SIGNIFICANCE-AWARE HAMMERSTEIN GROUP MODELS FOR NONLINEAR ACOUSTIC ECHO CANCELLATION

Christian Hofmann, Christian Huemmer, and Walter Kellermann

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

ABSTRACT

In this work, a novel approach for nonlinear acoustic echo cancellation is proposed. The main innovative idea of the proposed method is to model only the small region of the echo path around the direct path by a group of parallel Hammerstein models, to estimate a nonlinear preprocessor by correlations between the linear kernels of the Hammerstein submodels, and to describe the remaining echo path by a simple Hammerstein model with the preprocessor determined in the aforementioned way. While the computational complexity of such a system increases only slightly in comparison to a linear echo canceller, experiments with speech recordings from a smartphone in different environments confirm a significantly increased echo cancellation performance.

Index Terms— nonlinear AEC, NLAEC, SA-HGM, HGM

1. INTRODUCTION

Since the introduction of the first adaptive echo cancellers [1], a plethora of adaptation algorithms have been proposed [2–4], some of which even dedicate special attention to characteristic impulse response shapes [5–8]. Nevertheless, the performance of classical, linear Acoustic Echo Cancellation (AEC) is limited if nonlinearities of the echo path, as they typically occur for loudspeakers and amplifiers used in cheap or miniaturized devices [9], become significant. Although the suppression of nonlinear residual echos [10–12] is a very powerful tool, distortions of the desired signal are unavoidable, so that nonlinear echo cancellation is of vital importance.

Within the last two decades, numerous nonlinear echo-path models have been investigated. Featuring prominently, adaptive Volterra filters are very general, but also computationally expensive models [13, 14], for which computationally cheap approximations [15], efficient implementations [16], and self-configuring evolutionary structures exist [17, 18]. Considering only the main diagonals of a Volterra filter's kernels results in so-called power filters [19], which are again very similar to the Fourier-series-expansion-motivated group of parallel Hammerstein models in [20], where sinusoïds replace the monomial blocks of a power filter. Systems with this common structure will be denoted as Hammerstein Group Models (HGMs) in the following (see Section 3.2). Similar to the latter, the Functional-Link Adaptive Filters (FLAFs) in [21] employ a linear echo path model, adaptively augmented with a sinusoïdal model for the nonlinearities. All those nonlinear models have in common that the models' outputs linearly depend on the filter coefficients to be identified, so that the adaptation can be performed in the same way as for multichannel linear filters. When giving up the constraint of output linearity, simple nonlinear preprocessors followed by a linear filter can be realized, such as memoryless polynomials [22] or piece-wise-linear functions [23], adapted by a gradient-descent algorithm. Another computationally efficient but slowly adapting model was proposed in [24], which combines two filters out of a set of linear filters, each of which aims at capturing nonlinearities resulting from different input levels. Furthermore, regression techniques widely employed in machine learning, such as artificial neural networks, have been proposed for modeling the nonlinearities [25, 26].

In this work, we propose a computationally efficient trade-off between power-filter-like HGMs on the one hand, and a linear filter augmented with a very efficiently adapted preprocessor on the other hand. This novel structure is denoted as Significance-Aware HGM (SA-HGM).

2. NOTATIONS

Vectors are written as boldface lowercase letters, matrices are typeset with capital boldface letters, and $\mathrm{tr}\{\mathbf{A}\}$ denotes the trace of a matrix $\mathbf{A}$. Furthermore, $\mathbf{A} \odot \mathbf{B}$ stands for the element-wise multiplication (Hadamard product) of $\mathbf{A}$ and $\mathbf{B}$, $\Sigma\{\cdot\}$ sums over all elements of the vector or matrix on which Σ is applied to, as well as $\langle \mathbf{a}, \mathbf{b} \rangle$ and $\langle \mathbf{A}, \mathbf{B} \rangle$ denote the scalar product $\mathbf{a}^{\mathrm{T}}\mathbf{b}$ or the Hilbert-Schmidt scalar product $\mathrm{tr}\{\mathbf{A}^{\mathrm{T}}\mathbf{B}\} = \sum\{\mathbf{A} \odot \mathbf{B}\}$ (also known as Frobenius inner product), respectively. The operator $\mathrm{diag}(\mathbf{a})$ transforms a vector $\mathbf{a} = [a_i]$ into a diagonal matrix with diagonal elements $d_{i,i} = a_i$.

3. ECHO PATH MODELS

3.1. Linear FIR Models

A very simple echo path model is an adaptive linear Finite-Impulse-Response (FIR) system. Such a system is completely described by $y(k) = x(k) * h(k)$, where k is the discrete-time sample index, $x(k)$ is the input signal, $h(k)$ is the system's impulse response of length L, and where $*$ denotes linear convolution. By introducing the input-signal and impulse-response vectors

$$\mathbf{x}(k) = [x(k), x(k-1), \ldots, x(k-L+1)]^{\mathrm{T}} \quad (1)$$

$$\mathbf{h} = [h(0), h(1), \ldots, h(L-1)]^{\mathrm{T}}, \quad (2)$$

the input-output relation can be written as

$$y(k) = \langle \mathbf{h}, \mathbf{x}(k) \rangle . \quad (3)$$

For practical applications, the filter coefficients of such models are typically adapted by Least-Mean-Square (LMS)-type algorithms, such as the Normalized Least-Mean-Square (NLMS) algorithm (see [3] for an extensive review of such algorithms).

3.2. Hammerstein Group Models

A simple nonlinear model is a Hammerstein model, consisting of a memoryless nonlinearity and a subsequent linear system. Such a structure can approximate the behavior of nonlinearly distorting playback equipment (amplifier and loudspeaker) followed by the propagation of the radiated sound waves through the medium air to a microphone.

978-1-4799-2893-4/14/$31.00 ©2014 IEEE

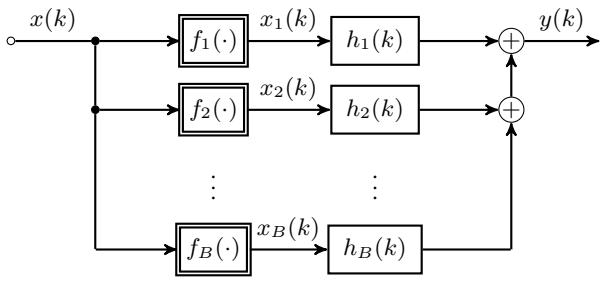

Fig. 1. Block diagram of a Hammerstein group model.

Hammerstein Group Models (HGMs) are comprised of a group of B parallel Hammerstein models, denoted as B branches. The block diagram of this structure is depicted in Fig. 1 where the input-output relation can be written as

$$y(k) = \sum_{b=1}^{B} x_b(k) * h_b(k), \tag{4}$$

where b is the branch index,

$$x_b(k) = f_b\{x(k)\} \tag{5}$$

is the b^{th} branch signal, and $h_b(k)$ is called the linear kernel and $f_b(\cdot)$ the nonlinear base function of branch b. Traditionally employed HGMs are the so-called power filters [19]. More recently, Fourier-base HGMs have been proposed in [20]. The HGM's output can be expressed as

$$y(k) = \langle \mathbf{H}, \mathbf{X}(k) \rangle \tag{6}$$

with the branch signal matrix

$$\mathbf{X}(k) = [\mathbf{x}_1(k), \ldots, \mathbf{x}_B(k)] = [f_1(\mathbf{x}(k)), \ldots, f_B(\mathbf{x}(k))] \tag{7}$$

and the kernel matrix

$$\mathbf{H} = [\mathbf{h}_1, \ldots, \mathbf{h}_B], \mathbf{h}_b = [h_b(0), h_b(1), \ldots, h_b(L-1)]^{\mathrm{T}}. \tag{8}$$

As the output of an HGM linearly depends on its kernel coefficients $\mathbf{H}$ and the branch signals $\mathbf{X}(k)$, classical LMS-type algorithms can be employed to adapt $\hat{\mathbf{H}}(k)$ of an adaptive HGM with a fixed set of base functions $f_1(\cdot), \ldots, f_B(\cdot)$ to model a physical system, employing the physical system's input and output signals.

Furthermore, it is important to note, that a linear combination of branch-signal vectors according to

$$\mathbf{x}_{\text{NL}}(k) = \mathbf{X}(k)\mathbf{w}, \ \mathbf{w} = [w_1, \ldots, w_B]^{\mathrm{T}} \tag{9}$$

with arbitrary $\mathbf{w}$ describes a memoryless nonlinearity. Subsequent linear filtering of such an $\mathbf{x}_{\text{NL}}(k)$ represents a single Hammerstein model with input $x(k)$, which is equivalent to an HGM with linearly dependent kernel vectors in $\mathbf{H}$.

3.3. Significance-Aware Hammerstein Group Models

For introducing the novel echo cancellation scheme, consider a partitioning of a Room Impulse Response (RIR) vector $\mathbf{h}$, as outlined in Fig. 2, into a direct-path part $\mathbf{h}_{\text{p}}$ (green) and the remaining part $\mathbf{h}_{\text{r}}$ (blue), where the splitting is performed according to the peak of the RIR energy smoothed along the taps (red). This can be expressed as $\mathbf{h}_{\text{p}} = \mathbf{S}_{\text{p}}\mathbf{h}$ and $\mathbf{h}_{\text{r}} = \mathbf{S}_{\text{r}}\mathbf{h}$ employing the selection matrices

$$\mathbf{S}_{\text{p}} = \text{diag}([s_{\text{p},i}]), \ s_{\text{p},i} = \begin{cases} 1 & |i - i_{\text{peak}}| \leq r \\ 0 & \text{otherwise} \end{cases} \tag{10}$$

$$\mathbf{S}_{\text{r}} = \mathbf{I} - \mathbf{S}_{\text{p}}, \tag{11}$$

where i_{peak} is the time lag of the RIR energy peak, marked in Fig. 2 with a black dot. Consequently, $\mathbf{h}_{\text{p}}$ captures an interval of $\mathbf{h}$ of length $L_{\text{p}} = 2r + 1$ which is centered around i_{peak}.

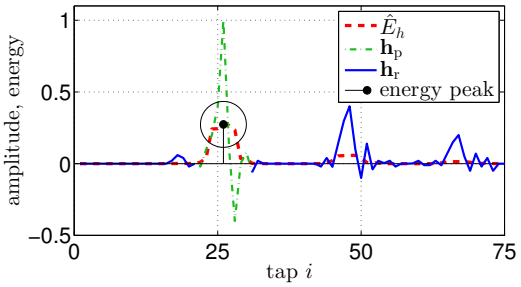

Fig. 2. Splitting of impulse response vector $\mathbf{h}$ in an energy-peak-centered direct-path part $\mathbf{h}_{\text{p}}$ and the remaining rest part $\mathbf{h}_{\text{r}}$.

Due to the dominant linear component of the Loudspeaker-Enclosure-Microphone (LEM) system, it seems reasonable to assume that a significant part of the system can be approximated as a single Hammerstein model and that the nonlinearity is easiest to observe for input-output time lags where a significant amount of signal energy is transmitted. Based on this assumption and employing the recently introduced temporal decomposition of the LEM system according to Fig. 2, we propose the adaptive structure depicted in Fig. 3 for system identification and echo cancellation. For illustration purposes, we emphasize the different model components throughout the following text by colors complying with Fig. 3. We denote this structure with output $\hat{y}_{\text{SA}}(k)$ as Significance-Aware HGM (SA-HGM), because

a) it observes the nonlinearity of the system only in a short, well identifiable region of the RIR (preferably in the direct-path-peak region) by means of an HGM (block b) with fixed base functions $f_b(\cdot)$ and adaptive kernel matrix $\hat{\mathbf{H}}(k) = [\hat{\mathbf{h}}_1(k), \ldots, \hat{\mathbf{h}}_B(k)]$,

b) it approximates this HGM by a Hammerstein model, which contains the memoryless preprocessor $\mathbf{x}_{\text{pp}}(k) = \mathbf{X}(k)\hat{\mathbf{w}}_{\text{pp}}(k)$ with normalization $w_{\text{pp},1} = 1$, where $\hat{\mathbf{w}}_{\text{pp}}(k)$ is derived from $\hat{\mathbf{H}}(k)$, and

c) it efficiently employs the memoryless preprocessor (b) extracted from the peak-modeling HGM as preprocessor part (block c) of another Hammerstein model (block d) with linear kernel $\hat{\mathbf{h}}(k) = \hat{\mathbf{h}}_{\text{p}}(k) + \hat{\mathbf{h}}_{\text{r}}(k)$ to model a long echo path.

The advantage of this approach is that the HGM models only the direct-path region of the echo path and the remaining, long nonlinear echo path is approximated by a low-complexity Hammerstein model (blue-colored taps in $\hat{\mathbf{h}}_{\text{r}}$ in block d). The system in Fig. 3 provides the following partial-path echo signal estimates

$$\hat{y}_{\text{r}}(k) = \left\langle \hat{\mathbf{h}}_{\text{r}}(k), \mathbf{x}_{\text{pp}}(k) \right\rangle \tag{12}$$

$$\hat{y}_{\text{p}}(k) = \left\langle \hat{\mathbf{h}}_{\text{p}}(k), \mathbf{x}_{\text{pp}}(k) \right\rangle \tag{13}$$

$$\hat{y}_{\text{HGM}}(k) = \left\langle \hat{\mathbf{H}}(k), \mathbf{X}(k) \right\rangle \tag{14}$$

and the complete-path echo signal estimates

$$\hat{y}_{\text{HM}}(k) = \left\langle \hat{\mathbf{h}}(k), \mathbf{x}_{\text{pp}}(k) \right\rangle = y_{\text{p}}(k) + y_{\text{r}}(k) \tag{15}$$

$$\hat{y}_{\text{SA}}(k) = y_{\text{HGM}}(k) + y_{\text{r}}(k) = \left\langle \hat{\mathbf{H}}(k) + \hat{\mathbf{H}}_{\text{r}}(k), \mathbf{X}(k) \right\rangle, \tag{16}$$

where $\hat{\mathbf{H}}(k)$ is the sparsely populated peak-HGM kernel matrix and $\hat{\mathbf{H}}_{\text{r}}(k) = [\hat{w}_{\text{pp},1}(k)\hat{\mathbf{h}}_{\text{r}}(k), \ldots, \hat{w}_{\text{pp},B}(k)\hat{\mathbf{h}}_{\text{r}}(k)]$ is the extension efficiently implemented as the blue component of the low-complexity Hammerstein model. The involved linear subsystems, all captured

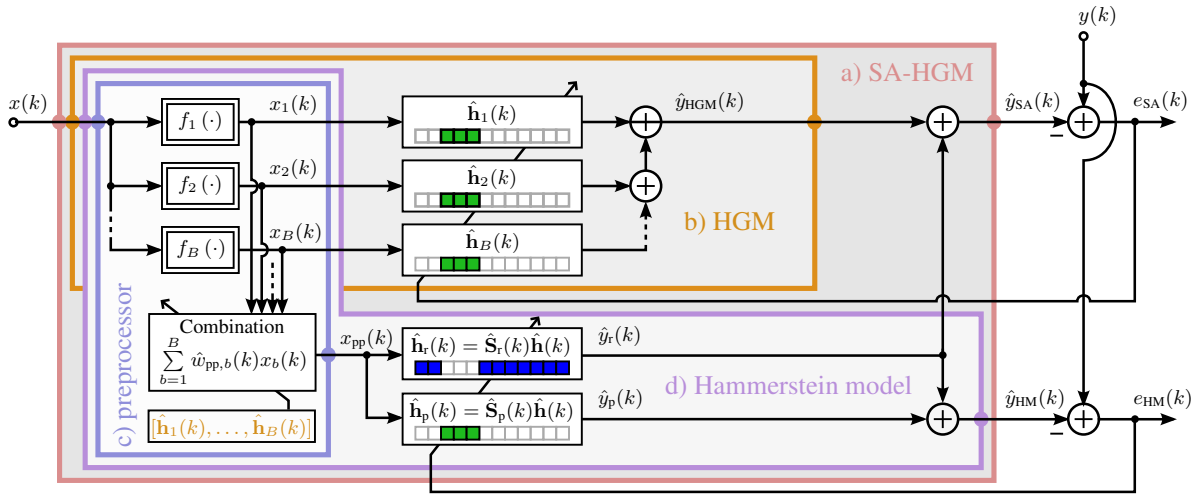

Fig. 3. System identification by an SA-HGM (block a), consisting of the peak-modeling HGM (block b), sharing signals with the memoryless preprocessor (block c) for an adaptive Hammerstein model (block d) extrapolating the peak-HGMs nonlinearities out of the peak region.

by $\hat{\mathbf{H}}(k)$ and $\hat{\mathbf{h}}(k)$, can be adapted by classical LMS-type algorithms employing the error signals $e_{\mathrm{SA}}(k)$ and $e_{\mathrm{HM}}(k)$, respectively. Thus, online reclustering into a peak-region and the remaining part by means of $\hat{\mathbf{S}}_{\mathrm{p}}(k)$ and $\hat{\mathbf{S}}_{\mathrm{r}}(k)$ can be performed according to the energy of $\hat{\mathbf{h}}(k)$. Consequently, the sparsity of the kernel matrices is reflected by $\hat{\mathbf{H}}(k) = \hat{\mathbf{S}}_{\mathrm{p}}(k)\hat{\mathbf{H}}(k)$ and $\hat{\mathbf{H}}_{\mathrm{r}}(k) = \hat{\mathbf{S}}_{\mathrm{r}}(k)\hat{\mathbf{H}}_{\mathrm{r}}(k)$.

If the underlying physical system has Hammerstein structure and is described by $y(k) = \langle \mathbf{h},\, \mathbf{X}(k)\mathbf{w}_{\mathrm{pp}} \rangle$, the optimal kernel matrix $\hat{\mathbf{H}}_{\mathrm{opt}}(k)$ is time-invariant and has linearly dependent columns as $\hat{\mathbf{H}}_{\mathrm{opt}}(k) = \mathbf{H} = [w_{\mathrm{pp},1}\mathbf{h}_{\mathrm{p}}, \ldots, w_{\mathrm{pp},B}\mathbf{h}_{\mathrm{p}}]$ with the true direct-path RIR vector $\mathbf{h}_{\mathrm{p}}$. In this case,

$$w_{\mathrm{pp},b} = \tfrac{\langle \mathbf{h}_b, \mathbf{h}_1 \rangle}{\langle \mathbf{h}_1, \mathbf{h}_1 \rangle} \tag{17}$$

holds with equality.

Of course, in practice, a slight model mismatch is very likely, as well as only suboptimal estimates $\hat{\mathbf{w}}_{\mathrm{pp}}(k)$,$\hat{\mathbf{h}}(k)$,$\hat{\mathbf{S}}_{\mathrm{p}}(k)$ and $\hat{\mathbf{H}}(k)$ are available, especially for a time-varying system. Nevertheless, introducing the decomposition

$$\hat{\mathbf{h}}_b(k) = \tilde{w}_{\mathrm{pp},b}(k)\tilde{\mathbf{h}}_{\mathrm{p}}(k) + \boldsymbol{\epsilon}_b(k) \tag{18}$$

with the intermediate estimates $\tilde{w}_{\mathrm{pp},b}(k)$ and $\tilde{\mathbf{h}}_{\mathrm{p}}(k)$ and an error vector $\boldsymbol{\epsilon}_b(k)$, intermediate least-squares estimates

$$\tilde{\mathbf{h}}_{\mathrm{p}}^{(\mathrm{LS})}(k) = \underset{\tilde{\mathbf{h}}_{\mathrm{p}}(k)}{\operatorname{argmin}} \, \|\boldsymbol{\epsilon}_1(k)\|_2^2 \,\big|_{\tilde{w}_{\mathrm{pp},1}(k)=1} = \hat{\mathbf{h}}_1(k) \tag{19}$$

and

$$\tilde{w}_{\mathrm{pp},b}^{(\mathrm{LS})}(k) = \underset{\tilde{w}_{\mathrm{pp},b}(k)}{\operatorname{argmin}} \, \|\boldsymbol{\epsilon}_b(k)\|_2^2 \,\big|_{\tilde{\mathbf{h}}_{\mathrm{p}}(k)=\tilde{\mathbf{h}}_{\mathrm{p}}^{(\mathrm{LS})}(k)} \tag{20}$$

$$= \tfrac{\langle \hat{\mathbf{h}}_b(k), \hat{\mathbf{h}}_1(k) \rangle}{\langle \hat{\mathbf{h}}_1(k), \hat{\mathbf{h}}_1(k) \rangle} \tag{21}$$

can be obtained. These $\tilde{w}_{\mathrm{pp},b}^{(\mathrm{LS})}(k)$ are employed to refine the overall system as mixing weights for the memoryless preprocessor after temporal smoothing by

$$\hat{w}_{\mathrm{pp},b}(k+1) = \gamma \cdot \hat{w}_{\mathrm{pp},b}(k) + (1-\gamma) \cdot \tilde{w}_{\mathrm{pp},b}^{(\mathrm{LS})}(k). \tag{22}$$

Thereby, the adaptive linear filter $\hat{\mathbf{h}}(k)$ can perform better and $\hat{\mathbf{H}}(k)$ is consequently adapted on a signal containing less unmodeled signal components, affecting the adaptation equivalently to noise. This also leads to improved estimates $\tilde{w}_{\mathrm{pp},b}^{(\mathrm{LS})}(\kappa)$, $\kappa > k$.

The benefit of the SA-HGM is that the complex HGM has to model only a small subset of the nonlinear system for observing significant nonlinear components. This region of the system response describes the transmission of a significant amount of signal energy and is therefore less affected by gradient noise of the adaptation algorithm than other, lower-energy taps would be (see [27] for corresponding experiments with Volterra filters).

Embedding the core idea into a practical framework: As the adaptation of the HGM component depends on the adaptation of the linear model and vice versa, the system has to start in a well-defined state and iteratively converge to a good solution.

To this end, three adaptation phases are distinguished:

- **phase 1**: initial convergence of K_1 updates of $\hat{\mathbf{h}}(k)$ with $\hat{\mathbf{S}}_{\mathrm{r}}(k) = \mathbf{I}$ and employing a linearly configured preprocessor with $\hat{w}_{\mathrm{pp},b}(k) = 0\,\forall b > 1$, implying $f_1(x(k)) = x(k)$,
- **phase 2**: convergence phase for the HGM's $\hat{\mathbf{H}}(k)$ with K_2 updates during which the preprocessor stays frozen, and
- **phase 3**: continuous adaptation of both the linear model $\hat{\mathbf{h}}(k)$ in block d and the preprocessor by means of the HGM.

Of course, hard switching of the preprocessor coefficients at the transition from phase 2 to phase 3 requires rescaling the RIR with

$$g = \tfrac{\langle \hat{\mathbf{h}}_1(k), \hat{\mathbf{h}}_{\mathrm{p}}(k) \rangle}{\langle \hat{\mathbf{h}}_{\mathrm{p}}(k), \hat{\mathbf{h}}_{\mathrm{p}}(k) \rangle} \tag{23}$$

to prevent unnecessary readaptation of $\hat{\mathbf{h}}(k)$. Considering the rare case of an abrupt change of the RIR, such that the peak of the RIR energy lies outside of the previous direct-path interval, a reconvergence phase for the HGM with frozen preprocessor (phase 2) has to be triggered, again.

Comparison of computational complexity: The calculation of the energy peak within $\hat{\mathbf{h}}(k)$ and also rescaling of $\hat{\mathbf{h}}(k)$ after switching the preprocessor can be performed tap-wise, distributed across multiple sampling intervals without loss of AEC performance and hence, their contribution to the computational load can be neglected. Thereby, the number of real-valued multiplications for the different algorithms scale with L, L_{p}, and B as given in Table 1. As can be seen, the complexity of the adaptation of the nonlinear preprocessor is completely decoupled from the final length of the RIR and therefore only has a minor contribution to the computational complexity as long as $L_{\mathrm{p}} \ll L$. Note that the effort for computing the branch

Table 1. Comparison of approximate computational complexity in terms of real-valued multiplications.

	filtering and filter update	energy estimation
linear (NLMS)	$O(2L)$	$O(1)$
HGM (NLMS)	$O(2LB)$	$O(B)$
SA-HGM (NLMS)	$O(2L + 3L_\mathrm{p}B)$	$O(B+1)$

signals is excluded from Table 1 because this complexity depends on the branch nonlinearities themselves and does not contribute much to the overall computational load for typical power filters or for the branch nonlinearities employed later on.

Relation to prior work: As the HGM to be estimated is much smaller than in the full-HGM approach of [19], the proposed algorithm is dramatically less complex (recall Table 1). Nevertheless, the observed nonlinearities are employed for a preprocessor extending beyond the time-lags of the echo path modeled by the HGM, a property which is also not included in the FLAFs of [21].

The pure preprocessors in [22, 23] are updated by correlations between a long impulse response vector, distorted signal vectors (corresponding to branch signals for an HGM), and the error signal. For the proposed SA-HGM, we employ the branch signals to perform more accurate AEC by means of an HGM and employ correlations between the correspondingly short branch kernels directly as estimates for the preprocessor coefficients. Thereby, the pure preprocessors described in [22,23] even result in a higher computational complexity than the proposed method with $L_\mathrm{p} \ll L$.

4. EXPERIMENTS

4.1. Experimental Setup

The experiments are conducted with real recordings of about 80 s of double-talk-free speech played back and recorded by a smartphone device in a living-room-like environment (Room A), as well as in an anechoic environment (Room B). Employing the anechoic signal, the echo signals for two further rooms are synthesized by convolution of the anechoic microphone signal with impulse responses measured in the respective environments. The rooms are denoted as Room C, having a reverberation time $T_{60} \approx 200$ ms, and Room D, characterized by $T_{60} \approx 400$ ms.

In the following experiments, signal processing is performed at the sampling rate $f_s = 16$ kHz. The models are parametrized by the filters' length $L = 300$ taps, $B = 5$ branches for the HGMs, and a peak region width $L_\mathrm{p} = 11$ taps for the SA-HGM. The base-function set of the HGMs comprises the odd-order Legendre polynomials [28] of the orders 1 to 9, because these base functions revealed better modeling capabilities than monomes or sinusoïds in the considered scenario. Furthermore, the SA-HGM adaptation time constants are chosen as $K_1 = 3L$ and $K_2 = 100 \cdot L_\mathrm{p}B$ and the filter coefficients of all filters are adapted by the NLMS algorithm with a stepsize of 0.1.

As AEC performance measure, the commonly used Echo-Return-Loss Enhancement

$$\mathrm{ERLE} = 10 \log_{10} \left(\frac{P_y}{P_{e_\nu}} \right) \text{ dB} \tag{24}$$

is employed, where P_y and P_{e_ν} denote short-time estimates of the power of the microphone signal and error signal, respectively.

4.2. Experimental Results

The adaptation of the aforementioned filter structures is demonstrated in Fig. 4 for Room A by visualization of the reference signal and the microphone signal next to ERLE curves for different algorithms, where the instantaneous ERLE is computed over short-time intervals of 200 ms. Obviously, the linear filter (blue curve) performs worst. The use of a 5-branch Legendre-base HGM (light green) yields a better approximation of the nonlinear system, resulting in an increased ERLE after convergence, as well as a steeper initial convergence due to the multiple branch filters. When employing an SA-HGM (solid red), both convergence speed and modeling accuracy after convergence resemble rather the behavior of the full HGM than of the linear filter. Interestingly, the SA-HGM even outperforms the full HGM in some time intervals.

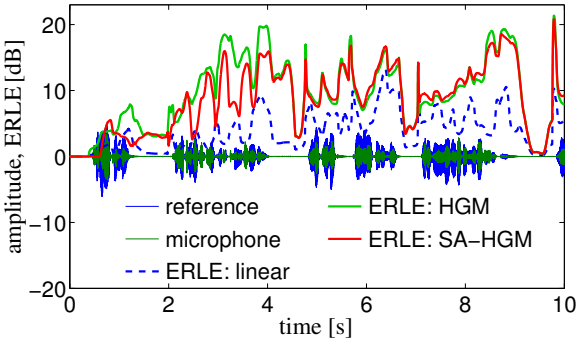

Fig. 4. Initial convergence of different AEC structures in Room A during the first 10 s.

Table 2. Average ERLE for different rooms and algorithms along with their relative number of Real-valued MULtiplications (RMUL), where all HGMs have 5 branches.

	Room A	Room B	Room C	Room D	RMUL
linear	9.8 dB	9.27 dB	9.45 dB	8.95 dB	1
full HGM	13.8 dB	16.6 dB	15.6 dB	13.6 dB	5
SA-HGM	14.3 dB	16.4 dB	15.1 dB	13.7 dB	1.3

The average ERLE value for the entire sequence is given in Table 2 along with the corresponding results for Room B, Room C, and Room D. In all cases, the full HGM and the SA-HGM perform similar and both lead to significantly better results than the linear AEC. For Rooms A and D, the SA-HGM even performs slightly better than the full HGM. But, compared to the linear AEC, the SA-HGM setup only demands for about 30% increased computational effort, which is just a small fraction of the effort needed for a full HGM. Hence, the HGM can be approximated successfully for a smartphone AEC application by an SA-HGM. The computational complexity of the adaptation of such a nonlinear AEC system is increased only insignificantly with respect to a linear filtering operation.

5. CONCLUSION

For typical acoustic echo paths, where the playback components are a major source of nonlinear distortions, HGMs (including power filters) can be efficiently approximated by the proposed SA-HGM structure. Thereby, a slight loss of nonlinear modeling capabilities is accepted in exchange for significantly reduced computational complexity and reduced gradient noise, because less coefficients have to be adapted. These SA-HGMs perform very similar to the complete HGMs for the investigated real recordings stemming from a smartphone in different acoustic environments, and are therefore very suitable for practical low-cost nonlinear AEC applications, such as mobile devices with hard processing-power or power constraints. This is possible because the method focuses only on significant nonlinearities and observes them from time lags of significant input-output energy transmission.

6. REFERENCES

[1] M. M. Sondhi, "An adaptive echo canceller," *Bell System Technical Journal*, vol. 46, no. 3, pp. 497–511, 1967.

[2] C. Breining *et al.*, "Acoustic echo control. an application of very-high-order adaptive filters," *IEEE Signal Processing Magazine*, vol. 16, no. 4, pp. 42–69, July 1999.

[3] S. Haykin, *Adaptive Filter Theory*, 4th ed. Upper Saddle River (NJ), USA: Prentice Hall, 2002.

[4] J. Benesty, M. M. Sondhi, and Y. Huang, *Springer handbook of speech processing*. Springer, 2008.

[5] S. Makino and Y. Kaneda, "Acoustic echo canceller algorithm based on the variation characteristics of a room impulse response," in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 2, 1990, pp. 1133–1136.

[6] J. Benesty and S. L. Gay, "An improved PNLMS algorithm," in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 2, May 2002, pp. 1881–1884.

[7] D. Duttweiler, "Proportionate normalized least-mean-squares adaptation in echo cancelers," *IEEE Transactions on Speech and Audio Processing*, vol. 8, no. 5, pp. 508–518, September 2000.

[8] Y. Chen, Y. Gu, and A. Hero, "Sparse LMS for system identification," in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, April 2009, pp. 3125–3128.

[9] A. Birkett and R. Goubran, "Limitations of handsfree acoustic echo cancellers due to nonlinear loudspeaker distortion and enclosure vibration effects," in *IEEE ASSP Workshop on Applications of Signal Processing to Audio and Acoustics*, October 1995, pp. 103–106.

[10] O. Hoshuyama and A. Sugiyama, "An acoustic echo suppressor based on a frequency-domain model of highly nonlinear residual echo," in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 5, May 2006, pp. 269 – 272.

[11] O. Hoshuyama, "An update algorithm for frequency-domain correlation model in a nonlinear echo suppressor," in *International Workshop on Acoustic Signal Enhancement (IWAENC)*, Aachen, Germany, September 2012.

[12] A. Schwarz, C. Hofmann, and W. Kellermann, "Spectral feature-based nonlinear residual echo suppression," in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, October 2013.

[13] A. Stenger, L. Trautmann, and R. Rabenstein, "Nonlinear acoustic echo cancellation with 2nd order adaptive Volterra filters," in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, no. 0. Phoenix, USA: IEEE, 1999, pp. 877–880.

[14] V. J. Mathews and G. L. Sicuranza, *Polynomial Signal Processing*. Wiley, 2000.

[15] W. A. Frank, "An efficient approximation to the quadratic Volterra filter and its application in real-time loudspeaker linearization," *Signal Processing*, vol. 45, no. 1, pp. 97 – 113, 1995.

[16] F. Küch and W. Kellermann, "Partitioned block frequency-domain adaptive second-order Volterra filter," *IEEE Transactions on Signal Processing*, vol. 53, no. 2, pp. 564–575, 2005.

[17] M. Zeller, L. Azpicueta-Ruiz, J. Arenas-Garcia, and W. Kellermann, "Adaptive Volterra filters with evolutionary quadratic kernels using a combination scheme for memory control," *IEEE Transactions on Signal Processing*, vol. 59, no. 4, pp. 1449–1464, April 2011.

[18] M. Zeller and W. Kellermann, "Evolutionary adaptive filtering based on competing filter structures," in *Proceedings of the European Signal Processing Conference (EUSIPCO)*, Barcelona, Spain, August 2011, pp. 1264–1268.

[19] F. Küch, A. Mitnacht, and W. Kellermann, "Nonlinear acoustic echo cancellation using adaptive orthogonalized power filters," in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 3, March 2005, pp. 105– 108.

[20] S. Malik and G. Enzner, "Fourier expansion of Hammerstein models for nonlinear acoustic system identification," in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, May 2011, pp. 85–88.

[21] D. Comminiello *et al.*, "Functional link adaptive filters for nonlinear acoustic echo cancellation," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 21, no. 7, pp. 1502–1512, 2013.

[22] A. Stenger and W. Kellermann, "Adaptation of a memoryless preprocessor for nonlinear acoustic echo cancelling," *Signal Processing*, vol. 80, no. 9, pp. 1747 – 1760, 2000.

[23] S. Shimauchi and Y. Haneda, "Nonlinear acoustic echo cancellation based on piecewise linear approximation with amplitude threshold decomposition," in *International Workshop on Acoustic Signal Enhancement (IWAENC)*, Aachen, Germany, September 2012.

[24] S. Saito, A. Nakagawa, and Y. Haneda, "Dynamic impulse response model for nonlinear acoustic system and its application to acoustic echo canceller," in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, October 2009, pp. 201–204.

[25] A. Birkett and R. Goubran, "Acoustic echo cancellation using NLMS-neural network structures," in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 5, May 1995, pp. 3035–3038.

[26] A. Birkett and R. Goubran, "Nonlinear loudspeaker compensation for hands free acoustic echo cancellation," *Electronics Letters*, vol. 32, no. 12, pp. 1063–1064, June 1996.

[27] L. Azpicueta-Ruiz *et al.*, "Enhanced adaptive Volterra filtering by automatic attenuation of memory regions and its application to acoustic echo cancellation," *IEEE Transactions on Signal Processing*, vol. 61, no. 11, pp. 2745–2750, 2013.

[28] I. S. Gradshteyn and I. M. Ryzik, *Table of Integrals, Series, and Products*, 7th ed., A. Jeffrey and D. Zwillinger, Eds. Elsevier, February 2007.