2019 27th European Signal Processing Conference (EUSIPCO)

Bayesian Model Selection for Nonlinear Acoustic Echo Cancellation

Mhd Modar Halimeh, Andreas Brendel, and Walter Kellermann
Multimedia Communications and Signal Processing
Friedrich-Alexander-Universität Erlangen-Nürnberg,
Cauerstr. 7, D-91058 Erlangen, Germany.
{mhd.m.halimeh, andreas.brendel, walter.kellermann}@fau.de

Abstract—**In this paper, we introduce a Bayesian framework to perform model selection for nonlinear acoustic echo cancellation. This is especially important for scenarios where the functional form of the underlying nonlinear distortion is time-varying and/or is unknown, e.g., nonlinear distortions that vary with the volume level of the loudspeakers. To this end, the proposed method evaluates the model probabilities, or what is known as the evidence density, in a Bayesian manner. Thus, unlike convex and affine combination schemes of adaptive filters, the proposed method optimizes both the model complexity as well as the model performance by a single criterion. Moreover, by using the significance-aware principle, the proposed framework is realized in a computationally efficient way. The method is validated by three experiments using synthesized time-invariant nonlinearities, synthesized time-varying nonlinearities, and using real recorded nonlinearities.**

I. Introduction

Acoustic Echo Cancellation (AEC) is a crucial task for reliable acoustic human-machine full-duplex interfaces [1]. Due to the wide-spread use of mobile phones, nonlinear distortions are one of the most relevant challenges for acoustic echo cancellation research [2], [3]. Numerous publications address the task of Nonlinear Acoustic Echo Cancellation (NLAEC) by proposing a wide range of basis functions for Hammerstein models that try to model the unknown nonlinear distortions [4]–[8]. To estimate the model parameters, many algorithms have been proposed [9]–[12]. It should be noted that for alternative approaches, such as neural networks, e.g., [13], [14], hyperparameters have to be chosen as well.

A common strategy for NLAEC is to pre-assume a few model characteristics, e.g., the basis functions for the Hammerstein model-based NLAEC, the maximum order of the basis functions, or the number of neurons in the neural networks approaches. This can lead to a suboptimal choice of these characteristics and thus a reduction of the NLAEC performance. Moreover, such a strategy is prone to failure when the functional form of the nonlinear distortion is subject to change. The goal of this paper is to propose a Bayesian model selection scheme to allow for an autonomous choice of an appropriate nonlinear echo path model.

Known model selection strategies for NLAEC, e.g., [15]–[19], propose adaptive convex or affine combinations of candidate models which converge to the optimum model over time [15]–[19] by evaluating a signal error-based criterion only. Given sufficient convergence time, these strategies favor complex models which generalize poorly to changing input signal statistics and suffer from slow tracking of time-variant echo paths. Heuristic measures could be applied to [15]–[19] to consider the model complexity. However, it is not intuitive how to quantify the complexity of, e.g., Hammerstein models with different basis functions. Genetic algorithms [20] and correlation distance metrics-based methods [21] have also been proposed, but would result in prohibitively complex schemes for AEC where the number of degrees of freedom is typically on the order of hundreds to thousands.

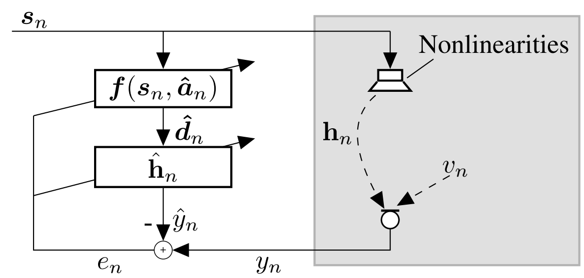

Fig. 1: NLAEC scenario.

In this paper, we introduce the Significance-Aware Evidence Maximization (SA-EM) to perform an efficient model selection for NLAEC. This framework is an extension of [22], where model selection is used for neural networks. The proposed framework enables a numerical comparison between different competing models for NLAEC that takes into consideration not only the errors produced by each model, but also the complexity of the model and the uncertainty associated with its parameter estimation process. Note that the robustness of the proposed approach against non-stationary noise, such as double-talk or burst-noise, is not considered as we aim at verifying the optimality of the model selection scheme. The robustness of the model selection should be addressed in a later work.

This paper is structured as follows: in Section II the NLAEC problem is introduced. In Section III, the evidence maximization framework is derived. In Section IV the proposed framework is evaluated for synthesized and real nonlinearities. Finally, conclusions are drawn in Section V.

978-9-0827-9703-9/19/$31.00 ©2019 IEEE

II. Problem Formulation

A typical NLAEC scenario is depicted in Figure 1. At time instance n, a block of length M of the input signal $\boldsymbol{s}_n = [s(n), s(n-1), ..., s(n-M-1)]^{\mathrm{T}}$ is nonlinearly distorted by the loudspeaker. The nonlinearly distorted signal is approximated by $\hat{\boldsymbol{d}}_n = \boldsymbol{f}_n(\hat{\boldsymbol{a}}_n, \boldsymbol{s}_n)$, where the nonlinear memoryless preprocessor $\boldsymbol{f}_n(\hat{\boldsymbol{a}}_n, \boldsymbol{s}_n)$ approximates the nonlinearity arising from the loudspeaker and is characterized by the estimated parameter vector $\hat{\boldsymbol{a}}_n = [\hat{a}_{0,n}, ..., \hat{a}_{Q,n}]^{\mathrm{T}}$ in

$$\boldsymbol{f}_n(\hat{\boldsymbol{a}}_n, \boldsymbol{s}_n) = \sum_{i=0}^{Q} \hat{a}_{i,n} \boldsymbol{g}_i(\boldsymbol{s}_n), \tag{1}$$

and where $\boldsymbol{g}_i$ denotes the i-th basis function, while Q denotes the highest order of the basis functions. It is common to assume that the optimum family of time-invariant basis functions and order Q are known a priori when modeling the nonlinear distortions of the far-end signal $\boldsymbol{s}_n$ [5], [6], [23], [24]. Moreover, the linear acoustic echo path between the loudspeaker and the microphone at time instant n is modeled by an estimated linear finite impulse response (FIR) filter of length M, $\hat{\mathbf{h}}_n = [\hat{h}_{0,n}, ..., \hat{h}_{M-1,n}]^{\mathrm{T}}$. Consequently, the estimated microphone signal is then obtained by

$$\hat{y}_n = \hat{\mathbf{h}}_n^{\mathrm{T}} \boldsymbol{f}_n(\hat{\boldsymbol{a}}_n, \boldsymbol{s}_n). \tag{2}$$

In this paper, we consider scenarios where the family of basis functions $\boldsymbol{g}_i$ and their maximum order Q are unknown and/or time-varying and propose an algorithm to estimate both in an optimal way.

III. The Evidence Maximization Framework

This paper introduces the use of the evidence maximization framework [22] and adapts it for NLAEC. In the following, an overview of the evidence maximization framework is given, where the relation describing the model probabilities is derived. Afterwards, the evidence maximization framework is adapted for the task of NLAEC by using the Significance-Aware principle which keeps the computational load of the overall framework small.

A. Overview

Denoting the j-th model defined by a family of basis functions and maximum order Q_j, $\mathcal{M}_j = \{\boldsymbol{g}_{ij} | i \in \{1, ..., Q_j\}\}$, the posterior probability of this model is

$$P(\mathcal{M}_j|\mathcal{D}) \propto P(\mathcal{M}_j) P(\mathcal{D}|\mathcal{M}_j), \tag{3}$$

where $\mathcal{D} = \{\boldsymbol{s}_k, y_k | n \in \{1, ..., n\}\}$ denotes the set containing the observed far-end and microphone signals.

Unless there is a reason to favor one model over another, so that $P(\mathcal{M}_j)$ is equal for all j, maximizing the model posterior probability $P(\mathcal{M}_j|\mathcal{D})$ becomes equivalent to maximizing the model likelihood $P(\mathcal{D}|\mathcal{M}_j)$. In the context of Bayesian model fitting, the model likelihood is known as the evidence, and is obtained by

$$P(\mathcal{D}|\mathcal{M}_j) = \int P(\mathcal{D}|\mathcal{M}_j, \hat{\boldsymbol{z}}_n) P(\hat{\boldsymbol{z}}_n|\mathcal{M}_j) \mathrm{d}\hat{\boldsymbol{z}}_n, \tag{4}$$

where $\hat{\boldsymbol{z}}_n = [\hat{z}_{1,n}, \hat{z}_{2,n}, ..., \hat{z}_{L,n}]^{\mathrm{T}}$ denotes the model parameters vector, i.e., $\hat{\boldsymbol{z}}_n = [\hat{\boldsymbol{a}}_n^{\mathrm{T}}, \hat{\mathbf{h}}_n^{\mathrm{T}}]^{\mathrm{T}}$, while $P(\hat{\boldsymbol{z}}_n|\mathcal{M}_j)$ is the prior density of the parameter vector given the model used.

By denoting $\hat{\boldsymbol{z}}_{\mathrm{MP},n}$ as the most probable parameter vector and assuming that (4) describes an integral over a Gaussian density, we arrive (by using Laplace's method) at [25]

$$P(\mathcal{D}|\mathcal{M}_j) \approx P(\mathcal{D}|\mathcal{M}_j, \hat{\boldsymbol{z}}_{\mathrm{MP},n}) \underbrace{P(\hat{\boldsymbol{z}}_{\mathrm{MP},n}|\mathcal{M}_j)(2\pi)^{\frac{L}{2}} \det^{-\frac{1}{2}} \mathbf{C}}_{\text{Occam factor}}, \tag{5}$$

where $\mathbf{C} = -\nabla\nabla \log P(\hat{\boldsymbol{z}}_n|\mathcal{D}, \mathcal{M}_j)$ is the Hessian matrix of the parameters posterior. Equation (5) is used throughout the rest of this paper to evaluate the evidence.

To provide an intuition into how the so-called 'Occam factor' [26] in (5) quantifies the model complexity, assume that the parameters posterior

$$P(\hat{\boldsymbol{z}}_n|\mathcal{D}, \mathcal{M}_j) \propto P(\mathcal{D}|\mathcal{M}_j, \hat{\boldsymbol{z}}_n) P(\hat{\boldsymbol{z}}_n|\mathcal{M}_j)$$

has a sufficiently dominant and narrow peak at the most probable parameter vector $\hat{\boldsymbol{z}}_{\mathrm{MP},n}$. Then, the evidence can be approximated by the product of the peak's height $P(\mathcal{D}|\mathcal{M}_j, \hat{\boldsymbol{z}}_{\mathrm{MP},n}) P(\hat{\boldsymbol{z}}_{\mathrm{MP},n}|\mathcal{M}_j)$ and volume $\Delta\hat{\boldsymbol{z}}_n$ [22]

$$P(\mathcal{D}|\mathcal{M}_j) \approx P(\mathcal{D}|\mathcal{M}_j, \hat{\boldsymbol{z}}_{\mathrm{MP},n}) \underbrace{P(\hat{\boldsymbol{z}}_{\mathrm{MP},n}|\mathcal{M}_j)\Delta\hat{\boldsymbol{z}}_n}_{\text{Occam factor}}. \tag{6}$$

Additionally assuming a uniform prior covering the volume $\Delta_0 \boldsymbol{z}$ simplifies the Occam factor to $\frac{\Delta\hat{\boldsymbol{z}}_n}{\Delta_0 \boldsymbol{z}}$, which describes the ratio between the volumes of the posterior and the prior. This implies that the Occam factor provides a measure for the model-specific simplicity of the parameterization and is thus not just responsible for penalizing a high model order.

B. The Significance-Aware Evidence Maximization

An obvious and immediate problem of using (5) for AEC lies in the term $\det^{-\frac{1}{2}} \mathbf{C}$, where $\mathbf{C}$ tends to be a rank-deficient matrix for long Room Impulse Responses (RIRs), due to the tails having low energy. To this end, we employ the Significance-Aware (SA) [11] principle, which in addition to overcoming the aforementioned difficulty eases the computational load of the algorithm. The SA concept divides the RIR into two parts $\hat{\mathbf{h}}_n = [\hat{\mathbf{h}}_{\text{direct},n}^{\mathrm{T}}, \hat{\mathbf{h}}_{\text{comp},n}^{\mathrm{T}}]^{\mathrm{T}}$. The direct echo path component $\hat{\mathbf{h}}_{\text{direct},n}$ of length $2R+1$ is centered around the highest energy peak, which contains the most significant part of the RIR. The complementary component $\hat{\mathbf{h}}_{\text{comp},n}$ contains the remaining coefficients. The SA principle is employed as follows: an estimate of the RIR vector is obtained using the Normalized Least Mean Square (NLMS) algorithm

$$\hat{\mathbf{h}}_n = \hat{\mathbf{h}}_{n-1} + \frac{\mu}{\hat{\boldsymbol{d}}_n^{\mathrm{T}} \hat{\boldsymbol{d}}_n + \epsilon} \hat{\boldsymbol{d}}_n e_n, \tag{7}$$

where μ is a scalar step size, ϵ is a positive constant to prevent division by zero, and $e_n = y_n - \hat{y}_n$ is the error signal. The adaptive FIR filter (7) is then subdivided into the components $\hat{\mathbf{h}}_{\text{direct},n}$, and $\hat{\mathbf{h}}_{\text{comp},n}$ as a starting point for the next step. The parameter vector is chosen to consist of the direct acoustic path only $\hat{\boldsymbol{z}}_n = [\hat{\mathbf{h}}_{\text{direct},n}^{\mathrm{T}}, \hat{\boldsymbol{a}}_n^{\mathrm{T}}]^{\mathrm{T}}$, and the nonlinearly distorted

			No nonlinear distortions	Underlying nonlinearity (16)		Underlying nonlinearity (17)	
			$Q_{\texttt{true}} = 0$	$Q_{\texttt{true}} = 1$	$Q_{\texttt{true}} = 2$	$Q_{\texttt{true}} = 1$	$Q_{\texttt{true}} = 2$
Linear AEC	$Q_1 = 0$	Probability	**0.39**	0.17	0.18	0.1	0.05
		ERLE/[dB]	**25.7**	18.2	15.3	13	12.5
PF-based	$Q_2 = 1$	Probability	0.21	**0.29**	0.18	0.25	0.2
		ERLE/[dB]	25.7	**24.1**	19.4	22.3	20.7
	$Q_3 = 2$	Probability	0.12	0.17	**0.25**	0.14	0.27
		ERLE/[dB]	25.6	23.9	**22.1**	22.2	24.0
Legendre-based	$Q_4 = 1$	Probability	0.17	0.24	0.23	**0.38**	0.17
		ERLE/[dB]	24.8	23.4	20.0	**23.8**	14.5
	$Q_5 = 2$	Probability	0.11	0.12	0.14	0.13	**0.31**
		ERLE/[dB]	25.2	22.9	20.8	22.2	**24.5**

TABLE I: ERLE and average model probabilities for static nonlinearities, see Subsection IV-A

signal is decomposed analogously into the subvectors $\hat{\boldsymbol{d}}_{\text{direct},n}$ and $\hat{\boldsymbol{d}}_{\text{comp},n}$ [11]. Based on this, the direct path component of the microphone signal is obtained through

$$y_{\text{direct},n} = y_n - y_{\text{comp},n} = y_n - \hat{\mathbf{h}}^{\text{T}}_{\text{comp},n}\hat{\boldsymbol{d}}_{\text{comp},n}, \tag{8}$$

while the error associated with the direct path is estimated via

$$e_{\text{direct},n} = y_{\text{direct},n} - \hat{y}_{\text{direct},n} = y_{\text{direct},n} - \hat{\mathbf{h}}^{\text{T}}_{\text{direct},n}\hat{\boldsymbol{d}}_{\text{direct},n} \tag{9}$$

The direct path signals $y_{\text{direct},n}$, $\hat{\boldsymbol{d}}_{\text{direct},n}$ and $e_{\text{direct},n}$ are then used to adapt the parameter vector $\hat{\boldsymbol{z}}_n = [\hat{\mathbf{h}}^{\text{T}}_{\text{direct},n}, \hat{\boldsymbol{a}}^{\text{T}}_n]^{\text{T}}$ using the NLMS algorithm.

In order to evaluate (5) for a specific model $\mathcal{M}_j$, we adopt a block-based scheme: at the end of block l, we start by evaluating the parameter vector likelihood

$$P(\mathcal{D}_l|\mathcal{M}_j, \hat{\boldsymbol{z}}_{\text{MP},l}) = \mathcal{N}(||\boldsymbol{e}_{\text{direct},l}||_2, \sigma_v^2), \tag{10}$$

where $\mathcal{D}_l$ denotes the set containing the l-th frame of the far-end and microphone signal. $\hat{\boldsymbol{z}}_{\text{MP},l}$ denotes the latest estimate of the parameter vector $\hat{\boldsymbol{z}}$, and σ_v^2 denotes the variance of the additive white noise in the microphone signal. $||\cdot||_2$ is the l_2 norm, and $\boldsymbol{e}_{\text{direct},l}$ is the direct path error signal for the l-th frame.

Evaluating (10) implicitly assumes that the latest estimate of the parameter vector is not too far from the initial estimate at the start of the l-th frame. Finally, the Hessian $\hat{\mathbf{C}}_n$ matrix estimate is updated by

$$\hat{\mathbf{C}}_n = (1-\lambda)\hat{\mathbf{C}}_{n-1} + \lambda\hat{\mathbf{C}}_{\text{instant},n}, \tag{11}$$

where $0 < \lambda < 1$ is a forgetting factor, while $\hat{\mathbf{C}}_{\text{instant.},n}$ is the instantaneous estimate of the Hessian obtained, following the notation in [26], via

$$\frac{\partial^2 \log P(\hat{\boldsymbol{z}}_n|\mathcal{D},\mathcal{M}_j)}{\partial \mathbf{h}_{\text{direct},n}\partial \mathbf{h}^{\text{T}}_{\text{direct},n}} = \hat{\boldsymbol{d}}_{\text{direct},n}\hat{\boldsymbol{d}}^{\text{T}}_{\text{direct},n}, \tag{12}$$

$$\frac{\partial^2 \log P(\hat{\boldsymbol{z}}_n|\mathcal{D},\mathcal{M}_j)}{\partial \boldsymbol{a}_n\partial \boldsymbol{a}^{\text{T}}_n} = \boldsymbol{G}^{\text{T}}_n\mathbf{h}_{\text{direct},n}\mathbf{h}^{\text{T}}_{\text{direct},n}\boldsymbol{G}_n, \tag{13}$$

$$\frac{\partial^2 \log P(\hat{\boldsymbol{z}}_n|\mathcal{D},\mathcal{M}_j)}{\partial \boldsymbol{a}_n\partial \mathbf{h}^{\text{T}}_{\text{direct},n}} = \boldsymbol{G}^{\text{T}}_n \cdot e_{\text{direct},n}, \tag{14}$$

where $\boldsymbol{G}_n = [\boldsymbol{g}_1(\boldsymbol{s}_n), \boldsymbol{g}_2(\boldsymbol{s}_n), ..., \boldsymbol{g}_{Q_j}(\boldsymbol{s}_n)]$.

IV. Experiments

The proposed SA-EM is evaluated for three scenarios. The first scenario aims at verifying that the SA-EM is capable of selecting the underlying time-invariant nonlinearity in Subsection IV-A. In the second experiment, we evaluate the performance of the SA-EM in tracking an abrupt change in the underlying echo path in Subsection IV-B. Finally, an experiment with real recordings using a commercial mobile phone is conducted. This experiment aims at demonstrating the benefit of using the SA-EM in scenarios such as double-talk where continuous adaptation of the competing models is not possible in Subsection IV-C.

To evaluate the performance of each model, the Echo Return Loss Enhancement (ERLE) is used [12]

$$\text{ERLE}_n = 10\log_{10}\frac{\text{E}\{y_n^2\}}{\text{E}\{e_n^2\}}, \tag{15}$$

where $\text{E}(\cdot)$ denotes the expectation operator.

In evaluating the SA-EM, the direct path is modeled using 11 coefficients, i.e., $R = 5$. In addition, $\lambda = 0.005$ is used for updating the Hessian matrix in (11). The parameter vector's prior $P(\hat{\boldsymbol{z}}_n|\mathcal{M}_j)$ in (5) is identical for the different models and is set as a uniform distribution over the range $0 < z_i < 5; \forall i$. The SA-EM uses frames of length $K = 512$ samples.

A. SA-EM for a Static Nonlinearity

In this experiment, we aim to verify that the SA-EM is indeed able to select the correct basis functions for nonlinearly distorted and noisy speech signals in reverberant environments. To this end, we use a female speech signal $\boldsymbol{s}_n$ with a sampling frequency of 16kHz. The signal $\boldsymbol{s}_n$ is nonlinearly distorted via $\boldsymbol{d}_n = \boldsymbol{f}(\boldsymbol{a}_n, \boldsymbol{s}_n)$. Synthesized nonlinearities are used in this experiment as for nonlinearities in real-world recordings, no ground-truth information is available to judge the performance of the selection method.

The SA-EM is evaluated for two different families of basis functions, namely the monomial basis of odd order, which is also known as Power filters (PF) [6]

$$\boldsymbol{f}(\boldsymbol{a}_n, \boldsymbol{s}_n) = \sum_{i=0}^{Q_j} a_i \boldsymbol{s}_n^{2i+1}, \tag{16}$$

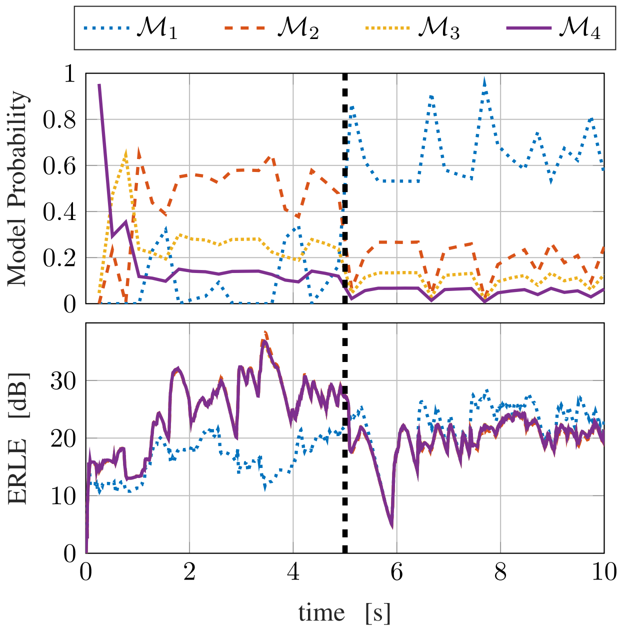

Fig. 2: ERLE and model probabilities, see Subsection IV-B

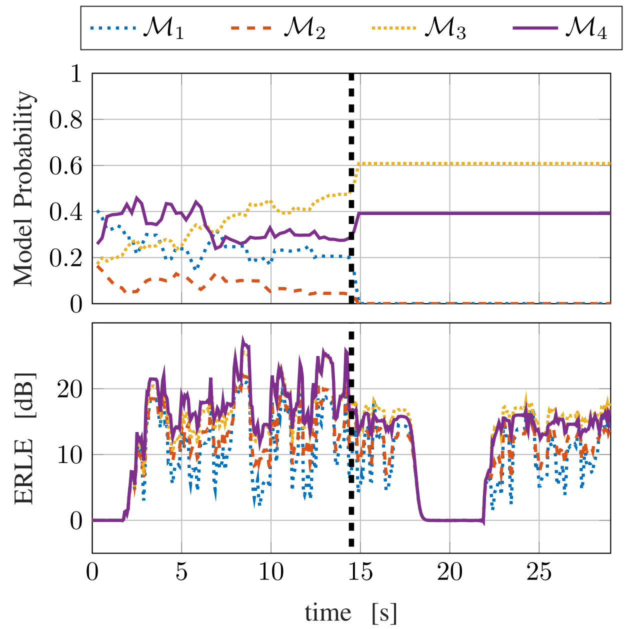

Fig. 3: ERLE and model probabilities, see Subsection IV-C

where $\boldsymbol{s}_n^{2i+1} = [s(n)^{2i+1}, ..., s(n-M-1)^{2i+1}]^{\mathrm{T}}$ and the Legendre polynomials (LP) of odd order $\mathcal{L}_{2i+1}(\cdot)$ [5]

$$\boldsymbol{f}(\boldsymbol{a}_n, \boldsymbol{s}_n) = \sum_{i=0}^{Q_j} a_i \mathcal{L}_{2i+1}(\boldsymbol{s}_n). \tag{17}$$

In addition to being evaluated for two different families of basis functions, the proposed framework is also evaluated for different maximum orders Q_j of the nonlinear functions. After being distorted by one of the nonlinear functions, the signal $\boldsymbol{d}_n$ is convolved by a recorded (RIR) $\mathbf{h}$, and corrupted by white Gaussian noise v_n

$$y_n = \mathbf{h}^{\mathrm{T}} \boldsymbol{d}_n + v_n, \tag{18}$$

at a Signal-to-Noise Ratio (SNR) of 30dB.

In Table I, the ERLE and model probabilities, averaged over all frames, are summarized for the different models. For the case of a purely linear echo path, i.e., no nonlinear distortion, it can be seen that even though the different models and orders perform similarly in terms of the ERLE values, the model probability clearly favors the purely linear AEC. Furthermore, when introducing the nonlinear distortion to the signal, the SA-EM is able to successfully identify the underlying functional form of the nonlinearity and the correct order. Moreover, it can be observed that the SA-EM does not simply favor the model with the lowest error, where despite having several models yielding the same performance level for the purely linear echo path, the SA-EM still favors the purely linear AEC model heavily.

B. Tracking a Time-variant Model

In this experiment, the SA-EM is evaluated for an echo path that goes through a drastic change due to the sudden disappearance of the nonlinear distortion. Such a scenario can occur in miniaturized loudspeakers for an abrupt change in the volume levels. To this end, a female speech signal $\boldsymbol{s}_n$ with a sampling frequency of 16kHz is nonlinearly distorted via (16) with $Q = 1$ and $a_1 = 1.2$ and then convolved by a recorded impulse response $\mathbf{h}$. Finally, the microphone signal is corrupted by white noise resulting in an SNR level of 30dB. To simulate a sudden change in the underlying echo path, the nonlinear distortion is removed at $t = 5$s, i.e., $a_1 = 0$ in (16), and the underlying model is turned into a purely linear one.

The competing models in this experiment are a purely linear model $\{\mathcal{M}_1, Q_1 = 0\}$ and three models based on (16) with $\{\mathcal{M}_2, Q_2 = 1\}$, $\{\mathcal{M}_3, Q_3 = 2\}$, and $\{\mathcal{M}_4, Q_4 = 3\}$.

The ERLE of the different models and probabilities are depicted in Figure 2, where the change in the echo path is marked by the dashed line. As it can be seen in the figure, the SA-EM is capable of quickly identifying the change in the model. Moreover, a stable performance can be observed, where other than during the first second, which reflects the model parameters convergence interval, the SA-EM is stable in favoring the correct model even during a speech pause around $t = 6$s. Furthermore, the benefit of using the SA-EM for selecting the underlying model is clear, where by simply choosing the most probable model at the end of each frame, a gain of 3 dB is obtained compared to using a nonlinear model which is still adapted for the entire duration.

Finally, the SA-EM resulted in a 20% runtime overhead compared to adapting the competing models by the NLMS algorithm.

C. Real System Performance

In this experiment a speech signal, sampled at $f_s = 16$kHz, is emitted and recorded by a commercial mobile phone at a SNR level of 30dB. The signal is almost 30s long.

The different models considered in this experiment are a purely linear model $\{\mathcal{M}_1, Q_1 = 0\}$ and 3 models based on Legendre polynomials of odd orders (17), i.e., $\{\mathcal{M}_2, Q_2 = 1\}$, $\{\mathcal{M}_3, Q_3 = 2\}$,and $\{\mathcal{M}_4, Q_4 = 3\}$. The different models are configured with a linear filter $\hat{\mathbf{h}}_n$ of 265 taps and the different coefficients $\hat{\mathbf{h}}_n$ and $\hat{\boldsymbol{a}}_n$ are adapted via the NLMS algorithm.

The adaptation of different models and of the SA-EM is stopped at $t \approx 15$s corresponding to freezing the filters in a double-talk scenario.

The ERLE resulting from each model is depicted in Figure 3, where the model of highest order $\{\mathcal{M}_4, Q_4 = 3\}$ performs best during the online adaptation period. Once the adaptation is stopped, i.e., represented by the dashed line, a model of a lower order $\{\mathcal{M}_3, Q_3 = 2\}$ actually generalizes better and yields higher ERLE values.

As seen from the model probabilities in the figure, the SA-EM increasingly favors the model that generalizes best, i.e., $\{\mathcal{M}_3, Q_3 = 2\}$. As a consequence, using the SA-EM, one is capable of pre-selecting the model that performs best in cases such as double-talk where continuous adaptation is not always possible.

V. Conclusion

In this paper, we introduced the SA-EM, a Bayesian framework for model selection in NLAEC. A task that is relevant in cases where the nonlinear distortion are either entirely unknown and several proposed models need to be compared, or in cases where the nonlinear distortion is subject to an abrupt and drastic change.

The proposed framework is realized efficiently by using the SA principle, resulting in a limited computational overhead. In order to verify the proposed framework, three experiments were conducted. The first experiment was conducted using synthesized nonlinear distortions where the SA-EM ability of selecting the correct underlying models was verified. A second experiment, also using synthesized nonlinear distortions is carried out, where the ability of the SA-EM was capable of tracking an abrupt change in the underlying nonlinear model. Finally, an experiment using real recordings was conducted to demonstrate the benefit of using the SA-EM in a double-talk situation, where the SA-EM correctly selects the model that generalizes best.

VI. Acknowledgment

The authors would like to thank the Deutsche Forschungsgemeinschaft (DFG) for supporting this work (contract number KE 890/9-1).

References

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

[2] S. Theodoridis and R. Chellappa, *Academic press library in signal processing, volume 4: image, video processing and analysis, hardware, audio, acoustic and speech processing*, Academic Press, Inc., Orlando, FL, USA, 1st edition, 2014.

[3] D. Comminiello and J. Principe, *Adaptive learning methods for nonlinear system modelling*, Butterworth-Heinemann, 1st edition, 2018.

[4] A. Stenger *et al.*, "Nonlinear acoustic echo cancellation with 2nd order adaptive Volterra filters," in *Int. Conf. on Acoust., Speech, and Signal Process. (ICASSP)*, Mar. 1999, pp. 15–19.

[5] A. Carini *et al.*, "Introducing Legendre nonlinear filters," in *ICASSP*, May 2014, pp. 7939–7943.

[6] F. Kuech *et al.*, "Nonlinear acoustic echo cancellation using adaptive orthogonalized power filters," in *ICASSP*, Mar. 2005, pp. 105–108.

[7] J. Malek and Z. Koldovsky, "Hammerstein model-based nonlinear echo cancelation using a cascade of neural network and adaptive linear filter," in *Int. Workshop on Acoust. Signal Enhancement (IWAENC)*, Sept. 2016, pp. 1–5.

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

[9] M. M. Halimeh *et al.*, "Nonlinear acoustic echo cancellation using elitist resampling particle filter," in *ICASSP*, April 2018, pp. 236–240.

[10] C. Huemmer *et al.*, "Estimating parameters of nonlinear systems using the elitist particle filter based on evolutionary," *IEEE/ACM Trans. on Audio, Speech and Language Process.*, vol. 26, pp. 595–608, 2018.

[11] C. Hofmann *et al.*, "Significance-aware Hammerstein group models for nonlinear acoustic echo cancellation," in *ICASSP*, May 2014, pp. 5934–5938.

[12] E. Haensler and G. Schmidt, *Topics in acoustic echo and noise control*, Springer-Verlag Berlin Heidelberg, 2006.

[13] A. Birkett and R. Goubran, "Acoustic echo cancellation using NLMS-neural network structures," in *ICASSP*, May 1995, pp. 3035–3038.

[14] H. Zhang and D. Wang, "Deep learning for acoustic echo cancellation in noisy and double-talk scenarios," in *Proc. Interspeech 2018*, 2018, pp. 3239–3243.

[15] M. Zeller *et al.*, "Adaptive Volterra filters with evolutionary quadratic kernels using a combination scheme for memory control," *IEEE Trans. Signal Process.*, vol. 59, no. 4, pp. 1449–1464, Apr. 2011.

[16] M. Zeller and W. Kellermann, "Evolutionary adaptive filtering based on competing filter structures," in *EUSIPCO*, Aug 2011, pp. 1264–1268.

[17] J. Arenas-Garcia *et al.*, "Combinations of adaptive filters: performance and convergence properties," *IEEE Signal Process. Mag.*, vol. 33, no. 1, pp. 120–140, Jan 2016.

[18] M. Zeller *et al.*, "Adaptive FIR filters with automatic length optimization by monitoring a normalized combination scheme," in *IEEE Workshop on Applications of Signal Process. to Audio and Acoust.*, Oct 2009, pp. 149–152.

[19] S. Van Vaerenbergh *et al.*, "A split kernel adaptive filtering architecture for nonlinear acoustic echo cancellation," in *EUSIPCO*, Aug 2016, pp. 1768–1772.

[20] H. Takata T. Hachino, "Structure selection and identification of Hammerstein type nonlinear systems using automatic choosing function model and genetic algorithm," *IEICE Trans. on Fundamentals of Electronics, Communications and Computer Sci.*, vol. E88, no. 10, pp. 2541–2547, Oct. 2005.

[21] J. Solares and H. Wei, "Nonlinear model structure detection and parameter estimation using a novel bagging method based on distance correlation metric," *Nonlinear Dynamics*, vol. 82, pp. 201215, May 2015.

[22] D. J. C. MacKay, *Bayesian methods for adaptive models*, Ph.D. thesis, California Institute of Technology, 1991.

[23] S. Shimauchi and Y. Haneda, "Nonlinear acoustic echo cancellation based on piecewise linear approximation with amplitude threshold decomposition," in *IWAENC*, Sept. 2012, pp. 1–4.

[24] P. Shah *et al.*, "Nonlinear acoustic echo cancellation using voltage and current feedback," *IEEE/ACM Trans. Audio, Speech and Language Process.*, vol. 23, no. 10, pp. 1589–1599, Oct. 2015.

[25] C. Bishop, *Pattern recognition and machine learning*, Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2006.

[26] D. MacKay, *Information theory, inference & learning algorithms*, Cambridge University Press, New York, USA, 2002.