n_p := w_p + n$. Using these definitions, the p-th order output reads

$$y_p(k) = \sum_{w_2=0}^{W_p-1} \cdots \sum_{w_p=w_{p-1}}^{W_p-1} \sum_{n=0}^{N_p-1-w_p} h^{(\mathrm{VF})}_{p,n,w_2+n,\ldots,w_p+n} \times x(k-n) \cdot \prod_{q=2}^{p} x(k-w_q-n). \quad (8)$$

which represents a processing scheme where computations are carried out over distinct diagonals with length $N_p - w_p$. As only a width W_p from the main diagonal of the higher-order kernels is covered, this allows for a flexible configuration of the nonlinear filter memory and may reduce algorithmic demands, whenever $W_p < N_p$ is a reasonable choice. For example, consider again Fig. 2 where W_2 can be identified with N_g of the LNL model and the elements of the diagonal $w_2 = 4$ are shaded in grey.

In order to simplify the notation of (8) in the following, we define the index vector $\underline{\mathbf{w}} := [w_2, \ldots, w_p]^{\mathrm{T}}$, the filter coefficients $h^{(\mathrm{VF})}_{p,\underline{\mathbf{w}},n} := h^{(\mathrm{VF})}_{p,n,w_2+n,\ldots,w_p+n}$ of each diagonal and the diagonal input signals $x_{p,\underline{\mathbf{w}}}(k-n) := x(k-n) \cdot \prod_{q=2}^{p} x(k-w_q-n)$ accordingly. This yields the Volterra kernel outputs in the form of:

$$y_p(k) = \sum_{w_2=0}^{W_p-1} \cdots \sum_{w_p=w_{p-1}}^{W_p-1} \sum_{n=0}^{N_p-1-w_p} h^{(\mathrm{VF})}_{p,\underline{\mathbf{w}},n} \cdot x_{p,\underline{\mathbf{w}}}(k-n). \quad (9)$$

For the sake of completeness, we point out that in both of the above representations, the symmetry of higher-order kernels [3] has been exploited by invoking only the unique coefficients in (7) and (8).

4. NONLINEAR ACOUSTIC ECHO CANCELLATION

Let us now consider an AEC task as depicted in Fig. 4 where the total echo path comprises an amplifier, a loudspeaker, the room impulse reponse (RIR) and a microphone. Since such a setup will exhibit a considerable amount of nonlinear distortions in presence of low-cost hardware, the performance of a purely linear adaptive filter is severely hampered. Thus we employ an adaptive VF in diagonal coordinates as nonlinear acoustic echo canceller (NLAEC).

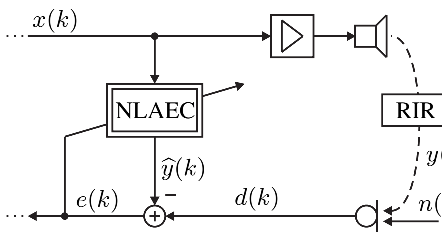

Fig. 3. Scenario for nonlinear acoustic echo cancellation (NLAEC)

Using the output $\widehat{y}(k)$ of the adaptive VF after (6) and (9) with the current coefficients $\widehat{h}^{(\mathrm{VF})}_{p,\underline{\mathbf{w}},n}(k)$ yields the residual error

$$e(k) = d(k) - \widehat{y}(k), \quad (10)$$

where the microphone reference is given by $d(k) = y(k) + n(k)$. The coefficient updates for all kernels p and diagonals $\underline{\mathbf{w}}$ are then performed according to [5]

$$\widehat{h}^{(\mathrm{VF})}_{p,\underline{\mathbf{w}},n}(k+1) = \widehat{h}^{(\mathrm{VF})}_{p,\underline{\mathbf{w}},n}(k) + \mu_p(k) \cdot e(k) \cdot x_{p,\underline{\mathbf{w}}}(k-n) \quad (11)$$

where the kernel-dependent step sizes are given by

$$\mu_p(k) := \frac{\alpha_p}{S_p(k) + \delta}. \quad (12)$$

Hence, the effective step size includes both a (fixed) parameter α_p and the normalization to the instantaneous kernel input energy

$$S_p(k) = \sum_{w_2=0}^{W_p-1} \cdots \sum_{w_p=w_{p-1}}^{W_p-1} \sum_{n=0}^{N_p-1-w_p} \Big(x_{p,\underline{\mathbf{w}}}(k-n) \Big)^2 \quad (13)$$

as well as some small regularization constant δ. The given algorithm thus represents an *SNLMS* adaptation scheme which performs a *separate normalization* of all kernels [6].

Inspection of the Volterra filtering equations in Sec. 3 reveals that the number of coefficients in the higher-order kernels is large and grows exponentially with the size N_p of the kernels. Regarding acoustic applications (like an NLAEC scenario), the challenges in adaptive nonlinear filtering become apparent as the length of the linear RIR also requires a great amount of memory for the nonlinear kernels.

5. PRUNING OF HIGHER-ORDER DIAGONAL COEFFICIENTS

In order to overcome the drawback of a potentially very large number of filter coefficients which have to be estimated by the adaptation algorithm, we will now consider the relation between cascaded LNL models and VFs. Comparing the filter coefficients in (5) and (7) which process the same input products, we see that $n_q = l + \kappa_q$ holds for all $q \in \{1, \ldots, p\}$ within the p-dimensional kernel of size $N_p = N_g + N_h - 1$. Thus, the following relations are found for the kernel coefficients and the LNL model parameters

$$h^{(\mathrm{VF})}_{p,n_1,n_2,\ldots,n_p} \equiv a^{(\mathrm{sym})}_p \cdot \sum_{l=0}^{N_h-1} h_l \cdot g_{n_1-l,n_2-l,\ldots,n_p-l} \quad (14)$$

$$h^{(\mathrm{VF})}_{p,\underline{\mathbf{w}},n} \equiv a^{(\mathrm{sym})}_p \cdot \sum_{l=0}^{N_h-1} h_l \cdot g_{n-l,(w_2+n)-l,\ldots,(w_p+n)-l} \quad (15)$$

where, for brevity, $a^{(\mathrm{sym})}_p$ captures another factor accounting for the number of possible index permutations due to the symmetry in the Volterra filtering [3]. Essentially, these identities imply that each LNL model has an equivalent P-th order Volterra representation whose coefficients are given by products of the LNL parameters g_κ, a_p and h_l. For illustration, consider the marked coefficient in the second-order kernel of Fig. 2 where $n_1 = 2$, $n_2 = 3$ and therefore $w_2 = 1$ and $n = 2$. Using (14), it can be obtained by

$$h^{(\mathrm{VF})}_{2,2,3} = a_2 \cdot 2 \cdot (h_0 \cdot g_{2,3} + h_1 \cdot g_{1,2} + h_2 \cdot g_{0,1}) \quad (16)$$

which represents the weighted superposition of three basic kernels and includes a factor of 2, since the coefficient appears twice in the LNL structure.

Without loss of generality, we can assume that the magnitudes of the linear filter coefficients are bounded, i.e. $0 \leq |g_\kappa|, |h_l| < 1$ for all valid κ, l. Note that this can always be ascertained, since the $a^{(\mathrm{sym})}_p$ may absorb any additional scaling. Regarding the multiplications in the kernels $g_{\kappa_1,\ldots,\kappa_p}$, we therefore notice that relatively sparse linear filters g_κ will result in even sparser basic kernels $g_{\kappa_1,\ldots,\kappa_p}$, as the magnitudes of the products tend to zero. Taking the size of these basic kernels into account, it can moreover be seen that the summation in (14) or (15) is limited to a maximum number of N_g contributions. Hence, under the assumption that the first linear stage of the LNL structure is sparse and relatively short compared to N_h, we conclude that the general shape of higher-order diagonals in the equivalent VF resembles the envelope of the linear kernel:

$$h^{(\mathrm{VF})}_{1,n} = a_1 \cdot \sum_{l=0}^{N_h-1} h_l \cdot g_{n-l} = a_1 \cdot \big(h_n * g_n\big). \quad (17)$$

Note that the above assumption $N_g < N_h$ is valid for an NLAEC scenario, if g_κ models the amplifier and the loudspeaker.

In order to exploit this similarity for complexity reduction, we now propose a simple method for coarsely estimating the importance of higher-order diagonal elements from the linear branch of the VF. Since the VF (and thus also the linear kernel) is realized adaptively in the regarded NLAEC scenario, there is a need for a robust quantity which can be used to infer the general shape of all higher-order diagonals $\widehat{h}^{(\mathrm{VF})}_{p,\underline{\mathbf{w}},n} (p \geq 2)$ from $\widehat{h}^{(\mathrm{VF})}_{1,n}$. For this purpose, we monitor the *coefficient energy* of the linear filter taps by calculating

$$\widetilde{E}_n(k) = \Big(\widehat{h}^{(\mathrm{VF})}_{1,n}(k)\Big)^2 + 0.9 \cdot \widetilde{E}_{n-1}(k) \quad (18)$$

which applies some 1st-order IIR low-pass filtering along n, in order to obtain a smoother shape. Note that due to the relatively fast convergence of the linear kernel, the time-variance of $\widetilde{E}_n(k)$ can be considered negligible for a coarse estimation of the envelope. Through analysis of the measure, the significant coefficients of the higher-order diagonals can be inferred from the shape of the coefficient energy envelope in the linear kernel. Introducing the binary masking

$$\beta_n(k) := \begin{cases} 1, & \text{if } \widetilde{E}_n(k) \geq \chi \cdot \max\limits_i \Big\{\widetilde{E}_i(k)\Big\} \\ 0, & \text{else} \end{cases} \quad (19)$$

reduces the total number of VF coefficients as it serves as a switch for in- or exclusion of the corresponding diagonal elements. This *coefficient pruning* can be controlled by the threshold parameter χ which denotes the relevant fraction of the maximum tap energy.

Consequently, for all higher-order kernels ($p \geq 2$) the filtering from (9) is modified as follows

$$\widehat{y}_p(k) = \sum_{w_2=0}^{W_p-1} \cdots \sum_{w_p=w_{p-1}}^{W_p-1} \sum_{n=0}^{N_p-1-w_p} \beta_n(k) \cdot \widehat{h}^{(\mathrm{VF})}_{p,\underline{\mathbf{w}},n}(k) \cdot x_{p,\underline{\mathbf{w}}}(k-n) \quad (20)$$

and the SNLMS updates are performed according to:

$$\widehat{h}^{(\mathrm{VF})}_{p,\underline{\mathbf{w}},n}(k+1) = \widehat{h}^{(\mathrm{VF})}_{p,\underline{\mathbf{w}},n}(k) + \beta_n(k) \cdot \mu_p(k) \cdot e(k) \cdot x_{p,\underline{\mathbf{w}}}(k-n). \quad (21)$$

As can be seen, all diagonal elements corresponding to $\beta_n(k) = 0$ are removed from the filtering and updating steps. Therefore, this technique selectively regards only those Volterra kernel coefficients with a defined significance and will be referred to as *pruning of higher-order diagonals (PHD)*.

6. EXPERIMENTAL RESULTS

To demonstrate the effectiveness of the proposed pruning method, several experiments with simulated LNL models and Volterra systems measured from real hardware have been conducted. All results have been obtained by examining the NLAEC setup from Fig.4 where the echo path has been simulated by means of the extracted kernels. Moreover, it has been ascertained that the signal components of $y(k)$ exhibit a linear-to-nonlinear power ratio of 10 dB and some white noise $n(k)$ is added with an SNR of 30 dB which constitutes a realistic scenario for up-to-date mobile phones. The step

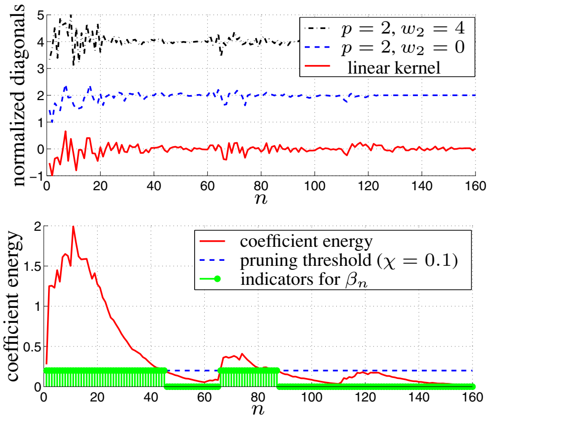

Fig. 4. Illustration of kernel diagonals (top), coefficient energy and pruning threshold for a simulated LNL system (bottom)

sizes of the adaptation have been set to $\alpha_{1/2} = 0.3/0.2$ for all experiments.

In the first experiment, an LNL model with $N_g = 16$, $N_h = 241$ has been simulated and the nonlinearity was given by a polynomial with $a_1 = 1.3$ and $a_2 = -0.5$. According to these parameters, the sizes of the adaptive second-order VF has been chosen as $N_{1/2} = 256/128$ and $W_2 = 16$. Fig. 4 visualizes both the similar shapes of the linear kernel and second-order diagonals as well as the corresponding energy measure $\widehat{E}_n$ and the resulting pruning to the significant coefficients for a threshold level $\chi = 0.1$. Since non-zero values of β_n only occur in highly populated regions of $h_{1,n}^{(\mathrm{VF})}$, the number of coefficients in the adaptation is lowered from a total of 2184 to 1312. Although this implies computational savings of approx. 40%, both the full and the pruned adaptation yield comparable echo cancellation performance as shown in Fig. 5 for both speech-like coloured noise and real speech.

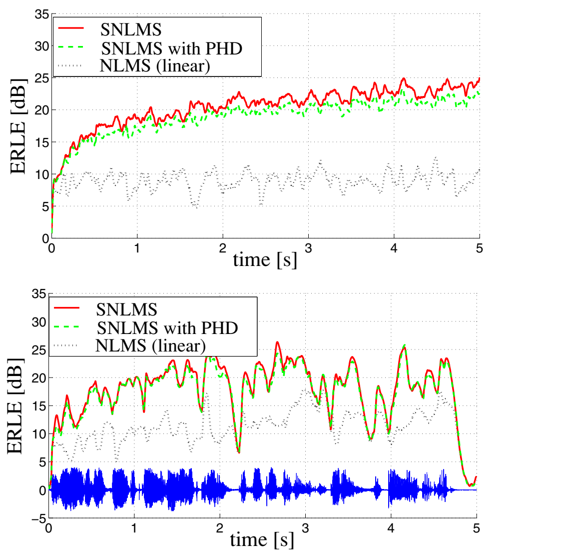

Fig. 5. ERLE for a simulated LNL system with speech-like coloured noise (top) and male speech (bottom)

Finally, Fig. 6 illustrates the resulting ERLE if the PHD method is applied to a real Volterra sytem with $N_{1/2} = 320/64$ and $W_2 = 32$ as obtained by measurements from a small loudspeaker. Clearly, the performance of the PHD method lacks some of its potential in this situation as the assumed Wiener-Hammerstein model seems to be an inappropriate model for the real nonlinearity. Nevertheless the VF using the PHD still provides noticeable gains for segments of highly nonlinear distortions compared to a linear approach. Furthermore, considering the reduction of coefficients (again approx. 40%), the pruning can also be seen as a means for trading filter complexity against performance.

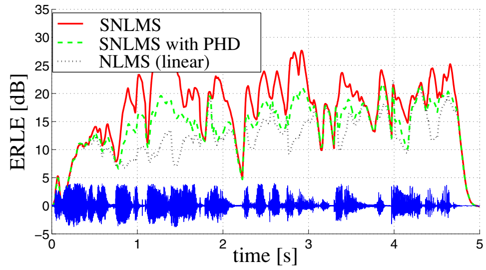

Fig. 6. ERLE for a real Volterra system with male speech

7. CONCLUSIONS

We discussed the relation between Wiener-Hammerstein or LNL models and Volterra filters in diagonal coordinates. Based on the similarity of the coefficient envelopes in the linear kernel and on higher-order diagonals, a simple method has been proposed for detecting the most important filter coefficients in the nonlinear kernels. Although pruning the less significant parts of the nonlinear adaptive filter results in a reduction of the algorithmic complexity, experiments for simulated and measured Volterra systems have shown that the echo cancellation performance is largely unaffected. Future work will focus on the refinement of this technique and on the possibility of translating these results to DFT-domain implementations.

8. REFERENCES

[1] R. Niemistö and T. Mäkelä, "On performance of linear adaptive filtering algorithms in acoustic echo cancellation in presence of distorting loudspeakers," in *Proc. Int. Workshop on Acoustic Echo and Noise Control (IWAENC)*, Kyoto, Sept. 2003.

[2] A. Guérin, G. Faucon, and R. Le Bouquin-Jeannès, "Nonlinear acoustic echo cancellation based on Volterra filters," *IEEE Trans. on Acoustics, Speech, and Signal Processing*, vol. 11, no. 6, pp. 672–683, Nov. 2003.

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

[4] J.-P. Costa, A. Lagrange, and A. Arliaud, "Acoustic echo cancellation using nonlinear cascade filters," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, Hongkong, April 2003.

[5] F. Küch and W. Kellermann, "Nonlinear acoustic echo cancellation," in *E. Hänsler and G. Schmidt (eds.), Topics in Acoustic Echo and Noise Control*, Heidelberg: Springer, June 2006, pp. 205–257.

[6] M. Zeller and W. Kellermann, "Framewise repeated coefficient updates for enhanced nonlinear AEC by diagonal coordinate Volterra filters," in *Proc. Joint Workshop on Hands-free Speech Communication and Microphone Arrays (HSCMA)*, Trento, May 2008.