RESEARCH   **Open Access**

CrossMark

Significance-aware filtering for nonlinear acoustic echo cancellation

Christian Hofmann[*], Christian Huemmer, Michael Guenther and Walter Kellermann

Abstract

This article summarizes and extends the recently proposed concept of Significance-Aware (SA) filtering for nonlinear acoustic echo cancellation. The core idea of SA filtering is to decompose the estimation of the nonlinear echo path into beneficially interacting subsystems, each of which can be adapted with high computational efficiency. The previously proposed SA Hammerstein Group Models (SA-HGMs) decompose the nonlinear acoustic echo path into a direct-path part, modeled by a Hammerstein Group Model (HGM) and a complementary part, modeled by a very efficient Hammerstein model. In this article, we furthermore propose a novel Equalization-based SA (ESA) structure, where the echo path is equalized by a linear filter to allow for an estimation of the loudspeaker nonlinearities by very small and efficient models. Additionally, we provide a novel in-depth analysis of the computational complexity of the previously proposed SA and the novel ESA filters and compare both SA filtering approaches to each other, to adaptive HGMs, and to linear filters, where fast partitioned-block frequency-domain realizations of the competing filter structures are considered. Finally, the echo reduction performance of the proposed SA filtering approaches is verified using real recordings from a commercially available smartphone. Beyond the scope of previous publications on SA-HGMs, the ability of the SA filters to generalize for double-talk situations is explicitly considered as well. The low complexity as well as the good echo reduction performance of both SA filters illustrate the potential of SA filtering in practice.

Keywords: Significance-aware, Equalization-based significance-aware, Efficiency, Nonlinear acoustic echo cancellation, Hammerstein group model, Hammerstein model

1 Introduction

Since the first adaptive linear echo canceler for network echoes in telephone lines [1], linear echo cancellation has evolved to a key ingredient of almost any full-duplex speech communication system. This has resulted in a multitude of approaches to efficiently model, parametrize, and estimate even complex linear systems, such as the acoustic echo paths in hands-free wideband telecommunication scenarios [2]. With increasingly common nonlinear distortions produced by miniaturized amplifiers and loudspeakers in modern portable devices, dedicated nonlinear echo path models have emerged as an important topic of research and motivated sophisticated approaches for nonlinear Acoustic Echo Cancellation (AEC) based on Volterra filters [3–5], artificial neural networks [6, 7], Functional Link Adaptive Filters (FLAFs) [8, 9], or kernel

*Correspondence: hofmann@LNT.de
[1]Multimedia Communications and Signal Processing, Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), Cauerstraße 7, Erlangen, Germany

methods [10, 11]. A very simple, yet effective model for nonlinear acoustic echo paths is the cascade of a memoryless preprocessor (modeling loudspeaker signal distortions) and a subsequent linear system (modeling sound propagation through air) [12]. Due to its simplicity, this so-called Hammerstein Model (HM) has been frequently employed [13–22] and will also be used in this contribution. So will be a group of *B* parallel HMs, referred to as Hammerstein Group Model (HGM) in the following. The recently proposed efficient Significance-Aware HGM (SA-HGM) [20] combines the advantages of HMs and HGMs and was extended to an efficient partitioned-block frequency-domain realization in [22].

Beyond previous work, this article introduces a novel variant of the Significance-Aware (SA) filtering concept denoted as Equalization-based SA (ESA) filtering which complements the existing efficient frequency-domain realization in [22]. Thereby, highly efficient SA filters can be derived for higher-order nonlinear systems—even without block partitioning. Furthermore, a novel in-depth

analysis of the computational complexity of the previously proposed SA filtering concepts is provided and contrasted with the computational complexity of the novel ESA filtering concept, adaptive HGMs, and conventional adaptive linear filters. Beyond previous investigations [22], this article also assesses the ability of the considered echo path models to generalize for double-talk situations, in which the models cannot be adapted to the current input signal. The performance in such situations reflects how well the estimated system models the physical system to be identified.

The remainder of this paper is structured as follows: after introducing the notation in Section 2, frequently used echo path models and their adaptation are reviewed in Sections 3.1 and 3.2, respectively. Afterwards, the recently proposed Partitioned-Block SA-HGM (PBSA-HGM) is summarized in Section 4.1, before introducing the novel ESA filtering concept in Section 4.2, which results in an Equalization-based Significance-Aware HM (ESA-HM). Then, the computational complexity of the SA filters (PBSA-HGM and ESA-HM) is analyzed and compared to the complexity for adapting a linear model and an HGM in Section 5.1. An experimental verification of the efficacy of the novel ESA-HM in comparison to other approaches in terms of echo reduction performance is given in Section 5.2. Finally, the manuscript is complemented by a summary of the main results in Section 6.

2   Notation

Throughout this article, vectors will be typeset in boldface lowercase letters, e.g., $\mathbf{a}$, and matrices in uppercase boldface, e.g., $\mathbf{A}$. The complex conjugate, transposed, and Hermitian-transposed of a vector $\mathbf{a}$ (column vector by default) will be written as $\mathbf{a}^*$, $\mathbf{a}^{\mathrm{T}}$, and $\mathbf{a}^{\mathrm{H}}$, respectively. Besides, $\mathbf{A} \odot \mathbf{B}$, $\mathbf{A} \oslash \mathbf{B}$, and $|\mathbf{A}|^2$ denote element-wise multiplication (Hadamard product), element-wise division, and element-wise magnitude-squaring of matrices or vectors of the same size, respectively. Furthermore, $\langle \mathbf{a}, \mathbf{b} \rangle$ stands for the scalar product $\mathbf{a}^{\mathrm{H}} \mathbf{b}$. Special matrices are the $M \times M$ identity matrix $\mathbf{I}_M$, the $M \times M$ all-zero matrix $\mathbf{0}_M$, and the windowing matrices

$$\mathbf{W}_{01} = \begin{bmatrix} \mathbf{0}_M & \mathbf{0}_M \\ \mathbf{0}_M & \mathbf{I}_M \end{bmatrix} \quad \text{and} \quad \mathbf{W}_{10} = \begin{bmatrix} \mathbf{I}_M & \mathbf{0}_M \\ \mathbf{0}_M & \mathbf{0}_M \end{bmatrix}, \quad (1)$$

setting the first or second half of a length-$2M$ vector to zero, respectively. Furthermore, $\mathbf{F}$ and $\mathbf{F}^{\mathrm{H}}$ denote the Nth-order Discrete Fourier Transform (DFT) matrix and its inverse, respectively. Besides, $a(k) * b(k)$ and $a(k) \circledast b(k)$ denote linear and cyclic convolution between time series $a(k)$ and $b(k)$, respectively, where k is the discrete-time sample index. Analogously to vectors, $\langle a(k), b(k) \rangle = \sum_{k=-\infty}^{\infty} a(k) b^*(k)$ denotes the scalar product between the time series $a(k)$ and $b(k)$. For a real-valued scalar a, $\lceil a \rceil$

refers to the smallest integer number larger than or equal to a, also known as ceiling function.

## 3   Fundamentals of linear and nonlinear acoustic echo cancellation

The acoustic echo path of a full-duplex communication system can be described as illustrated in Fig. 1.

Therein, the discrete-time loudspeaker signal $x(k)$, where k is the sample index, is interpolated by the D/A converter, amplified, and played-back via a loudspeaker, yielding the analog loudspeaker signal $x_{\mathrm{a}}(t)$, where t is the continuous time variable. The subsequent propagation of the sound waves through the medium air to the microphone can be modeled very accurately by a linear system with impulse response $h_{\mathrm{a}}(t)$. These acoustic echoes (far-end components) superimpose with local speakers and interferences (near-end components), evoking the continuous-time microphone signal $y_{\mathrm{a}}(t)$ and its sampled, discrete-time representation $y(k)$. In order to provide good estimates of local (near-end) speakers, two technologies are typically employed: an AEC unit for removing the far-end signal components from $y(k)$ and a postfilter suppressing residual echoes and near-end interferences [23–32].

In this work, we will not address the latter but focus on the AEC unit, which provides an echo estimate $\hat{y}(k)$ by identifying the acoustic echo path adaptively. The AEC error signal

$$e(k) = y(k) - \hat{y}(k)$$

is an estimate for the near-end signal, even during double-talk periods, where both far-end and near-end components are present. In periods of vanishing near-end signal (single-talk periods), $e(k)$ can be used for refining the echo path estimate. Thus, the AEC unit typically requires a double-talk detection [33–38]. As double-talk detection
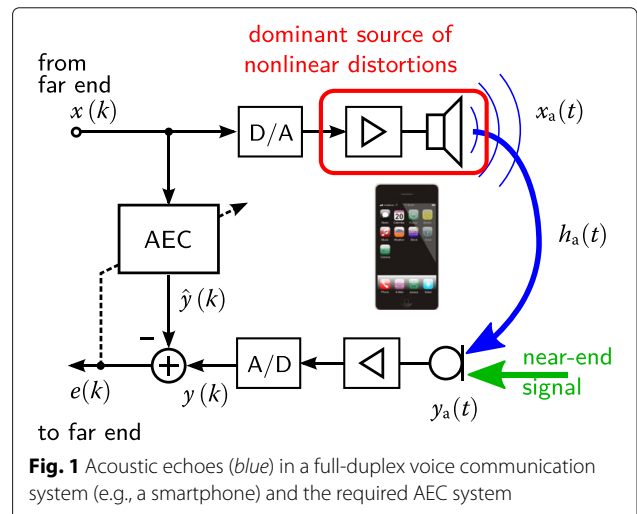
Fig. 1 Acoustic echoes (*blue*) in a full-duplex voice communication system (e.g., a smartphone) and the required AEC system

Hofmann *et al. EURASIP Journal on Advances in Signal Processing* (2016) 2016:113

Page 3 of 18

is out of the scope of this article, we will assume this to be handled separately, such that the microphone signal during time intervals used for echo path estimation only consists of far-end signal components (acoustic echoes). Furthermore, the AEC unit requires the choice of a suitable echo path model and of an adaptation strategy. This will be addressed in Sections 3.1 and 3.2, respectively.

3.1 From linear to nonlinear echo path models

This section contains a brief overview of frequently employed echo path models. Strategies for estimating the model parameters (i.e., filter coefficients) will be described in Section 3.2.

3.1.1 Linear models

The most simple, yet frequently used echo path model is the linear model, depicted in Fig. 2a.

Therein, the Loudspeaker-Enclosure-Microphone System (LEMS) is modeled by a linear FIR filter with the input/output relation

$$y(k) = x(k) * h(k), \tag{2}$$

where $h(k)$ is the discrete-time impulse response of the LEMS. Such models are most suitable for high-quality audio equipment. On the other hand, energy-efficient and miniaturized portable devices operating at the limit of their capabilities (such as portable navigation devices or

a

$x(k)$ $h(k)$ $y(k)$

Linear echo path model

b

$x(k)$ $f\{\cdot\}$ $x_{\mathrm{NL}}(k)$ $h(k)$ $y(k)$

Hammerstein Model (HM)

c

$x(k)$ $\sum_{b=1}^{B} w_b \cdot f_b\{\cdot\}$ $x_{\mathrm{NL}}(k)$ $h(k)$ $y(k)$

Hammerstein Model (HM) with parametric preprocessor

d

$x(k)$ $f_1\{\cdot\}$ $x_1(k)$ $h_1(k)$ $y(k)$

$f_2\{\cdot\}$ $x_2(k)$ $h_2(k)$

$f_B\{\cdot\}$ $x_B(k)$ $h_B(k)$

Hammerstein Group Model (HGM)

Fig. 2 a–d Frequently used echo path models for AEC, where each of the red blocks represents a fixed number of coefficients of a linear Finite Impulse Response (FIR) filter

smartphones in hands-free mode) lead to nonlinear distortions in the played-back signal [12] (see red box in Fig. 1), which render linear echo path models insufficient.

3.1.2 Hammerstein models

Due to the cascaded nature of the acoustic echo path between $x(k)$ and $y(k)$ (first nonlinear playback equipment, then linear transmission through air), a very simple, yet effective nonlinear echo path model is given by the cascade of a memoryless nonlinearity (memoryless preprocessor) with a subsequent linear system [12–22]. Such a structure is depicted in Fig. 2b and is referred to as HM or nonlinear-linear cascade in the literature [39–43]. This corresponds to the input/output relation

$$\begin{aligned} y(k) &= h(k) * x_{\mathrm{NL}}(k) \\ &= h(k) * f\{x(k)\}. \end{aligned} \tag{3}$$

In AEC, the memoryless preprocessor $x_{\mathrm{NL}}(k) = f\{x(k)\}$ typically has to model a saturation introduced by the amplifier and the loudspeaker. To this end, $f\{x(k)\}$ can be chosen from a simple hard limiter to a parametric preprocessor [14]

$$x_{\mathrm{NL}}(k) = \sum_{b=1}^{B} w_b f_b\{x(k)\} \tag{4}$$

with nonlinearity basis functions $f_b\{\cdot\}$ and preprocessor weights (parameters) w_b, which corresponds to the block diagram in Fig. 2c. Considering a Bth-order polynomial preprocessor, intermodulation products up to the order B may be modeled. Examples for such polynomial preprocessors can be found in [13, 14].

3.1.3 Hammerstein group models

A more complex and more general class of models will be referred to as HGMs. They consist, as depicted in Fig. 2d, of a group of B parallel HMs and have individual impulse responses $h_b(k)$ (referred to as kernels) for each branch $b = 1, \ldots, B$ after the branch's nonlinearity basis function $f_b\{\cdot\}$. This corresponds to the input/output relation

$$y(k) = \sum_{b=1}^{B} h_b(k) * \underbrace{f_b\{x(k)\}}_{x_b(k)}, \tag{5}$$

where $x_b(k)$ will be referred to as branch signals in the following.

Practical examples for such HGMs employ, e.g., monomes as nonlinearity basis functions [44, 45] (so-called power filters) and Legendre polynomials [19–22], as well as sinusoidal functions [8, 9, 42]. Some of the publications employing sinusoidal functions also refer to HGMs as FLAF without memory [8, 9]. Inspired by machine learning, FLAFs typically express the input/output relation of (5) differently: instead of individual branch signals, all branch signals are interleaved and concatenated to a
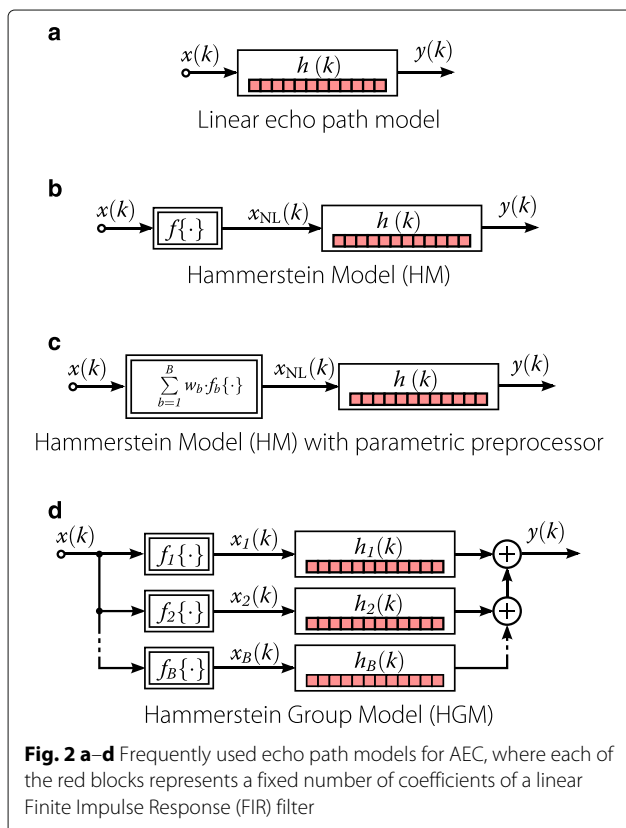
larger expanded vector, which is mapped to an echo estimate by linear combination (covers both convolution and summation in (5)). However, with focus on the applicability of fast convolution methods in this article, the formulation according to a group of parallel Hammerstein systems is advantageous.

Furthermore, note that the special case of an HGM where all $h_b(k) = w_b h_{b_{\mathrm{ref}}}(k) \forall b = 1, \ldots, B$ are scaled versions of a reference impulse response $h_{b_{\mathrm{ref}}}(k)$ in branch b_{ref} can be expressed as an HM. One possible HM representation of such an HGM has the reference impulse response $h(k) = h_{b_{\mathrm{ref}}}(k)$ and a parametric preprocessor according to (4) with the HGM's nonlinearity basis functions $f_b\{\cdot\}$ and preprocessor weights

$$w_b = \frac{\langle h_{b_{\mathrm{ref}}}(k), h_b(k) \rangle}{\langle h_{b_{\mathrm{ref}}}(k), h_{b_{\mathrm{ref}}}(k) \rangle}. \tag{6}$$

This correspondence will be essential for the SA filters in Section 4, where $b_{\mathrm{ref}} = 1$ (implying $w_1 = 1$) will be considered by default.

3.1.4 Models with dynamic nonlinearities

A further generalization of HGMs are models, where the nonlinearity basis functions are nonlinear functions $f_b\{\mathbf{x}(k)\}$ depending on a vector $\mathbf{x}(k)$, which is formed from samples of the input signal $x(k)$. The most prominent example for such filters are the so-called Volterra filters [3–5, 40, 46–48], for which $f_b\{\mathbf{x}(k)\}$ computes time-lagged products of samples of $x(k)$ (different elements of $\mathbf{x}(k)$)[1]. Note that power filters [44, 45] (having a memoryless nonlinearity) represent the special case of the Volterra filters where only the main diagonal of each Volterra kernel is populated with non-zero coefficients, corresponding to the respective kernels of the power filter. Volterra filters can be seen as a multidimensional Taylor series expansion of the function $\mathbf{x}(k) \mapsto y(k)$. Alternatively, the monomial basis functions of Volterra filters have also been replaced by Legendre polynomials [49, 50] or Fourier basis functions (sinusoids) [51]. Note that for the Fourier-basis nonlinear filters, the feed-forward structure has also been complemented by a feedback path, leading to bounded-input/bounded-output (BIBO)-stable recursive nonlinear filters [52].

Although the SA filters described in Section 4 will employ only HGMs and memoryless preprocessors, it is worth noticing that the SA filtering concept can also be applied to dynamic nonlinearities $f_b\{\mathbf{x}(k)\}$, involving, e.g., Volterra filters instead of HGMs.

3.2 Adaptation strategies

Adaptation of nonlinear models can be performed on different levels. On a basic level, parameters like filter coefficients of a model with a given structure are identified.

To this end, the parameters may be modeled as deterministic parameters, resulting in, e.g., Least-Mean-Square (LMS) algorithms [53] for adaptation, or as probabilistic parameters, leading to Kalman filter-like algorithms [15, 19, 21, 54]. On a higher level, the model structure (e.g., filter lengths of linear subsystems or numbers of diagonals of Volterra kernels) can be estimated via self-configuring evolutionary algorithms [5, 55–57]. In this article, the term "adaptation" will refer to the iterative estimation of filter coefficients modeled as deterministic parameters of a model with fixed structure.

In this section, the adaptation of the linear subsystems of the models described in Section 3.1 will be discussed. As all these models can be expressed as special cases of a parallel structure like an HGM, the adaptation will be described for an HGM. Then, the adaptation schemes for all other models can be derived as special cases of the HGM case. In particular, two common approaches for filter adaptation will be revisited: the direct adaptation of the impulse responses in the time domain by a Normalized Least-Mean-Square (NLMS) algorithm and the adaptation of partitioned versions of the impulse responses in the frequency domain, which allows a smooth trade-off between computational efficiency and algorithmic delay (latency) of the digital signal processing system.

3.2.1 Time-domain adaptive filtering

A straightforward identification of the linear filters of an HGM is possible by iteratively minimizing a quadratic cost function derived from the error signal $e(k)$. Most common algorithms are LMS-, affine projection-, or Recursive Least-Squares (RLS)-type algorithms [53]. Due to its computational efficiency, the adaptation of an HGM by an NLMS algorithm, which aims at minimizing the Mean Squared Error (MSE) in a gradient-descent manner, will be considered. The following description is independent of the actual choice of the nonlinearity basis functions $f_b\{\cdot\}$, as long as $f_b\{\cdot\}$ are time-invariant (Legendre polynomials will be employed for the evaluations in Section 5).

An adaptive HGM with such time-invariant nonlinearity basis functions $f_b\{\cdot\}$ first computes the branch signals

$$x_b(k) = f_b\{x(k)\} \tag{7}$$

and afterwards the echo estimate

$$\hat{y}(k) = \sum_{b=1}^{B} \sum_{l=0}^{L-1} x_b(k-l) \cdot \hat{h}_{b,l}(k-1), \tag{8}$$

where $\hat{h}_{b,l}(k-1)$ is the lth tap of a length-L estimate of the impulse response in branch b and has been obtained at time index $k-1$. Then, the error signal

$$e(k) = y(k) - \hat{y}(k) \tag{9}$$

has to be computed, before performing for each branch b and filter tap l an update according to

$$\hat{h}_{b,l}(k) = \hat{h}_{b,l}(k-1) + \mu_b \cdot \frac{e(k)}{E_{x_b}(k) + \delta} x_b(k-l), \quad (10)$$

where $0 < \mu_b < 1$ are branch-specific adaptation step-sizes, $E_{x_b}(k) = \sum_{l=0}^{L-1} |x_b(k-l)|^2$ are the branch signal energies, and δ is a regularization constant for numerical stability[2]. An efficient implementation (as assumed for the complexity analysis in Section 5.1) decomposes (10) by determining branch powers

$$P_{x_b}(k) = (x_b(k))^2 \quad (11)$$

to compute the branch energies

$$E_{x_b}(k) = E_{x_b}(k-1) + P_{x_b}(k) - P_{x_b}(k-L), \quad (12)$$

computes normalized errors

$$e_{\mathrm{norm},b}(k) = \mu_b e(k) \big/ \left(E_{x_b}(k) + \delta \right), \quad (13)$$

and finally updates the filter coefficients according to

$$\hat{h}_{b,l}(k) = \hat{h}_{b,l}(k-1) + e_{\mathrm{norm},b}(k) x_b(k-l). \quad (14)$$

This allows to adaptively identify the individual branches of HGMs but also covers the identification of single HMs and linear models as special cases.

3.2.2 Partitioned-block frequency-domain adaptive filtering

Partitioned-block convolution: A linear convolution can be realized despite large filter lengths L with a low input/output delay efficiently by block-based processing methods like Partitioned-Block Frequency-Domain Filtering (PBFDF) [56, 58–60], also known as multidelay convolution [61]. In the following, only a uniform partitioning with frame shift M and frame size $N = 2M$ will be considered.

In this case, the input signal $x(k)$, the impulse response $h(k)$, and the output signal $y(k)$ are partitioned into length-N vectors

$$\mathbf{x}(\kappa) = [x(\kappa M - N + 1), \ldots, x(\kappa M)]^{\mathrm{T}} \quad (15)$$

$$\mathbf{h}^{(p)} = \left[h(pM), \ldots, h(pM + M - 1), 0, \ldots, 0 \right]^{\mathrm{T}} \quad (16)$$

$$\mathbf{y}(\kappa) = \left[0, \ldots, 0, y(\kappa M - M + 1), \ldots, y(\kappa M) \right]^{\mathrm{T}}, \quad (17)$$

respectively, where κ is the frame index for block processing and p is the index of the impulse response partition. After a DFT, represented by the DFT matrix $\mathbf{F}$, the transformed versions of the signal vectors and impulse response partitions are referred to as

$$\underline{\mathbf{x}}(\kappa) = \mathbf{F}\mathbf{x}(\kappa) \quad (18)$$

$$\underline{\mathbf{h}}^{(p)} = \mathbf{F}\mathbf{h}^{(p)}. \quad (19)$$

As illustrated in Fig. 3 for $P = 2$ partitions, performing fast DFT-domain convolution between each pair of $\mathbf{h}^{(p)}$
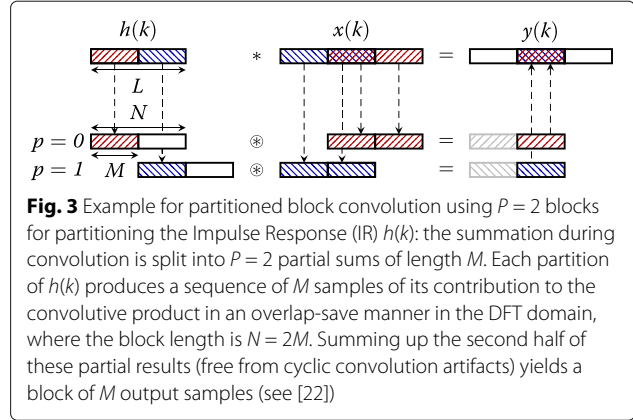
Fig. 3 Example for partitioned block convolution using $P = 2$ blocks for partitioning the Impulse Response (IR) $h(k)$: the summation during convolution is split into $P = 2$ partial sums of length M. Each partition of $h(k)$ produces a sequence of M samples of its contribution to the convolutive product in an overlap-save manner in the DFT domain, where the block length is $N = 2M$. Summing up the second half of these partial results (free from cyclic convolution artifacts) yields a block of M output samples (see [22])

and $\mathbf{x}(\kappa - p)$ and summing up the respective partial results yields

$$\mathbf{y}(\kappa) = \mathbf{W}_{01} \overbrace{\mathbf{F}^{\mathrm{H}} \underbrace{\sum_{p=0}^{P-1} \underline{\mathbf{x}}(\kappa - p) \odot \underline{\mathbf{h}}^{(p)}}_{\underline{\mathbf{y}}^{\circ}(\kappa)}}^{\mathbf{y}^{\circ}(\kappa)}, \quad (20)$$

where $P = \left\lceil \frac{L}{M} \right\rceil$ is the number of non-zero impulse response partitions and where the windowing matrix $\mathbf{W}_{01}$ according to (1) suppresses additionally computed samples in $\mathbf{y}^{\circ}(\kappa)$, which result from the previous frame and may contain cyclic convolution artifacts.

Note that such a PBFDF scheme is computationally efficient because each input signal frame's DFT has to be computed only once.

Partitioned-block frequency-domain adaptive filtering (PBFDAF): In the following, partitioned-block convolution will be utilized during the adaptation of an HGM with DFT-domain filter estimates $\hat{\underline{\mathbf{h}}}_b^{(p)}(\kappa)$ for partition p in branch b at time frame κ. As for the time-domain description in Section 3.2.1, we will assume that the non-linearity basis functions $f_b \{\cdot\}$ are chosen in advance (e.g., as monomials or Legendre polynomials)—the validity of the following section is independent of the actual choice. Iteratively minimizing the MSE by adapting $\hat{\underline{\mathbf{h}}}_b^{(p)}(\kappa)$ in a gradient-descent manner with instantaneous estimates for the gradient can be realized by the NLMS algorithm. All computations required for modeling a Partitioned-Block HGM (PB-HGM) and adapting it via a Frequency-domain Normalized Least-Mean-Square (FNLMS) algorithm can be summarized as follows: first, the set of new samples of each branch signal $x_b(k)$ has to be computed and appended to the already available samples of the branch signal for a given frame (for overlap-save processing), yielding the branch signal vectors

Hofmann *et al. EURASIP Journal on Advances in Signal Processing* (2016) 2016:113

Page 6 of 18

$$\mathbf{x}_b(\kappa) = \begin{bmatrix} [\mathbf{0}_M, \mathbf{I}_M]\, \mathbf{x}_b(\kappa - 1) \\ [f_b\{x(\kappa)\}, \ldots, f_b\{x(\kappa + M - 1)\}]^{\mathrm{T}} \end{bmatrix}, \tag{21}$$

which are then transformed into the DFT domain according to

$$\underline{\mathbf{x}}_b(\kappa) = \mathbf{F}\mathbf{x}_b(\kappa). \tag{22}$$

This allows the computation of the DFT-domain error signal vector

$$\underline{\mathbf{e}}(\kappa) = \mathbf{F}\left(\mathbf{y}(\kappa) - \underbrace{\mathbf{W}_{01}\mathbf{F}^{\mathrm{H}}\hat{\underline{\mathbf{y}}}^{\circ}(\kappa)}_{\hat{\mathbf{y}}(\kappa)}\right), \tag{23}$$

where, analogously to (20),

$$\hat{\underline{\mathbf{y}}}^{\circ}(\kappa) = \sum_{b=1}^{B} \sum_{p=0}^{P-1} \underline{\mathbf{x}}_b(\kappa - p) \odot \hat{\underline{\mathbf{h}}}_b^{(p)}(\kappa - 1) \tag{24}$$

is the intermediate DFT-domain microphone signal estimate with cyclic convolution artifacts of the partitioned-block convolution and $\hat{\mathbf{y}}(\kappa)$ is the final time-domain estimate containing M zeros and the M most recent samples of $\hat{y}(k)$. Similar to the time-domain adaptation, branch signal Power Spectral Density (PSD) vectors

$$\underline{\mathbf{s}}_{x_b}(\kappa) = \gamma_{\mathrm{E}}\underline{\mathbf{s}}_{x_b}(\kappa - 1) + (1 - \gamma_{\mathrm{E}})\left|\underline{\mathbf{x}}_b(\kappa)\right|^2 \tag{25}$$

are calculated in the DFT domain by recursively smoothing the element-wise squared magnitudes $\left|\underline{\mathbf{x}}_b(\kappa)\right|^2$ with $0 \leq \gamma_{\mathrm{E}} < 1$. In the following, $\gamma_{\mathrm{E}} = 0.9$ will be employed as default. Based on $\underline{\mathbf{s}}_{x_b}(\kappa)$, normalized branch signals are computed according to

$$\underline{\mathbf{x}}_{\mathrm{norm},b}(\kappa) = \left(\underline{\mathbf{x}}_b(\kappa)\right)^* \odot \left(\mu_b \oslash \left(\underline{\mathbf{s}}_{x_b}(\kappa) + \delta\right)\right), \tag{26}$$

before performing the actual filter update according to

$$\tilde{\underline{\mathbf{h}}}_b^{(p)}(\kappa) = \hat{\underline{\mathbf{h}}}_b^{(p)}(\kappa - 1) + \underline{\mathbf{x}}_{\mathrm{norm},b}(\kappa - p) \odot \underline{\mathbf{e}}(\kappa). \tag{27}$$

Additionally, the temporal support of the filters can be limited explicitly to the partition length M

$$\hat{\underline{\mathbf{h}}}_b^{(p)}(\kappa) = \mathbf{F}\,\underbrace{\mathbf{W}_{10}\mathbf{F}^{\mathrm{H}}\left(\tilde{\underline{\mathbf{h}}}_b^{(p)}(\kappa)\right)}_{\hat{h}_b^{(p)}(\kappa)}. \tag{28}$$

This is equivalent to the classical formulation of the so-called constrained update, where the constraint (zeros in time domain) is imposed on the update (see [61]). Yet, the computation of $\hat{\mathbf{h}}_b^{(p)}(\kappa)$ as a byproduct of the filter constraint in (28) will be beneficial for the SA filtering later on.

As for time-domain adaptive filtering, the algorithm for adapting an HGM in the partitioned-block frequency domain covers the adaptation of an HM (an HGM with $B = 1$ and $f_1\{x(k)\} = f\{x(k)\}$) and a linear model (an

HGM with $B = 1$ and $x_1(k) = f_1\{x(k)\} = f\{x(k)\} = x(k)$). Furthermore, the non-partitioned FNLMS algorithm, derived in many textbooks, e.g., [53, 62], results from the aforementioned partitioned-block description for $P = 1$ with sufficiently large $N = 2L$. Adapting a model with the block partitioning will be referred to as Partitioned-Block Frequency-domain Normalized Least Mean Squares (PB-FNLMS) algorithm.

4 Significance-aware filtering
Significance-aware (SA) filtering is a generalized system identification concept which exploits prior knowledge about the physical system to be identified by decomposing the adaptive model in a divide-and-conquer manner into beneficially interacting subsystems.

In particular, the originally proposed SA-HGM [20] performs a temporal decomposition of the acoustic echo path, as depicted in Fig. 4 for a linear model, into a region describing the direct-path wavefront and a complementary region.

This decomposition is then employed, as depicted in Fig. 5, to model the short direct-path region (carrying a significant amount of energy) by an HGM and to model the large complementary region by a computationally efficient HM. Apart from the decomposition into subsystems, the interaction between the subsystems is a key feature of SA filtering as well. In particular, the HM subsystem's nonlinear preprocessor has the form

$$x_{\mathrm{pp}}(k) = f_{\mathrm{pp}}\{x(k)\} = \sum_{b=1}^{B} \hat{w}_b(\kappa)\, x_b(k) \tag{29}$$

and combines the branch signals $x_b(k)$ with frame-wise updated weights $\hat{w}_b(\kappa)$. These weights $\hat{w}_b(\kappa)$ are estimated from the kernels of the HGM subsystem (will be explained in Section 4.1.3). Thereby, a nonlinear model estimated from the direct path is extrapolated to the entire acoustic echo path.

These two features, the decomposition and the preprocessor coefficient estimation via an HGM, are key components of both the recently proposed PBSA-HGM [22], revisited in Section 4.1, and the novel alternative SA
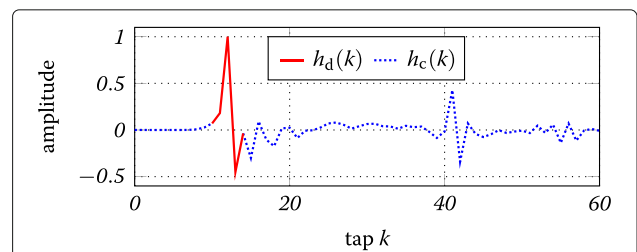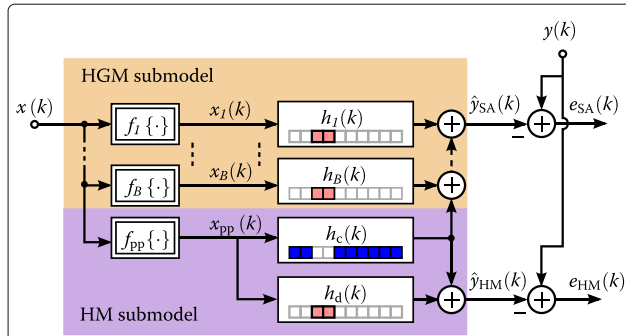

Fig. 4 Exemplary splitting of the Room Impulse Response (RIR) for SA filtering into a direct-path (*red*) and a complementary (*blue*) region

Hofmann *et al. EURASIP Journal on Advances in Signal Processing* (2016) 2016:113

Page 7 of 18

Fig. 5 SA-HGM for efficient nonlinear system identification: the temporal support modeled by the filters is color-coded by the *squares* below the IR symbols h_*. *White squares* indicate unmodeled temporal support and *red* and *blue squares* represent modeled support corresponding to the direct path and the complementary region, respectively. Typically, $\hat{y}_{\mathrm{SA}}(k)$ is employed as final echo estimate

filtering structure, which will be denoted by ESA filtering and will be introduced in Section 4.2.

4.1 Significance-aware Hammerstein group models

SA-HGMs are efficient because the potentially computationally expensive HGM only has to model a small temporal support of the acoustic echo path. To realize this concept while exploiting the benefit of fast frequency-domain convolution, a partitioned-block variant of the SA-HGMs, denoted as PBSA-HGM, has recently been introduced in [22][3]. This PBSA-HGM will be briefly reviewed in the following. Employing the uniform partitioning according to Section 3.2.2 leads to an adaptive HM with DFT-domain partition estimates $\hat{\underline{\mathbf{h}}}^{(p)}(\kappa)$ for partitions $p = 0, \ldots, P-1$ and an adaptive HGM with DFT-domain estimates $\hat{\underline{\mathbf{h}}}_b^{(p_\mathrm{d})}(\kappa)$ for branches $b = 1, \ldots, B$ but just a single partition with index p_d. This partition with index p_d should capture the direct path and thus a significant portion of the energy transmitted from the loudspeaker to the microphone. Note that $\hat{\underline{\mathbf{h}}}^{(p_\mathrm{d})}(\kappa)$ models $h_\mathrm{d}(k)$ in Fig. 5 and $\hat{\underline{\mathbf{h}}}^{(p)}(\kappa) \forall p \neq p_\mathrm{d}$ jointly model $h_\mathrm{c}(k)$ in Fig. 5. In the following, the estimation of $\hat{\underline{\mathbf{h}}}^{(p)}(\kappa)$ (covering both $h_\mathrm{d}(k)$ and $h_\mathrm{c}(k)$ of Fig. 5) and $\hat{\underline{\mathbf{h}}}_b^{(p_\mathrm{d})}(\kappa)$ will be described in Sections 4.1.1 and 4.1.2, respectively, and the estimation of the HM's preprocessor coefficients $\hat{w}_b(\kappa)$ based on $\hat{\underline{\mathbf{h}}}_b^{(p_\mathrm{d})}(\kappa)$ will be explained in Section 4.1.3.

4.1.1 Estimation of the RIR of the HM submodel

In the following, the identification of the HM subsystem's impulse response partitions will be described. For this purpose, the preprocessed loudspeaker signal can be expressed in vector notation as

$$\mathbf{x}_{\mathrm{pp}}(\kappa) = \begin{bmatrix} [\mathbf{0}_M, \mathbf{I}_M]\, \mathbf{x}_{\mathrm{pp}}(\kappa-1) \\ \sum_{b=1}^{B} \hat{w}_b(\kappa)[\mathbf{0}_M, \mathbf{I}_M]\, \mathbf{x}_b(\kappa) \end{bmatrix}, \quad (30)$$

where $\mathbf{x}_b(\kappa)$ are the branch signal vectors of the HGM submodel. The HM submodel with DFT-domain partition estimates $\hat{\underline{\mathbf{h}}}^{(p)}(\kappa)$ yields, analogously to (23), an echo signal estimate vector

$$\hat{\mathbf{y}}_{\mathrm{HM}}(\kappa) = \mathbf{W}_{01}\mathbf{F}^{\mathrm{H}} \sum_{p=0}^{P-1} \underline{\mathbf{x}}_{\mathrm{pp}}(\kappa-p) \odot \hat{\underline{\mathbf{h}}}^{(p)}(\kappa-1), \quad (31)$$

where $\underline{\mathbf{x}}_{\mathrm{pp}}(\kappa) = \mathbf{F}\mathbf{x}_{\mathrm{pp}}(\kappa)$ with $\mathbf{x}_{\mathrm{pp}}(\kappa)$ from (30).

The HM submodel's DFT-domain error vector can then be computed via

$$\underline{\mathbf{e}}_{\mathrm{HM}}(\kappa) = \mathbf{F}\left(\mathbf{y}(\kappa) - \hat{\mathbf{y}}_{\mathrm{HM}}(\kappa)\right). \quad (32)$$

Iteratively minimizing the two-norm of $\underline{\mathbf{e}}_{\mathrm{HM}}(\kappa)$ in a gradient-descent manner by an FNLMS algorithm leads to the update rule

$$\hat{\underline{\mathbf{h}}}^{(p)}(\kappa) = \mathbf{F}\mathbf{W}_{10}\mathbf{F}^{\mathrm{H}}\Big(\hat{\underline{\mathbf{h}}}^{(p)}(\kappa-1) \\ + \underline{\mathbf{x}}_{\mathrm{norm,pp}}(\kappa-p) \odot \underline{\mathbf{e}}_{\mathrm{HM}}(\kappa)\Big), \quad (33)$$

where $\underline{\mathbf{x}}_{\mathrm{norm,pp}}(\kappa-p)$ is the normalized DFT-domain signal vector computed analogously to (26) from the preprocessed input $\underline{\mathbf{x}}_{\mathrm{pp}}(\kappa)$.

4.1.2 Estimating the HGM submodel

The HGM submodel alone provides an echo estimate

$$\hat{\mathbf{y}}_{\mathrm{HGM}}(\kappa) = \mathbf{W}_{01}\mathbf{F}^{\mathrm{H}}\left(\sum_{b=1}^{B} \underline{\mathbf{x}}_b(\kappa-p_\mathrm{d}) \odot \hat{\underline{\mathbf{h}}}_b^{(p_\mathrm{d})}(\kappa-1)\right), \quad (34)$$

which is combined with the complementary-region partitions' echo estimate of the HM submodel according to

$$\hat{\mathbf{y}}_{\mathrm{SA}}(\kappa) = \mathbf{W}_{01}\mathbf{F}^{\mathrm{H}}\left(\sum_{b=1}^{B} \underline{\mathbf{x}}_b(\kappa-p_\mathrm{d}) \odot \hat{\underline{\mathbf{h}}}_b^{(p_\mathrm{d})}(\kappa-1) \\ + \sum_{p \in \{0,\ldots,P-1\}\backslash p_\mathrm{d}} \underline{\mathbf{x}}_{\mathrm{pp}}(\kappa-p) \odot \hat{\underline{\mathbf{h}}}^{(p)}(\kappa-1)\right). \quad (35)$$

Minimizing the two-norm of the error signal vector

$$\underline{\mathbf{e}}_{\mathrm{SA}}(\kappa) = \mathbf{F}\left(\mathbf{y}(\kappa) - \hat{\mathbf{y}}_{\mathrm{SA}}(\kappa)\right) \quad (36)$$

by an FNLMS algorithm leads to the update rule

$$\hat{\underline{\mathbf{h}}}_b^{(p_\mathrm{d})}(\kappa) = \mathbf{F}\mathbf{W}_{10}\mathbf{F}^{\mathrm{H}}\Big(\hat{\underline{\mathbf{h}}}_b^{(p_\mathrm{d})}(\kappa-1) \\ + \underline{\mathbf{x}}_{\mathrm{norm},b}(\kappa-p_\mathrm{d}) \odot \underline{\mathbf{e}}_{\mathrm{SA}}(\kappa)\Big), \quad (37)$$

where $\underline{\mathbf{x}}_{\mathrm{norm},b}(\kappa)$ is the normalized DFT-domain signal vector computed analogously to (26). The reason for the application of the windowing in (37) to the actual impulse

Hofmann *et al. EURASIP Journal on Advances in Signal Processing* (2016) 2016:113

Page 8 of 18

response partitions will become obvious in the following section.

4.1.3 Estimating the preprocessor of the HM

Finally, the HM's preprocessor coefficients can be recomputed employing the HGM. To avoid overhead due to complex-valued arithmetic, a direct use of the time-domain kernels $\hat{\mathbf{h}}_b^{(p_d)}(\kappa)$ (obtained as a byproduct of (37)) is possible to compute instantaneous least squares estimates for the preprocessor weights [20] according to

$$\tilde{w}_b(\kappa) = C_{b,1}(\kappa) \big/ C_{1,1}(\kappa) \tag{38}$$

with

$$C_{b,1}(\kappa) = \left\langle \hat{\mathbf{h}}_b^{(p_d)}(\kappa), \hat{\mathbf{h}}_1^{(p_d)}(\kappa) \right\rangle. \tag{39}$$

A subsequent temporal smoothing of these estimates leads to the final preprocessor coefficients

$$\hat{w}_b(\kappa + 1) = \gamma_w \hat{w}_b(\kappa) + (1 - \gamma_w)\tilde{w}_b(\kappa), \tag{40}$$

where $0 \leq \gamma_w < 1$. In the following, $\gamma_w = 0.95$ will be employed per default. Note that $\hat{w}_1(\kappa) = 1 \, \forall \kappa$ and therefore does not need to be computed at all. As in [20, 22], the first branch ($b = 1$) will be assumed to be linear ($f_1\{x(t)\} = x(t)$), such that an entirely linear echo path model results from a preprocessor with $\hat{w}_b(\kappa) = 0 \, \forall b > 1$.

By the method described in Section 4.1 and its subsections, the estimation of the nonlinear system has been split into two beneficially interacting subproblems (HM and HGM adaptation). The beneficial interaction is achieved by the preprocessor coefficient refinement based on the HGM and by the extension of the temporal support of the HGM employing partitions of the HM. An in-depth evaluation of the computational complexity of such a PBSA-HGM will be given in Section 5.1.

4.2 Equalization-based significance-aware Hammerstein models

Previous applications of the SA concept [20–22] split the echo path model along the time axis (using knowledge about a dominating direct-path component), to allow for an efficient estimation of nonlinear parameters (see also Section 4.1). This corresponded to a parallel decomposition of the acoustic echo path (see $h_c(k)$ and $h_d(k)$ of the HM submodel block in Fig. 5, where $f_{pp}\{\cdot\}$ can be shifted in the parallel branches of $h_c(k)$ and $h_d(k)$ as well).

In this section, we propose a novel realization of the SA filtering concept. This realization employs a serial decomposition (a cascade) of the echo path into a nonlinear loudspeaker and a subsequent linear RIR, in order to estimate the nonlinear parameters of an HM (which is used

as actual echo path model). The resulting ESA-HM structure is depicted in Fig. 6. Therein, the topmost branch contains the actual echo path, which is assumed to have Hammerstein structure. Below, there are three colored blocks comprising the novel structure denoted as ESA-HM. While block $\boldsymbol{\alpha}$ computes the actual echo estimate for AEC by an HM, and blocks $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ facilitate the estimation of the nonlinear preprocessor $f_{pp}\{\cdot\}$ of the HM in block $\boldsymbol{\alpha}$. In the following, the ESA-HM will be explained in a bottom-up manner, starting with block $\boldsymbol{\gamma}$ and ending with $\boldsymbol{\alpha}$. To avoid redundancy, the employed adaptive filtering algorithms, all of which have been described in previous sections, are not entirely re-written for the particular filters to be estimated in this section but are referenced instead.

4.2.1 Equalization of the RIR (block γ)

In block $\boldsymbol{\gamma}$ (Fig. 6), the ESA-HM directly exploits the cascaded nature of the nonlinear echo path (memoryless nonlinearity $f\{\cdot\}$, followed by an impulse response $h(k)$): the IR $h(k)$ is equalized by an adaptive linear filter with coefficients $\hat{h}_{eq,l}(k)$ for tap l at time k. This yields a discrete-time estimate

$$\hat{x}_{NL}(k - L) = \sum_{l=0}^{L-1} \hat{h}_{eq,l}(k - 1) \cdot y(k - l) \tag{41}$$

of the delayed nonlinearly distorted loudspeaker signal $x_{NL}(k - L)$. The linear equalizer $\hat{h}_{eq,l}(k)$ is identified by a PB-FNLMS algorithm (see Section 3.2.2) using the error signal
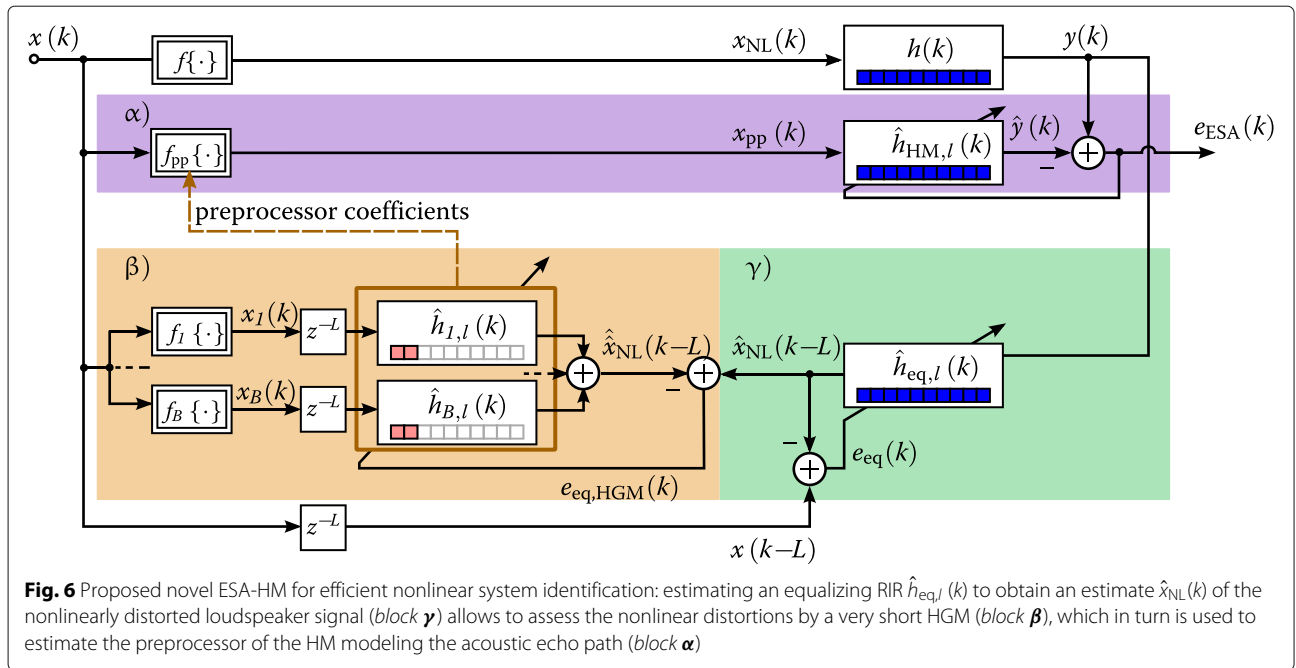
$$e_{eq}(k) = x(k - L) - \hat{x}_{NL}(k - L). \tag{42}$$

On the one hand, (41) suggests that $\hat{x}_{NL}(k-L)$ is an estimate for $x(k - L)$, obtained via linear filtering of $y(k)$. On the other hand, the linear equalizer $\hat{h}_{eq,l}(k)$ cannot equalize the nonlinear components of the LEMS, such that the nonlinear distortion remains in the equalizer output and $\hat{x}_{NL}(k-L)$ according to (41) can be seen as an estimate for the nonlinearly distorted signal $x_{NL}(k - L)$. Thus, while the inability of impulse responses to model nonlinear systems hampers the performance of the AEC system in the first place, this inability is exploited here to estimate an otherwise inaccessible intermediate signal to guide the estimation of the nonlinear components (described in the next paragraphs). Furthermore, the equalization is only as complex as a single linear AEC system (assuming identical filter lengths).

4.2.2 Estimating the nonlinearity (block β)

Note that both the RIR and the frequency response of the loudspeaker (the linear component of the loudspeaker) have been equalized by the inverse filtering

Fig. 6 Proposed novel ESA-HM for efficient nonlinear system identification: estimating an equalizing RIR $\hat{h}_{\text{eq},l}(k)$ to obtain an estimate $\hat{x}_{\text{NL}}(k)$ of the nonlinearly distorted loudspeaker signal (*block* $\boldsymbol{\gamma}$) allows to assess the nonlinear distortions by a very short HGM (*block* $\boldsymbol{\beta}$), which in turn is used to estimate the preprocessor of the HM modeling the acoustic echo path (*block* $\boldsymbol{\alpha}$)

in block $\boldsymbol{\gamma}$ (Fig. 6), such that $\hat{x}_{\text{NL}}(k - L)$ is time-aligned with $x(k - L)$. Thus, the nonlinear mapping between the discrete-time loudspeaker signal $x(k - L)$ and the estimate of the nonlinear loudspeaker signal $\hat{x}_{\text{NL}}(k - L)$ can be modeled by an adaptive time-domain HGM with branch kernels $\hat{h}_b(k)$ with a very small temporal support of L_{SA} taps (e.g., $L_{\text{SA}} = 3$). This corresponds to the input/output relation

$$\hat{\hat{x}}_{\text{NL}}(k - L) = \sum_{b=1}^{B} \sum_{l=0}^{L_{\text{SA}}-1} x(k - L - l) \cdot \hat{h}_{b,l}(k - 1). \quad (43)$$

The adaptation of $\hat{h}_{b,l}(k)$ is performed by a time-domain NLMS algorithm (see Section 3.2.1) operating on the error signal

$$e_{\text{eq,HGM}}(k) = \hat{x}_{\text{NL}}(k - L) - \hat{\hat{x}}_{\text{NL}}(k - L). \quad (44)$$

Due to the small number of modeled taps L_{SA}, such a time-domain algorithm is computationally inexpensive.

4.2.3 *Estimating the preprocessor of the HM (between blocks β and α)*

Analogously to the previously proposed SA filtering concept, the parameters of the adaptive HGM in block $\boldsymbol{\beta}$ in Fig. 6 (branch kernels $\hat{h}_{b,l}(k)$) can be employed to determine the coefficients of the nonlinear preprocessor of the HM in block $\boldsymbol{\alpha}$ in Fig. 6. Using a preprocessor with the structure of (29), this is done analogously to Eqs. (38) and (39) on a frame-wise basis (every M samples) according to

$$\tilde{w}_b(\kappa) = C_{b,1}(\kappa) \big/ C_{1,1}(\kappa) \quad (45)$$

with

$$C_{b,1}(\kappa) = \sum_{l=0}^{L_{\text{SA}}-1} \hat{h}_{b,l}(\kappa \cdot M) \hat{h}_{1,l}(\kappa \cdot M), \quad (46)$$

with the filter taps $\hat{h}_{b,l}(k)$ from the end of time frame κ, where $k = \kappa \cdot M$. As in (40), a subsequent temporal smoothing of these estimates leads to the final preprocessor coefficients

$$\hat{w}_b(\kappa + 1) = \gamma_{\text{w}} \hat{w}_b(\kappa) + (1 - \gamma_{\text{w}}) \tilde{w}_b(\kappa), \quad (47)$$

where $0 \leq \gamma_{\text{w}} < 1$.

4.2.4 *Computation of the echo estimate*

The HM's linear subsystem computes the echo estimate in block $\boldsymbol{\alpha}$ in Fig. 6 similar as in (3) by convolving $x_{\text{pp}}(k)$ with an impulse response estimate $\hat{h}_{\text{HM},l}(k)$ according to

$$\hat{y}(k) = \sum_{l=0}^{L-1} x_{\text{pp}}(k - l) \hat{h}_{\text{HM},l}(k - 1). \quad (48)$$

For computational efficiency, $\hat{h}_{\text{HM},l}(k)$ is estimated adaptively by a PB-FNLMS algorithm (see Section 3.2.2) operating on the error signal

$$e_{\text{ESA}}(k) = y(k) - \hat{y}(k). \quad (49)$$

The overall computational complexity of this algorithm is slightly higher than two adaptive linear filters (originating from blocks $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$). A detailed analysis of the computational complexity of this novel algorithm and of previously proposed algorithms will be given in the following in Section 5.1.

4.3   Structural comparison of SA and ESA filtering

In the following, the similarities and differences between the previously proposed SA filtering and the novel ESA filtering will be summarized. Note that in this section, the notation of time-domain filters will be employed, despite their potential realization as block-partitioned adaptive filters in the frequency domain.

Similarities Both the SA-HGM and ESA-HM decompose the acoustic echo path into subsystems to facilitate the efficient estimation of an HGM to estimate the nonlinear components of the echo path. Furthermore, both the SA-HGM and ESA-HM estimate the preprocessor coefficients of an HM from identified HGM kernels as a least squares estimate and use this preprocessor in an HM with a long subsequent impulse response. Therefore, both structures are considered as different variants of SA filtering.

Differences Despite their similarities, the previously proposed SA filtering and the novel ESA filtering employ a fundamentally different mechanism for the estimation of the nonlinearity by an HGM with small temporal support. As depicted in Fig. 7a, classical SA filtering employs a parallel decomposition of the echo path into a direct-path component and a complementary-path component. Canceling the complementary-part component of the echo signal yields the direct-path component of the echo path,

which is identified efficiently by an HGM for the SA-HGM (recall Section 4.1). As opposed to this, ESA filtering does not rely on any temporal decomposition of the echo path, such that it may be advantageous for non-acoustic system identification, where a dominant direct-path peak does not exist. Instead, the cascade of the Hammerstein system (a serial decomposition) is employed by augmenting the LEMS with an equalizer (recall block γ in Fig. 6), as depicted in Fig. 7b. The resulting overall system ideally consists of the first component of the Hammerstein echo path, namely the nonlinearity. This nonlinearity can be identified by an even shorter HGM than for the originally proposed SA filtering and does not require knowledge about any temporal structure of the Hammerstein-shaped echo path's linear subsystem. Another difference between the proposed SA and ESA filters is that the ESA-HM provides an echo path model with Hammerstein structure (output $e_{\text{ESA}}(k)$ in Fig. 6), while the SA-HGM allows to replace the direct-path region of an echo path model with Hammerstein structure by an HGM. This corresponds to the two SA-HGM outputs $e_{\text{HM}}(k)$ and $e_{\text{SA}}(k)$ in Fig. 5, respectively.
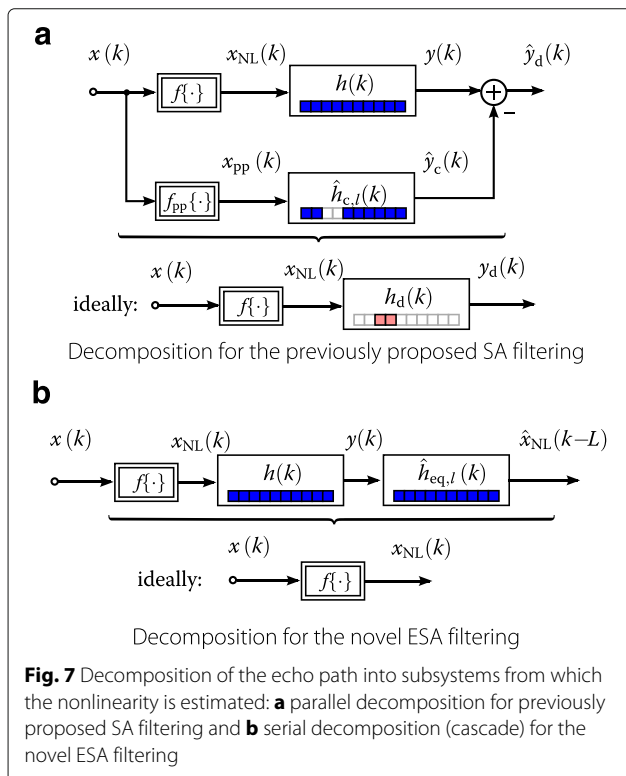
5   Evaluation

In this section, the novel and the previously proposed SA filtering concepts will be compared to each other and to classical adaptive filters in terms of computational complexity (Section 5.1) and echo cancellation performance (Section 5.2).

5.1   Computational complexity

This section contains an in-depth analysis of the computational complexity of the previously discussed adaptive filters. In practice, the actual computational load of an algorithm is determined by many platform-specific factors, such as the instruction set, the number of clock cycles for a particular arithmetic operation, pipelining, and caching abilities of the processor. Furthermore, especially with growing amount of data to be processed, the memory access pattern may significantly impact how well the processor exploits its capabilities or waits for new data from external memory. Nevertheless, the number of FLoating Point Operations (FLOPs) is still a commonly accepted indicator for the computational complexity of an algorithm on modern platforms, where the different arithmetic operations require similar execution time in principle. Operations counted as FLOPs will be real-valued multiplications (RMULs), additions (RADDs), and divisions (RDIVs).

As complex numbers typically have to be implemented by treating real and imaginary parts individually, complex-valued arithmetic operations are composed of multiple real-valued operations. In the following, we will assume the realization of a complex-valued multiplication

Fig. 7 Decomposition of the echo path into subsystems from which the nonlinearity is estimated: **a** parallel decomposition for previously proposed SA filtering and **b** serial decomposition (cascade) for the novel ESA filtering

Table 1 Computational effort for the identification of an HGM by a time-domain NLMS algorithm

Computed quantity	Equation	Multiplicity	Required operations					
			FFT	CMUL	CADD	RMUL	RADD	RDIV
Branch signals	(7)	$\forall b$	n/a	n/a	n/a	n/a	n/a	n/a
Output signal	(8)		0	0	0	BL	$B(L-1)$	0
Error signal	(9)		0	0	0	0	1	0
Input power	(11)	$\forall b$	0	0	0	B	0	0
Frame energy	(12)	$\forall b$	0	0	0	0	$2B$	0
Normalized error	(13)	$\forall b$	0	0	0	B	B	B
Filter coefficients	(14)	$\forall b, l$	0	0	0	BL	BL	0
		Accumulated:	0	0	0	$2B(L+1)$	$2B(L+1)+1$	B
					Total: $B(4L+5)+1$ FLOPs			

(CMUL) by four RMULs and two RADDs. Only for FFT algorithms, where operations performed on the twiddle factors can be cached, an alternative implementation with three RMULs and three RADDs will be considered [63]. Thereby, an Nth-order DFT for real-valued time-domain signals can be implemented by

$$\mathrm{RMUL}_{\mathrm{FFT}}(N) = \frac{1}{2}N\log_2 N - \frac{5}{4}N$$

RMULs and

$$\mathrm{RADD}_{\mathrm{FFT}}(N) = \frac{3}{2}N\log_2 N - \frac{1}{4}N - 4$$

RADDs [64].

Complexity of adaptive HGM Using these conventions, Table 1 lists the computational effort of the individual operations for the adaptation of an HGM in the time-domain (recall Section 3.2.1).

The first column contains the name of the computed quantity, the second column contains the corresponding equation number in this article, the third column

indicates how often the equation has to be evaluated, and the subsequent columns list the number of FFTs, complex-valued, and real-valued operations for the particular equation. Below this list, the accumulated number of individual operations and the total number of FLOPs are listed as well. Note that these operation counts are given per output sample for this time-domain algorithm. The frequency-domain algorithms considered in the following will be analyzed on a frame-wise basis.

Complexity of adaptive PB-HGM Analogously to Table 1, the operations for identifying a PB-HGM according to Section 3.2.2 in the frequency domain are listed in Table 2. Therein, $N_{\mathrm{N}} = \left\lceil \frac{N-1}{2} \right\rceil + 1$ denotes the number of non-redundant frequency bins (DC bin to Nyquist bin), for which signal processing is done—the remaining bins can be reconstructed due to the conjugate symmetry. Recall that the operations for a linear model are included in this analysis for the special case of $B = 1$. Furthermore, the case of a single-partition HGM results for $P = 1$.

Table 2 Computational effort for the identification of an HGM by a PB-FNLMS algorithm

Computed quantity	Equation	Multiplicity	Required operations					
			FFT	CMUL	CADD	RMUL	RADD	RDIV
Branch signals	(21)	$\forall b$	n/a	n/a	n/a	n/a	n/a	n/a
	(22)	$\forall b$	B	0	0	0	0	0
Output signal	(24)		0	BPN_{N}	$(BP-1)N_{\mathrm{N}}$	0	0	0
Error signal	(23)		2	0	0	0	$\frac{N}{2}$	0
Branch PSDs	(25)	$\forall b$	0	0	0	$4BN_{\mathrm{N}}$	$2BN_{\mathrm{N}}$	0
Normalized branch signals	(26)	$\forall b$	0	0	0	$2BN_{\mathrm{N}}$	BN_{N}	BN_{N}
Filter coefficients	(27)	$\forall b, p$	0	BPN_{N}	BPN_{N}	0	0	0
Filter constraint	(28)	$\forall b, p$	$2BP$	0	0	0	0	0
		Accumulated:	$B + 2BP + 2$	$2BPN_{\mathrm{N}}$	$N_{\mathrm{N}}(2BP-1)$	$6BN_{\mathrm{N}}$	$\frac{N}{2} + 3BN_{\mathrm{N}}$	BN_{N}
			Total: $(B(4P+2)+4)N\log_2 N + B\left(8P + \frac{7N}{2} + 5PN + 6\right) - \frac{7N}{2} - 10$ FLOPs					

Hofmann *et al. EURASIP Journal on Advances in Signal Processing*   (2016) 2016:113

Page 12 of 18

Table 3 Computational effort for the identification of a PBSA-HGM with a PB-FNLMS algorithm

Computed quantity	Equation	Multiplicity	Required operations					
			FFT	CMUL	CADD	RMUL	RADD	RDIV
Preprocessed input	(30)	$p = 0, \ldots, \frac{N}{2} - 1$	0	0	0	$(B-1)N/2$	$(B-1)N/2$	0
HM submodel	Table 2, $B=1$		$2P+3$	$2PN_N$	$N_N(2P-1)$	$6N_N$	$\frac{N}{2} + 3N_N$	N_N
HGM submodel	Table 2, $P=1$		$3B+2$	$2BN_N$	$N_N(2B-1) + N_N$	$6BN_N$	$\frac{N}{2} + 3BN_N$	BN_N
Kernel correlation	(39)	$\forall b$	0	0	0	$BN/2$	$B(N/2 - 1)$	0
Instantaneous weights	(38)	$\forall b > 1$	0	0	0	0	0	$B-1$
Smoothing weights	(40)	$\forall b > 1$	0	0	0	$2(B-1)$	$B-1$	0
Accumulated:			$3B+2P$ $+5$	$2N_N$ $\cdot(B+P)$	$N_N(2B$ $+2P-1)$	$2B - \frac{N}{2} +$ $6N_N + BN +$ $6BN_N - 2$	$\frac{N}{2} + 3N_N +$ $BN +$ $3BN_N - 1$	$B + N_N +$ $BN_N - 1$

Total: $(6B + 4P + 10)\, N \log_2 N + 8P - \frac{7N}{2} + B\left(\frac{21N}{2} + 17\right) + 5PN - 16$ FLOPs

Complexity of adaptive SA-HGM The two aforementioned PB-HGM special cases are building blocks for the complexity analysis of the PBSA-HGM (recall Section 4.1), which is listed in Table 3. The additional N_N additions for the HGM submodel stem from the fact that the HGM submodel's output has to be added to the complementary-part echo estimate obtained with the HM (recall (35)).
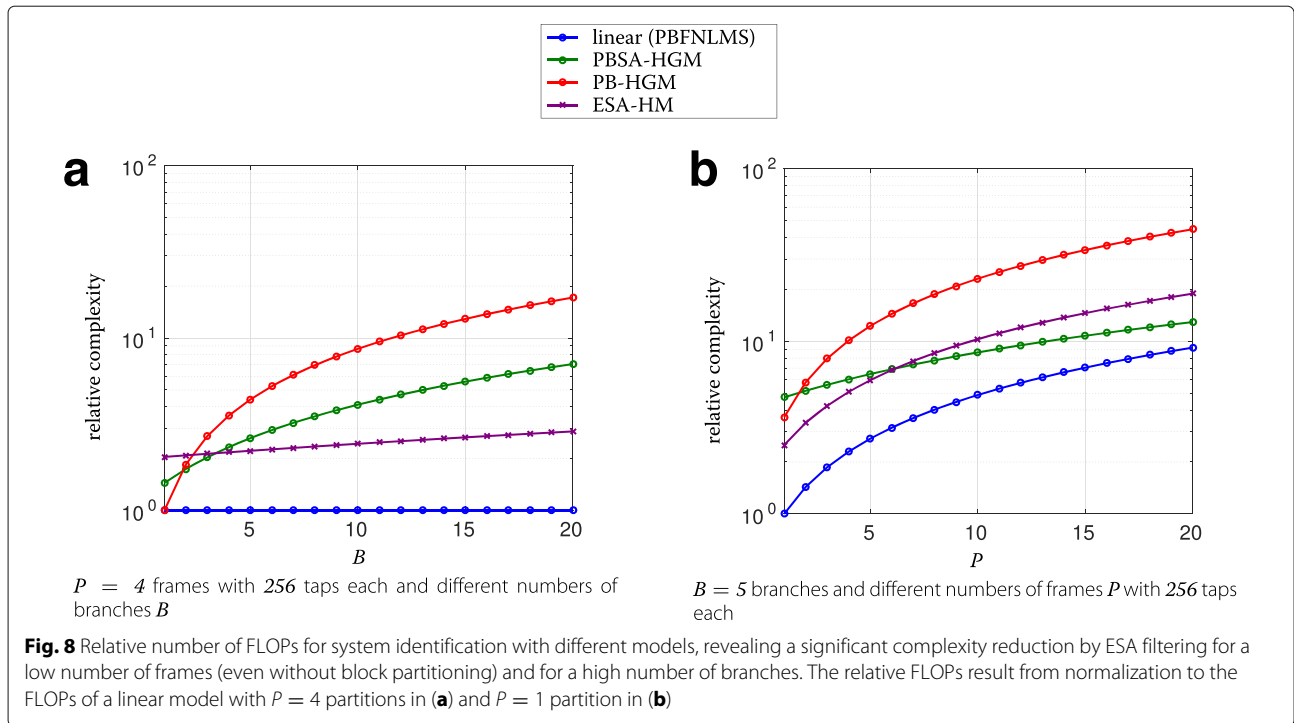
Complexity of adaptive ESA-HM In the same way, the operations for the novel ESA-HM are listed in Table 4. These operations consist of one linear frequency-domain adaptive filter for the HM and one for the echo path inversion, a very short time-domain HGM for assessing loudspeaker nonlinearities, determining the preprocessor weights from the HGM, and evaluating the preprocessor of the HM.

Comparison For consistency with the echo reduction performance evaluation later on, a frame shift of $M = 256$ taps, $B = 5$ branches, $P = 4$ frames, and only $L_{SA} = 3$ taps for the ESA-HM's HGM submodel are assumed as default. Varying the number of branches B of the nonlinear models and normalizing the resulting FLOPs to the FLOPs of a linear PB-FNLMS yields the relative FLOPs depicted in Fig. 8a.

Obviously, the ESA-HM has the constant offset of another linear filter for system inversion, which makes it unattractive in comparison to the previously proposed PBSA-HGM for only two or three branches. For two branches, even a full PB-HGM is marginally more efficient than an ESA-HM and only slightly more complex than a PBSA-HGM. However, with increasing number of branches, the benefit of an ESA-HM in comparison

Table 4 Computational effort for the identification of a ESA-HM with a PB-FNLMS algorithm

Computed quantity	Equation	Multiplicity	Required operations					
			FFT	CMUL	CADD	RMUL	RADD	RDIV
Preprocessed input	(30)	$p = 0, \ldots, \frac{N}{2} - 1$	0	0	0	$(B-1)N/2$	$(B-1)N/2$	0
HM submodel (forward)	Table 2, $B=1$		$2P+3$	$2PN_N$	$N_N(2P-1)$	$6N_N$	$\frac{N}{2} + 3N_N$	N_N
Inversion submodel	Table 2, $B=1$		$2P+3$	$2PN_N$	$N_N(2P-1)$	$6N_N$	$\frac{N}{2} + 3N_N$	N_N
Time-domain HGM submodel	Table 1, $L = L_{SA}$	$\frac{N}{2}$	0	0	0	$2B(L_{SA}+1)\frac{N}{2}$	$2B(L_{SA}+1)\frac{N}{2} + \frac{N}{2}$	$B\frac{N}{2}$
Kernel correlation	(46)	$\forall b$	0	0	0	BL_{SA}	$B(L_{SA} - 1)$	0
Instantaneous weights	(45)	$\forall b > 1$	0	0	0	0	0	$B-1$
Smoothing weights	(47)	$\forall b > 1$	0	0	0	$2(B-1)$	$B-1$	0
Accumulated:			$4P+6$	$4PN_N$	$N_N(4P-1)$	$2B + 12N_N +$ $\frac{N(B-1)}{2} + BL_{SA} +$ $BN(L_{SA}+1) - 2$	$N + 6N_N +$ $BL_{SA} + \frac{3BN}{2} +$ $BL_{SA}N - 1$	$B +$ $2N_N +$ $\frac{BN}{2} - 1$

Total: $(8P + 12)\, N \log_2 N + 16P + \frac{N}{2} + \left(2L_{SA} + \frac{7N}{2} + 2L_{SA}N + 3\right)B + 10PN - 10$ FLOPs

Hofmann *et al. EURASIP Journal on Advances in Signal Processing* (2016) 2016:113

Page 13 of 18

Fig. 8 Relative number of FLOPs for system identification with different models, revealing a significant complexity reduction by ESA filtering for a low number of frames (even without block partitioning) and for a high number of branches. The relative FLOPs result from normalization to the FLOPs of a linear model with $P = 4$ partitions in (**a**) and $P = 1$ partition in (**b**)

to a PB-HGM or even to the less complex PBSA-HGM becomes evident: an ESA-HM's complexity is mainly dominated by the two linear adaptive filters—even for $B = 20$ branches.

Varying the number of frames P and leaving $B = 5$ constant results in the relative FLOPs depicted in Fig. 8b, where relative FLOPs result from normalization to the FLOPs of an adaptive linear model with a single partition. As can be seen, the ESA-HM leads to a significantly reduced computational complexity for a low number of frames P. Thus, the ESA-HM always leads to a complexity reduction compared to the PB-HGM, even for $P = 1$, which corresponds to conventional frequency-domain adaptive filtering without block partitioning.

5.2 Echo reduction performance
In this section, the echo reduction performance of the novel ESA and the previously proposed SA filtering concept will be compared to classical adaptive filters based on linear models and HGMs.

5.2.1 Experimental setup
The following evaluation is based on double-talk free recordings of male and female speech, played-back and recorded with a smartphone in hands-free mode at very high volume. In total, five different physical setups, referred to as setups A to E, will be considered. The acoustically relevant characteristics of these setups are listed in Table 5.

Setups A and B correspond to recordings with a duration of about 130 s in a living room-like environment. Setups A and B differ in the placement of the smartphone within the same environment. Setup C corresponds to a recording in an anechoic environment and has a duration of about 80 s. The data for setups D and E are synthesized by convolving the anechoic nonlinear recordings obtained from setup C with measured RIRs. The employed RIRs are from a lab environment with variable acoustics, once with curtains at the walls (setup D) and once with open curtains (setup E). For all setups, a partitioned-block linear model (PB-HGM with a single linear branch), a full PB-HGM (full temporal support), a PBSA-HGM, and an ESA-HM will be compared.

In the following experiments, processing is done at a sampling rate of $f_S = 16$ kHz. The acoustic echo paths are estimated with FIR filters of length $L = 1024$ taps and the input signals and impulse responses are partitioned into $P = 4$ blocks with a relative frame shift of $M = 256$ taps

Table 5 Properties of setups A to E for the AEC experiments

	Room type	T_{60}	Room dimensions
Setup A	Living room	900 ms	$4 \times 5 \times 2.5$ m
Setup B	Living room	900 ms	$4 \times 5 \times 2.5$ m
Setup C	Anechoic chamber	n/a	$2 \times 2 \times 2$ m
Setup D	Lab environment with curtains on walls	250 ms	$5.5 \times 5.7 \times 3.1$ m
Setup E	Lab environment	400 ms	$5.5 \times 5.7 \times 3.1$ m

(implying $N = 512$). Note that, despite large values of T_{60} for some setups, the echo reduction performance in the experiments is not limited by the filter length but by the nonlinearity of the echo path (see Section 5.2.2 for a quantitative analysis of the amount of nonlinearity in the recordings).

As in [20], the HGMs consist of $B = 5$ branches, with odd-order Legendre polynomials up to order 9 as nonlinearity basis functions $f_b\{\cdot\}$ in (5). The algorithm parameters used in the experiments have been selected based on the recordings of setup C. In particular, a stepsize of $\mu_b = 0.1$ (see (26)) is used for the kernels of the PB-HGMs and a stepsize of $\mu_b = 0.2$ is used for the PBSA-HGMs and ESA-HMs. For the reference algorithm, consisting of the linear filter, a stepsize of $\mu_1 = 0.5$ is employed to account for the lack of multiple adaptation branches. Furthermore, the adaptation process is refined using the robust statistics described in [34, 59], which hardly affects the computational complexity but robustifies the system identification against outliers in the error signal. For the same purpose, the dynamics of the adaptive preprocessor weights is limited to a maximum stepsize of ± 0.001 per frame.

The AEC performance is quantified by computing the Echo-Return Loss Enhancement (ERLE) measure [62]

$$\text{ERLE} = 10 \log_{10} \left(\frac{\mathcal{E}\left\{y_{\text{echo}}^2(k)\right\}}{\mathcal{E}\left\{\left|y_{\text{echo}}(k) - \hat{y}(k)\right|^2\right\}} \right) \text{dB}, \quad (50)$$

where $y_{\text{echo}}(k)$ is the echo signal component of the microphone signal $y(k)$ and $\hat{y}(k)$ is the echo estimate obtained using the model to be evaluated (linear model, PB-HGM, PBSA-HGM, ESA-HM). For single talk, (50) becomes

$$\text{ERLE} = 10 \log_{10} \left(\frac{\mathcal{E}\left\{y^2(k)\right\}}{\mathcal{E}\left\{e^2(k)\right\}} \right) \text{dB}, \quad (51)$$

where the numerator is the power of the microphone signal $y(k)$ and the denominator is the power of the error signal $e(k)$ produced by the model to be evaluated. For a practical evaluation, the mathematical expectation (ensemble average) in (51) will be replaced by a time-averaging over the entire sequences. Note that the ERLE measure in its general definition of (50) does not depend on the near-end signal components (including the number and positions of the near-end sources) but requires knowledge of the actual far-end signal component during double-talk periods. This knowledge is not available in practice. Nonetheless, the ERLE performance during double-talk periods can be simulated by performing AEC on single-talk recordings with previously determined fixed filters (as will be done in Section 5.2.3 in experiment 2).

5.2.2 Quantification of nonlinearity

For monofrequent excitation signals, the nonlinearity of a system can easily be assessed by means of the Total Harmonic Distortion (THD), which measures the ratio of the energy of the excitation frequency to the energy of its harmonics. However, for nonstationary broadband excitation signals like speech, this measure is not meaningful anymore. In order to characterize the nonlinearity nevertheless, one can pursue noise-loading methods [65] or, as done in the following, methods based on the actual AEC excitation signal itself. To this end, consider the energy E_{in} which is recorded in the bandwidth excited by the digital loudspeaker signal $(0 - 8 \text{ kHz})$ and the energy E_{out} which is recorded above for frequencies exceeding the highest frequency contained in the excitation signal $(8 - 16 \text{ kHz})$. For a linear system, the out-of-band energy is only caused by noise. For a nonlinear system, E_{out} is significantly increased due to the harmonic distortion. The In-band/Out-of-band Ratio (IOR)

$$\text{IOR} = 10 \log_{10} \left(\frac{E_{\text{in}}}{E_{\text{out}}} \right) \text{dB}$$
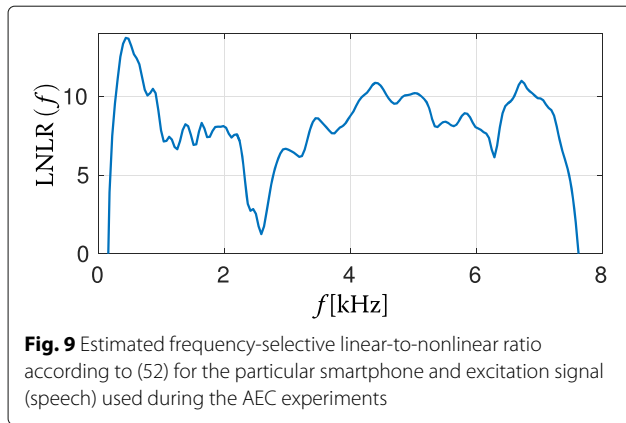
may therefore serve as a measure of the intensity of the nonlinearity in the echo signal.

Furthermore, the so-called Linear-to-Non-Linear Ratio (LNLR) is defined as the ratio between the power of the linear echo signal component and the power of the nonlinear echo signal component. Thus, a lower LNLR corresponds to stronger nonlinearities. While the THD is a description of how much nonlinearity a particular frequency causes at its harmonics, a frequency-selective LNLR quantifies the distortion present at a particular frequency[4]. Although an exact LNLR computation required exact knowledge of the linear and nonlinear system components to compute the respective signal components, the LNLR can be estimated using the echo estimate $\hat{y}(k)$ obtained by a converged long adaptive linear filter and the resulting error signal $e(k) = y(k) - \hat{y}(k)$. Employing the respective signal spectra estimates $S_{\hat{y}\hat{y}}(f)$ and $S_{ee}(f)$, the LNLR at a frequency f can be estimated as

$$\text{LNLR}(f) = 10 \log_{10} \left(\frac{S_{\hat{y}\hat{y}}(f)}{S_{ee}(f)} \right) \text{dB}. \quad (52)$$

The particular smartphone employed for the AEC experiments leads to an IOR of 27 dB for the AEC speech signal in an anechoic environment[5]. The LNLR estimated from the same measurement is depicted in Fig. 9 up to 8 kHz—for higher frequencies, the numerator of (52) is determined by noise only and, therefore, (52) cannot be employed to estimate the LNLR anymore.

As can be seen, the LNR (f) is strongly frequency-dependent and varies between 1.2 dB (at about 2.5 kHz) and 13.7 dB (at about 500 Hz). This minimum around 2.5 kHz and the maximum around 500 Hz coincide with

Fig. 9 Estimated frequency-selective linear-to-nonlinear ratio according to (52) for the particular smartphone and excitation signal (speech) used during the AEC experiments

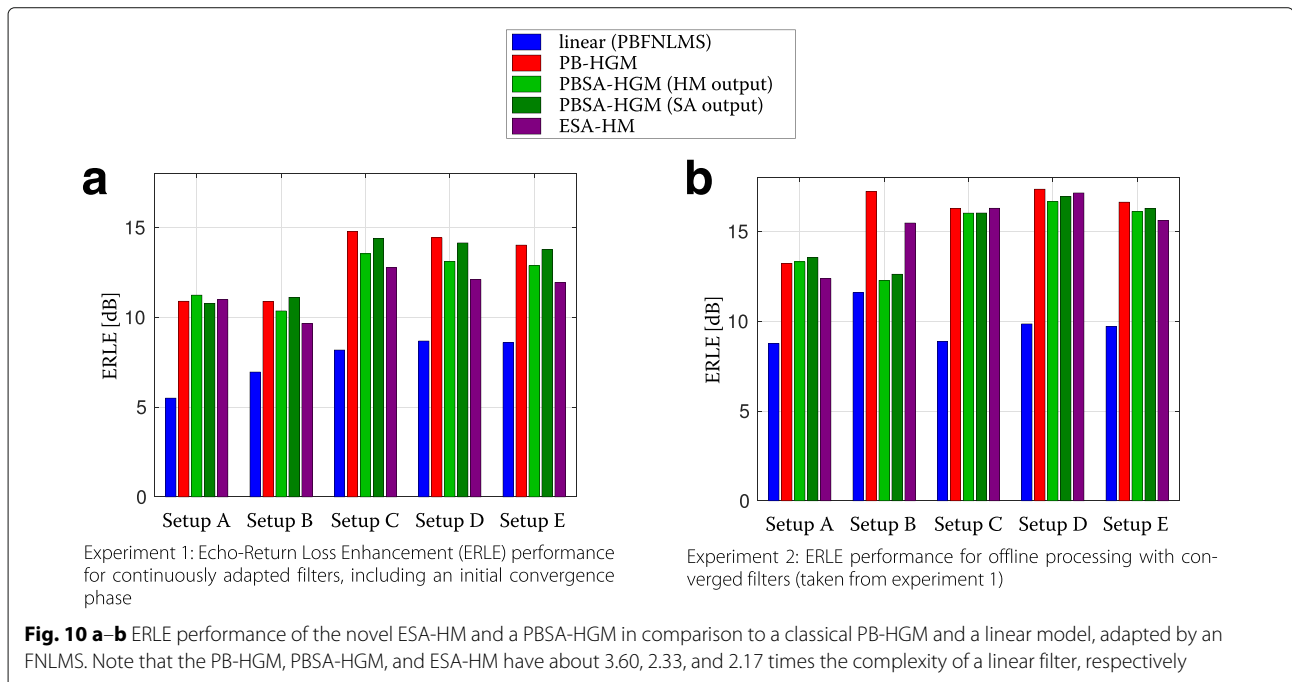a local minimum and the maximum of the far-end signal spectrum, respectively.

5.2.3   *Experimental results*

In the following, two experiments will be discussed. Experiment 1 considers an adaptive identification of the nonlinear systems according to Section 5.2.1, and experiment 2 considers an offline filtering without further adaptation with previously adapted models from the end of experiment 1. While experiment 1 is affected by the initial convergence phase, experiment 2 evaluates the performance achievable in double-talk situations after convergence of the filters, which is of vital importance for a full-duplex voice communication system as well.

Experiment 1 - continuously adapted filters Figure 10a shows the ERLE measure for different setups and models

for experiment 1. Note that the PBSA-HGM is listed twice: once with the previously considered output signal $e_{SA}(k)$ and once with the HM submodel's output $e_{HM}(k)$. Obviously, the linear filter performs worst for all setups. The other end of the scale is represented by the PB-HGM, which leads to an almost twice as high average ERLE value for setup C. The PBSA-HGM leads to a good approximation of the PB-HGM. In setups A and B, the PB-HGM with full temporal support is even marginally outperformed by the PBSA-HGM's HM and HGM outputs, respectively. On the other hand, when considering the HM submodel of the PBSA-HGM only, the average gap to the PB-HGM performance is larger. This is not surprising, as the classical PBSA-HGM with a peak-modeling HGM has more degrees of freedom and tracks the nonlinearities before providing a smoothed nonlinearity estimate to the HM submodel. The ESA-HM can also approximate the PB-HGM but leads to slightly less ERLE than the PBSA-HGM's HM output in experiment 1. Yet, the ESA-HM consistently outperforms the partitioned block linear filter. This clearly verifies the applicability of SA filtering in general and of the proposed novel ESA filtering concept for practically relevant scenarios with real-world smartphone recordings.

Experiment 2 - echo reduction during double talk In experiment 2 (depicted in Fig. 10b), when processing the entire sequence offline with the converged filters from the previous experiment, the ESA-HM performs as good as the full PB-HGM in setups C and D. Only in setups A and E, the PBSA-HGM can marginally outperform the

Experiment 1: Echo-Return Loss Enhancement (ERLE) performance for continuously adapted filters, including an initial convergence phase

Experiment 2: ERLE performance for offline processing with converged filters (taken from experiment 1)

Fig. 10 a–b ERLE performance of the novel ESA-HM and a PBSA-HGM in comparison to a classical PB-HGM and a linear model, adapted by an FNLMS. Note that the PB-HGM, PBSA-HGM, and ESA-HM have about 3.60, 2.33, and 2.17 times the complexity of a linear filter, respectively

Hofmann *et al. EURASIP Journal on Advances in Signal Processing* (2016) 2016:113

Page 16 of 18

ESA-HM in experiment 2. Overall, the ESA-HM generalizes slightly better for double-talk situations than the PBSA-HGM. Interestingly, the typically considered PBSA-HGM's HGM output and the PBSA-HGM's HM output do not differ significantly in experiment 2, which supports the initial assumption that an HM is a suitable efficient approximation of the echo path. The increased performance due to the PBSA-HGM's HGM submodel seems to originate from a quicker reaction to the instantaneous signal characteristics.

Clearly, the novel ESA-HM is a very well performing alternative to the PBSA-HGM, which leads to comparable or even better echo reduction in double-talk situations.

6 Conclusions

The adaptation of nonlinear echo paths for small portable devices requires efficient adaptive nonlinear echo path models. To this end, a novel variant of SA filtering has been introduced and compared to known concepts in this article. This novel ESA filtering method, leading to an adaptive ESA-HM, exploits the inability of IRs to model nonlinearities to obtain an estimate of the unobservable nonlinearly distorted loudspeaker signal by inverse filtering. While the previously proposed PBSA-HGM has been an efficient alternative to HGMs only for a block-partitioning of the filter, the novel ESA filtering concept is advantageous without block partitioning at all and for a very high number of branches (see complexity analyses in Section 5.1). For applications where very long RIRs need to be modeled and where a low input-output delay is required, a block partitioning and therefore the PBSA-HGM may computationally be more efficient than the ESA-HM. Both methods thereby complement each other very well in terms of computational efficiency for different application scenarios. A comparison of the echo reduction performance of the ESA-HM, the PBSA-HGM, a linear model, and a PB-HGM in Section 5.2 has emphasized the efficacy of the proposed ESA-HM, especially for double-talk situations, in which AEC is actually most important.

Endnotes

[1] Note that this representation of a Volterra filter is also referred to as diagonal coordinate representation [66].

[2] This update rule is also referred to as NLMS with kernel-specific normalization in [66].

[3] As opposed to time-domain adaptive filters, the complexity for filter adaptation in the frequency domain is not determined by the length of the time-domain support of the filter, but by the DFT size N and the number of partitions P. This disqualifies unpartitioned frequency-domain adaptive filters, as the HGM submodel in Fig. 5 would

have the same complexity as the HGM with full temporal support in Fig. 2d.

[4] Determining such a measure is also referred to as Schüßler's model in [67], as it goes back to [68, 69].

[5] Note that this number is caused by nonlinearity and not by background noise, as the SNR of the recorded signal is more than 24 dB in the $8 - 16$ kHz range, where E_{out} is computed.

References
1. MM Sondhi, An adaptive echo canceller. Bell Syst. Tech. J. **46**(3), 497–511 (1967)
2. P Dreiseitel, E Hänsler, H Puder, in *Proc. European Signal Process. Conf. (EUSIPCO)*. Acoustic echo and noise control—a long lasting challenge (EURASIP, Rhodes, 1998), pp. 945–952
3. EJ Thomas, Some considerations on the application of the Volterra representation of nonlinear networks to adaptive echo cancellers. Bell Syst. Tech. J. **50**(8), 2797–2805 (1971)
4. A Stenger, L Trautmann, R Rabenstein, in *Proc. IEEE Intl. Conf. Acoustics, Speech, Signal Process. (ICASSP)*. Nonlinear acoustic echo cancellation with 2nd order adaptive Volterra filters, vol. 2 (IEEE, Phoenix, 1999), pp. 877–880
5. M Zeller, LA Azpicueta–Ruiz, J Arenas-Garcia, W Kellermann, Adaptive Volterra filters with evolutionary quadratic kernels using a combination scheme for memory control. IEEE Trans. Signal Process. **59**(4), 1449–1464 (2011)
6. AN Birkett, RA Goubran, in *Proc. IEEE Workshop Neural Networks Signal Process. (NNSP)*. Nonlinear echo cancellation using a partial adaptive time delay neural network (IEEE, Cambridge, 1995), pp. 449–458
7. LSH Ngja, J Sjobert, in *Proc. IEEE Intl. Conf. Acoustics, Speech, Signal Process. (ICASSP)*. Nonlinear acoustic echo cancellation using a Hammerstein model (IEEE, Seattle, 1998), pp. 1229–1232
8. D Comminiello, M Scarpiniti, R Parisi, A Uncini, in *Proc. Int. Workshop Acoustic Echo, Noise Control (IWAENC)*. A functional link based nonlinear echo canceller exploiting sparsity (IWAENC, Tel Aviv, 2010)
9. D Comminiello, M Scarpiniti, LA Azpicueta-Ruiz, J Arenas-Garcia, A Uncini, Functional link adaptive filters for nonlinear acoustic echo cancellation. IEEE Trans. Audio Speech Lang. Process. **21**(7), 1502–1512 (2013)
10. G Li, C Wen, WX Zheng, Y Chen, Identification of a class of nonlinear autoregressive models with exogenous inputs based on kernel machines. IEEE Trans. Signal Process. **59**(5), 2146–2159 (2011)
11. J Kivinen, AJ Smola, RC Williamson, Online learning with kernels. IEEE Trans. Signal Process. **52**(8), 165–176 (2004)
12. AN Birkett, RA Goubran, in *Proc. IEEE Workshop Applications Signal Process. Audio Acoustics (WASPAA)*. Limitations of handsfree acoustic echo cancellers due to nonlinear loudspeaker distortion and enclosure vibration effects (IEEE, New Paltz, 1995), pp. 103–106
13. A Stenger, R Rabenstein, in *Proc. European Signal Process. Conf. (EUSIPCO)*. An acoustic echo canceller with compensation of nonlinearities, vol. 98 (EURASIP, Island of Rhodes, 1998), pp. 969–972
14. A Stenger, W Kellermann, Adaptation of a memoryless preprocessor for nonlinear acoustic echo cancelling. Signal Process. **80**(9), 1747–1760 (2000)
15. S Malik, G Enzner, State-space frequency-domain adaptive filtering for nonlinear acoustic echo cancellation. IEEE Trans. Audio Speech Lang. Process. **20**(7), 2065–2079 (2012)
16. S Shimauchi, Y Haneda, in *Proc. Int. Workshop Acoustic Echo, Noise Control (IWAENC)*. Nonlinear acoustic echo cancellation based on piecewise linear

Hofmann *et al. EURASIP Journal on Advances in Signal Processing* (2016) 2016:113

Page 17 of 18

approximation with amplitude threshold decomposition (VDE, Aachen, 2012), pp. 1–4

17. S Malik, G Enzner, in *Proc. IEEE Intl. Conf. Acoustics, Speech, Signal Process. (ICASSP)*. Variational Bayesian inference for nonlinear acoustic echo cancellation using adaptive cascade modeling (IEEE, Kyoto, 2012), pp. 37–40

18. S Malik, G Enzner, A variational Bayesian learning approach for nonlinear acoustic echo control. IEEE Trans. Signal Process. **61**(23), 5853–5867 (2013)

19. C Huemmer, C Hofmann, R Maas, A Schwarz, W Kellermann, in *Proc. IEEE Intl. Conf. Acoustics, Speech, Signal Process. (ICASSP)*. The elitist particle filter based on evolutionary strategies as novel approach for nonlinear acoustic echo cancellation (IEEE, Florence, 2014), pp. 1315–1319

20. C Hofmann, C Huemmer, W Kellermann, in *Proc. IEEE Intl. Conf. Acoustics, Speech, Signal Process. (ICASSP)*. Significance-aware Hammerstein group models for nonlinear acoustic echo cancellation (IEEE, Florence, 2014), pp. 5934–5938

21. C Huemmer, C Hofmann, R Maas, W Kellermann, in *Proc. IEEE Global Conf. Signal Information Process. (GlobalSIP)*. The significance-aware EPFES to estimate a memoryless preprocessor for nonlinear acoustic echo cancellation (IEEE, Atlanta, 2014), pp. 557–561

22. C Hofmann, M Guenther, C Huemmer, W Kellermann, in *Proc. European Signal Process. Conf. (EUSIPCO)*. Efficient nonlinear acoustic echo cancellation by partitioned-block significance-aware Hammerstein group models (EURASIP, Budapest, 2016)

23. S Boll, Suppression of acoustic noise in speech using spectral subtraction. IEEE Trans. Acoust. Speech Signal Process. **27**(2), 113–120 (1979)

24. S Gustafsson, R Martin, P Vary, Combined acoustic echo control and noise reduction for hands-free telephony. Signal Process. **64**(1), 21–32 (1998)

25. K Linhard, T Haulick, in *Proc. European Conf. Speech Communication and Technology (EUROSPEECH)*. Noise subtraction with parametric recursive gain curves (ISCA, Budapest, 1999), pp. 2611–2614

26. I Cohen, Optimal speech enhancement under signal presence uncertainty using log-spectral amplitude estimator. IEEE Signal Process. Lett. **9**(4), 113–116 (2002)

27. O Hoshuyama, A Sugiyama, in *Proc. IEEE Intl. Conf. Acoustics, Speech, Signal Process. (ICASSP)*. An acoustic echo suppressor based on a frequency-domain model of highly nonlinear residual echo, vol. 5 (IEEE, Toulouse, 2006), pp. 269–272

28. F Küch, W Kellermann, in *Proc. IEEE Intl. Conf. Acoustics, Speech, Signal Process. (ICASSP)*. Nonlinear residual echo suppression using a power filter model of the acoustic echo path, vol. I (IEEE, Honolulu, 2007), pp. 73–76

29. DA Bendersky, JW Stokes, HS Malvar, in *Proc. IEEE Intl. Conf. Acoustics, Speech, Signal Process. (ICASSP)*. Nonlinear residual acoustic echo suppression for high levels of harmonic distortion (IEEE, Las Vegas, 2008), pp. 261–264

30. O Hoshuyama, in *Proc. Int. Workshop Acoustic Echo, Noise Control (IWAENC)*. An update algorithm for frequency-domain correlation model in a nonlinear echo suppressor (VDE, Aachen, 2012)

31. A Schwarz, C Hofmann, W Kellermann, in *Proc. IEEE Workshop Applications Signal Process. Audio Acoustics (WASPAA)*. Spectral feature-based nonlinear residual echo suppression (IEEE, New Paltz, 2013)

32. A Schwarz, C Hofmann, W Kellermann, in *ITG Conf. Speech Commun*. Combined nonlinear echo cancellation and residual echo suppression (VDE, Erlangen, 2014)

33. DL Duttweiler, A twelve-channel digital echo canceler. IEEE Trans. Commun. **26**(5), 647–653 (1978)

34. T Gänsler, A double-talk resistant subband echo canceller. Signal Process. **65**(1), 89–101 (1998)

35. T Gänsler, J Benesty, A frequency-domain double-talk detector based on a normalized cross-correlation vector. Signal Process. **81**, 1783–1787 (2001)

36. H Buchner, J Benesty, T Gänsler, W Kellermann, Robust extended multidelay filter and double-talk detector for acoustic echo cancellation. IEEE Trans. Audio Speech Lang. Process. **14**(5), 1633–1644 (2006)

37. K Shi, X Ma, GT Zhou, in *Annual Conf. on Information Sciences and Systems (CISS)*. A mutual information based double-talk detector for nonlinear systems (IEEE, Princeton, 2008), pp. 356–360

38. M Schneider, EAP Habets, in *Proc. European Signal Process. Conf. (EUSIPCO)*. Comparison of multichannel doubletalk detectors for acoustic echo cancellation (EURASIP, Nice, 2015), pp. 300–304

39. IW Hunter, MJ Korenberg, The identification of nonlinear biological systems: Wiener and Hammerstein cascade models. Biol. Cybern. **55**(2), 135–144 (1986). doi:10.1007/BF00341929

40. VJ Mathews, GL Sicuranza, *Polynomial Signal Processing*. (Wiley, New York, 2000)

41. GL Sicuranza, A Carini, in *Proc. IEEE Instrumentation and Measurement Technology Conf. (IMTC)*. On the accuracy of generalized Hammerstein models for nonlinear active noise control (IEEE, Sorrento, 2006), pp. 1411–1416

42. S Malik, G Enzner, in *Proc. IEEE Intl. Conf. Acoustics, Speech, Signal Process. (ICASSP)*. Fourier expansion of Hammerstein models for nonlinear acoustic system identification (IEEE, Prague, 2011), pp. 85–88

43. S Van Vaerenbergh, LA Azpicueta-Ruiz, in *Proc. IEEE Intl. Conf. Acoustics, Speech, Signal Process. (ICASSP)*. Kernel-based identification of Hammerstein systems for nonlinear acoustic echo-cancellation (IEEE, Florence, 2014), pp. 3739–3743

44. F Küch, A Mitnacht, W Kellermann, in *Proc. IEEE Intl. Conf. Acoustics, Speech, Signal Process. (ICASSP)*. Nonlinear acoustic echo cancellation using adaptive orthogonalized power filters, vol. 3 (IEEE, Philadelphia, 2005), pp. 105–108

45. F Küch, W Kellermann, Orthogonalized power filters for nonlinear acoustic echo cancellation. Signal Process. **86**(6), 1168–1181 (2006)

46. VJ Mathews, Adaptive polynomial filters. IEEE Signal Process. Mag. **8**(3), 10–26 (1991)

47. MJ Korenberg, IW Hunter, The identification of nonlinear biological systems: Volterra kernel approaches. Ann. Biomed. Eng. **24**(2), 250–268 (1996)

48. A Fermo, A Carini, GL Sicuranza, Low-complexity nonlinear adaptive filters for acoustic echo cancellation in GSM handset receivers. Eur. Trans. Telecommun. **14**(2), 161–169 (2003)

49. A Carini, S Cecchi, M Gasparini, GL Sicuranza, in *Proc. IEEE Intl. Conf. Acoustics, Speech, Signal Process. (ICASSP)*. Introducing Legendre nonlinear filters (IEEE, Florence, 2014), pp. 7939–7943

50. A Carini, S Cecchi, L Romoli, GL Sicuranza, in *Proc. European Signal Process. Conf. (EUSIPCO)*. Perfect periodic sequences for Legendre nonlinear filters (EURASIP, Lisbon, 2014), pp. 2400–2404

51. A Carini, GL Sicuranza, in *Proc. IEEE Intl. Conf. Acoustics, Speech, Signal Process. (ICASSP)*. Even mirror Fourier nonlinear filters (IEEE, Vancouver, 2013), pp. 5608–5612

52. GL Sicuranza, A Carini, in *Proc. IEEE Intl. Conf. Acoustics, Speech, Signal Process. (ICASSP)*. A novel class of BIBO stable recursive nonlinear filters (IEEE, Florence, 2014), pp. 7934–7938

53. S Haykin, *Adaptive Filter Theory*, 4th edn. (Prentice Hall, Upper Saddle River, 2002)

54. G Enzner. A Model-Based Optimum Filtering Approach to Acoustic Echo Control: Theory and Practice. Aachener Beiträge zu Digitalen Nachrichtensystemen (ABDN), vol. 22 (Verlag Mainz, Aachen, 2006)

55. M Zeller, LA Azpicueta-Ruiz, W Kellermann, in *Proc. IEEE Workshop Applications Signal Process. Audio Acoustics (WASPAA)*. Adaptive FIR filters with automatic length optimization by monitoring a normalized combination scheme (IEEE, New Paltz, 2009), pp. 149–152. doi:10.1109/ASPAA.2009.5346547

56. M Zeller, W Kellermann, in *Proc. Int. Workshop Acoustic Echo, Noise Control (IWAENC)*. Self-configuring system identification via evolutionary frequency-domain adaptive filters (IWAENC, Tel Aviv, 2010)

57. M Zeller, W Kellermann, in *Proc. European Signal Process. Conf. (EUSIPCO)*. Evolutionary adaptive filtering based on competing filter structures (EURASIP, Barcelona, 2011), pp. 1264–1268

58. WG Gardner, Efficient convolution without input-output delay. J. Audio Eng. Soc. **43**(2), 127–136 (1995)

59. H Buchner, J Benesty, T Gänsler, W Kellermann, in *Proc. Int. Workshop Acoustic Echo, Noise Control (IWAENC)*. An Outlier-Robust Extended Multidelay Filter with Application to Acoustic Echo Cancellation (IEEE Japan Council et al., Kyoto, 2003), pp. 19–22

60. F Küch, E Mabande, G Enzner, in *Proc. IEEE Intl. Conf. Acoustics, Speech, Signal Process. (ICASSP)*. State-space architecture of the partitioned-block-based acoustic echo controller (IEEE, Florence, 2014), pp. 1295–1299

61. J-S Soo, KK Pang, Multidelay block frequency domain adaptive filter. IEEE Trans. Acoust. Speech Signal Process. **38**(2), 373–376 (1990)

62. E Hänsler, G Schmidt, *Acoustic Echo and Noise Control: a Practical Approach*. (Wiley, Hoboken, 2004)

63. CSS Burrus, TW Parks, *DFT/FFT and Convolution Algorithms: Theory and Implementation*, 1st edn. (John Wiley & Sons, Inc., New York, 1985)

Hofmann *et al. EURASIP Journal on Advances in Signal Processing* (2016) 2016:113

Page 18 of 18

64. H Sorensen, D Jones, M Heideman, C Burrus, Real-valued fast fourier transform algorithms. IEEE Trans. Acoust. Speech Signal Process. **35**(6), 849–863 (1987)

65. HW Schüßler, An objective method for measuring the performance of weakly nonlinear and noisy systems. Frequenz. **41**(6–7), 147–154 (1987)

66. F Küch, *Adaptive polynomial filters and their application to nonlinear acoustic echo cancellation*. (PhD thesis, University Erlangen-Nuremberg, Germany, 2005)

67. G Enzner, in *ITG Conf. Speech Communication*. From acoustic nonlinearity to adaptive nonlinear system identification (VDE, Braunschweig, 2012), pp. 1–4

68. HW Schüßler, Y Dong, in *7. Aachener Symposium für Signaltheorie (ASST '90)*. Messung und Modellierung von schwach nichtlinearen Systemen (Springer, Aachen, 1990), pp. 2–7. (in German)

69. HW Schüßler, Y Dong, in *Proc. IEEE Intl. Conf. Acoustics, Speech, Signal Process. (ICASSP)*. A new method for measuring the performance of weakly nonlinear systems, vol. 4 (IEEE, Glasgow, 1989), pp. 2089–2092