HAMMERSTEIN MODEL-BASED NONLINEAR ECHO CANCELATION USING A CASCADE OF NEURAL NETWORK AND ADAPTIVE LINEAR FILTER

Jiri Malek and Zbyněk Koldovský

Faculty of Mechatronics, Informatics, and Interdisciplinary Studies, Technical University of Liberec, Studentská 2, 461 17 Liberec, Czech Republic.

jiri.malek@tul.cz

ABSTRACT

We present a novel nonlinear echo cancellation method that assumes the Hammerstein nonlinear system model. Model parameters are identified using a neural network followed by an adaptive linear filter. The parameters of both subsystems are estimated separately, which allows the utilization of computationally efficient conventional methods. The proposed method is verified on simulated as well as on real-world signals. In comparison to power-filter echo cancelers, the method achieves significantly higher echo suppression provided that the excitation signal is white noise, and it yields comparable performance when the signal is speech.

***Index Terms*—** Nonlinear echo cancelation, power filter, neural network, least mean squares algorithm.

1. INTRODUCTION

The echo is a detrimental phenomenon arising in the context of duplex communication, e.g., during phone calls. Here, speech originating from the far-end is played by a loudspeaker. Then it is convolved with the impulse response of the acoustic environment and is received by a microphone. The convolved signal is the acoustic echo, which is transmitted back to the far-end speaker, along with the near-end speech [1].

There are many studies where the echo path is modeled as time-(in)variant and linear. Then, only the acoustic path is taken into account, assuming a linear recording microphone. However, electronic circuits within the emitting device may introduce an additional *nonlinear* distortion to the far-end signal. For example, cell phones often contain low-cost components and loudspeakers with nonlinear characteristics.

Such distortions are modeled as nonlinear systems either with memory or memoryless. A memoryless system is described as a nonlinear function applied sample-wise to the input signal. Systems with memory are often described as Volterra filters [2], whose identification is computationally demanding [3]. A simplified model for nonlinearities with memory represents the Hammerstein model [4], which is a memoryless nonlinearity followed by a linear system. Alternative approaches to model nonlinear systems are neural networks [5].

Methods for the nonlinear echo removal are called Nonlinear Acoustic Echo Cancelers (NAEC). They attempt to identify the parameters of the models described above, assuming that the far-end signal is available as a reference. An adaptive technique based on Volterra filters was proposed in [6]. Considering only diagonal elements in Volterra kernels, linear-in-parameters power filters [7] are obtained. Adaptive identification procedure for Hammerstein models was proposed in [8], where it is assumed that the output of the loudspeaker is observable, which rarely happens in practice. Authors of [9,10] alleviate this need by optimizing the parameters of both Hammerstein subsystems simultaneously; using a single neural network. Another network topology called Functional Link Adaptive Filters was published in [11].

In this paper, we propose a method that assumes the Hammerstein model and identifies it using a cascade of a neural network followed by an adaptive linear filter. Compared to [8], the reference output of the loudspeaker is not required. Instead, an artificial reference signal is used to train the neural network, i.e., separate identification of the linear and the nonlinear subsystem is performed, unlike to works [9, 10]. To identify the linear subsystem of the Hammerstein model, three adaptive algorithms are tested: Normalized Least Mean Squares (NLMS [12]), Recursive Least Squares (RLS [12]), and the Affine Projection Algorithm (APA [13]). The utilization of conventional optimization methods enables fast adaptive filter implementations or updates of only one subsystem in the cascade at a time. We verify the functionality of the proposed method on simulated and on real-world signals.

2. PROPOSED ALGORITHM

2.1. Problem formulation

The echo cancellation scenario is illustrated in Fig. 1. The signal on the microphone $d(n)$ consists of the echo $y(n)$, the background noise $v(n)$ and the near-end signal $z(n)$. The

This work was supported by The Czech Science Foundation through Project No. 14-11898S.

978-1-5090-2007-2/16/$31.00

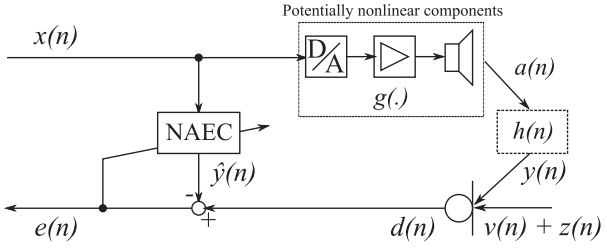

Fig. 1. Nonlinear echo creation (and cancellation) along with the Hammerstein model describing the process

Fig. 2. Proposed Nonlinear Acoustic Echo Canceler

echo $y(n)$ is a distorted version of reference signal $x(n)$. First, $x(n)$ is distorted by low-quality electronic components (modeled by a memoryless system $g(\cdot)$). Then, the resulting signal $a(n)$ is convolved with filter $h(n)$, which represents the influence of the acoustic environment. The NAEC block attempts to minimize the contribution of the echo signal $y(n)$ to the error signal $e(n)$. This is done by subtracting an estimate of the echo signal $\hat{y}(n)$ from $d(n)$.

In this paper, we consider the parameter identification only during intervals where the near-end signal is not active, i.e., when $z(n) = 0$ and $d(n) = v(n)+y(n)$. The presence of the near-end speech complicates the identification even when the system is only linear; see, for example, [14] and [15].

2.2. NAEC structure

The structure of the proposed NAEC shown in Fig. 2 assumes the Hammerstein model. The subsystem represented by $\hat{g}(\cdot)$ aims to embrace the nonlinearity in the echo path $g(\cdot)$. It is realized as a feed-forward neural network with two hidden layers. Next, the FIR filter $\hat{h}_1(n)$ of length L involves the linear part, and, ideally, it corresponds to the estimate of $h(n)$. The cascade of the subsystems suffers from the scaling ambiguity (multiplication of $\hat{g}(\cdot)$ by a constant β is canceled by multiplying the $\hat{h}_1(n)$ by $1/\beta$). Therefore, $\hat{h}_1(n)$ is normalized as $\hat{h}_1(n) \leftarrow \hat{h}_1(n)/\max|\hat{h}_1(n)|$ during each parameter update in order to cope with this indeterminacy.

Let $\hat{h}_1^n(k)$ be the value of the kth tap of the filter at the nth time instant. Let $\hat{g}^n(\cdot)$ represent the neural network at the nth time index. The adaptation proceeds by repeating the following three steps with each new input sample.

Step 1: The unknown signal $a(n)$ is estimated from the microphone signal $d(n)$ as the *backward* estimate $\hat{a}_b(n)$ given by

$$\hat{a}_b(n) = \hat{\mathbf{p}}^T(n-1)\mathbf{d}(n), \quad (1)$$

where

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

$$\mathbf{d}(n) = [d(n), d(n-1), \ldots, d(n-2L+1)]^T. \quad (3)$$

Here, $\hat{p}^{n-1}$ of length $2L$ stands for the approximate inverse filter of $\hat{h}_1^{n-1}$ in the least square sense given by

$$\hat{\mathbf{p}}(n-1) = [\mathbf{H}_1(n-1)^T\mathbf{H}_1(n-1)]^{-1}\mathbf{H}_1(n-1)^T\mathbf{b} \quad (4)$$

where $\mathbf{H}(n-1)$ is the $2L \times 2L$ Toeplitz matrix whose first column and row are, respectively, $[\hat{h}_1^{n-1}(0), \ldots, \hat{h}_1^{n-1}(L-1), 0, \ldots, 0]^T$ and $[\hat{h}_1^{n-1}(0), 0, \ldots, 0]$. The column vector $\mathbf{b}$ of length $2L$ is given by $[1, 0, \ldots, 0]^T$.

Step 2: The neural network weights are updated using the backpropagation algorithm [16] with gradient descent. The goal is to minimize the energy of the error signal defined as

$$e_g(n) = \hat{a}_b(n) - \hat{g}^{n-1}(x(n)) = \hat{a}_b(n) - \hat{a}_f^{n-1}(n), \quad (5)$$

where $\hat{a}_f^{n-1}(n)$ is the *forward* estimate of the unobservable signal $a(n)$ computed using neural network parameters at instant $n-1$. The update of weights results into new estimate of the forward signal $\hat{a}_f^n(n) = \hat{g}^n(x(n))$.

Step 3: The parameters of the linear subsystem $\hat{h}_1$ are updated. We study the applicability of three well-known adaptive algorithms to perform this step: NLMS [12], RLS [12] and APA [13].

The NLMS method is computationally cheap but exhibits slow convergence when the input signal is speech [13]. The RLS algorithm mitigates this disadvantage at the cost of a higher computational complexity. The APA stands for a compromise between the latter two algorithms. The algorithms are all based on the minimization of quadratic criteria measured on the error signal

$$e_1(n) = d(n) - \hat{\mathbf{h}}_1(n-1)^T\hat{\mathbf{a}}_f(n), \quad (6)$$

where

$$\hat{\mathbf{h}}_1(n-1) = [\hat{h}_1^{n-1}(0), \hat{h}_1^{n-1}(1), \ldots, \hat{h}_1^{n-1}(L-1)]^T, \quad (7)$$

$$\hat{\mathbf{a}}_f(n) = [\hat{a}_f^n(n), \hat{a}_f^n(n-1), \ldots, \hat{a}_f^n(n-L+1)]^T. \quad (8)$$

2.3. A versatile solution

In cases where the system is, in principle, linear or very close to being linear (that is, $g(x(n)) \approx x(n)$), the proposed NAEC is outperformed by conventional methods that are designed for linear AEC. In order to have the NAEC suitable also for the linear cases, we propose its final structure that is shown in Fig. 3. It consists of parallel connection of the proposed NAEC and of a conventional linear echo canceler that is based on the NLMS algorithm.

Let $\hat{y}_1(n)$ denote the output of the proposed NAEC and $\hat{y}_2(n)$ denote the output of the linear AEC. The goal here is

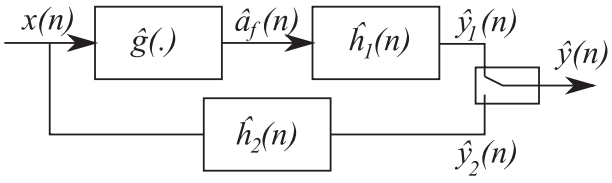

Fig. 3. Parallel version of the proposed algorithm

to select the one result that exhibits lower echo cancellation error. Hence, the final output is $\hat{y}(n) = \hat{y}_i(n)$ where

$$i = \arg \min_{j \in \{1,2\}} \sum_{k=0}^{C} e_j^2(n-k), \tag{9}$$

where $e_1(n)$ is defined in (6) while $e_2(n)$ is the error of the linear echo canceler; C is a free integer parameter.

2.4. Implementation details

To achieve an economical computational burden, the parameter updates are performed in a "mini-batch" manner. Compared to the sample-wise processing, this means that the changes in parameters are accumulated during short intervals of length B while the update proceeds only once at the end of each such interval. Similarly, the inverse filter $\hat{p}(n)$ is updated only once per every B samples.

It is worth to point to the fact that, in some situations, the nonlinearity within the echo path is time-invariant (e.g., a distortion due to electronic circuits). Then, once the nonlinear part appears to be identified, its parameters can be fixed and the computation of $\hat{p}(n)$ can be omitted.

3. SIMULATIONS

In experiments, the situation corresponding to Fig. 1 is considered, where $x(n)$ is either a Gaussian noise with zero mean and variance equal to $1/3$ or a speech signal whose average variance is 0.05; the sampling rate is 16 kHz. The close-talk signal is not active, i.e., $z(n) = 0$. Three nonlinearities are considered: (1) $g_1(x(n)) = \tanh(5 \cdot x(n))$, which is symmetric, (2) the identity $g_2(x(n)) = x(n)$, and (3) the asymmetric sigmoid nonlinearity $g_3(\cdot)$ simulating loudspeaker distortion, which is described in details in paper [11].

The impulse response $h(n)$ of length $L = 100$ is generated as $h(n) = 0.1 \cdot \xi \exp(1.1|n - D|^{0.2})$, where ξ is a random gaussian variable with zero mean and variance one, and $D = 5$ is the index of the main peak of the impulse response. The step-size parameters within NLMS and APA are $\mu = 0.03$ and the forgetting factor in RLS is $\lambda = 0.9999$. In APA, $K = 10$, which is the number of past errors that are taken into account (with $K = 1$, APA coincides with NLMS). The neural network consists of two hidden layers where each one contains 5 neurons. The input as well as the output layer contains single neuron. The learning rate is 0.05, mini-batch size $B = 50$ (Section 2.4) and parameter $C = 1000$ in (9).

The performance of NLMS/APA is sensitive to the noise present in $d(n)$. Therefore, a regularization proposed in [19] is used to alleviate this drawback. We select the regularization parameter $\delta = 0.55$ (equation (9) in [19]), which is recommended for filter length $L = 100$ and SNR 20 dB.

The cancellation performance is assessed through the Echo Return Loss Enhancement (ERLE), defined as

$$\text{ERLE} = \frac{\sum y(n)^2}{\sum [y(n) - \hat{y}(n)]^2}, \tag{10}$$

and its average value is computed over 50 trials of the experiment, each one with different $x(n)$ and $h(n)$. The performance of the adaptive algorithms is measured in the steady-state after processing 3s of signals, using a test (i.e., a distinct from processed) input reference.

The proposed method is compared with (generalized) power filter echo cancelers from [20]. We consider their two variants that differ in their sets of basis functions used to approximate the nonlinear function $g(\cdot)$. Namely, (1) Monomial power filters (MPF) utilize x, x^2, and x^3. (2) Generalized power filters (GPF) use the same monomials together with $\tanh(x)$. The optimization of both variants is based on the multichannel RLS algorithm, which achieves good performance but is computationally expensive.

Variants of the proposed method are denoted as PM:A where A refers to the adaptive method to identify the linear subsystem (e.g. PM:NLMS). Next, PPM:RLS denotes the parallel implementation proposed in Section 2.3 using RLS (note that the linear AEC connected in parallel to the proposed method is always NLMS). We compare PM:NLMS with an "oracle" version denoted as OPM:NLMS which uses the unknown $a(n)$ to learn the parameters of the neural network.

3.1. Echo cancellation in the noiseless case

Results achieved in the noiseless situation, i.e. when $v(n) = 0$, with different nonlinearities $g(\cdot)$ are shown in Table 1. With the exception when $g(x(n)) = x(n)$, the proposed techniques achieve higher ERLE compared to MPF, GPF, and NLMS. The oracle method OPM:NLMS outperforms PM:NLMS only by about $2 - 3$ dB, which points to the efficiency of the estimation of $a(n)$ within the proposed methods.

Considering the linear channel $g(x(n)) = x(n)$, the proposed method is significantly outperformed by MPF, GPF as well as by NLMS. The difference in performance is compensated by the parallel variant PPM:RLS that treats the linear case in its parallel branch.

Using speech input reference, the achieved ERLE is significantly lower than when the signal is white noise. There are two facts that explain this phenomenon. First, speech usually contains only a small number of samples whose amplitude is high. Therefore, the identification of $g(\cdot)$ is difficult as its

reference signal $[g(x)]$	OPM: NLMS	PM: NLMS	PM: APA	PM: RLS	PPM: RLS	NLMS	MPF	GPF
white noise $[\tanh(5x)]$	**32.0**	30.4	30.1	30.4	30.4	6.7	9.9	14.5
speech $[\tanh(5x)]$	16.8	13.4	16.4	**18.2**	**18.2**	6.8	11.8	14.9
white noise $[x(n)]$	31.3	29.7	29.5	29.7	105.4	105.6	**107.8**	82.6
speech $[g_3(\cdot)]$	11.2	8.3	11.5	11.4	**12.5**	4.4	9.8	10.6

Table 1. Averaged ERLE [dB] achieved in the noiseless scenario

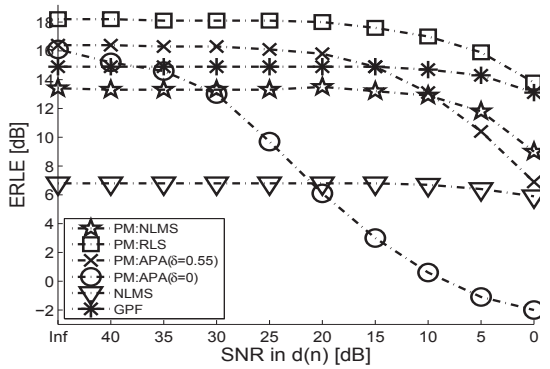

Fig. 4. ERLE [dB] achieved in scenario considering nonlinearity $g_1(\cdot)$ and speech reference

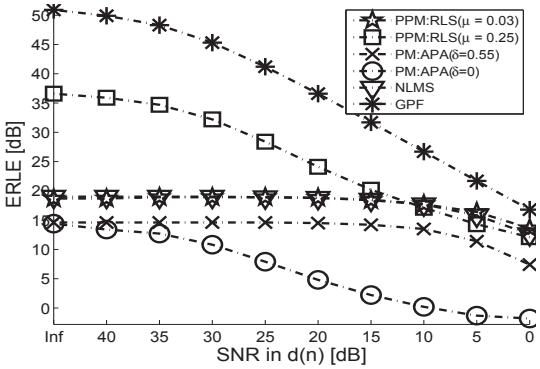

Fig. 5. ERLE [dB] achieved in scenario considering linear channel $g_2(x(n)) = x(n)$ and speech reference

PM:NLMS	PM:APA	PM:RLS	NLMS	GPF
11.2	**12.2**	10.0	7.7	10.9

Table 2. The ERLE [dB] achieved in the real-world scenario

values $g(x)$ when $|x|$ is "high" are rarely available. Second, speech is sparse in the frequency domain, so not all frequencies are sufficiently excited. The latter fact is known to cause slow convergence of the NLMS algorithm [13], which also confirm the results in in Table 1. All the algorithms utilizing NLMS achieve lower ERLE compared to the variants based on RLS or APA.

3.2. Echo cancellation in the noisy case

Results of experiments with additive gaussian noise are presented in Figures 4 and 5 as functions of the signal-to-noise ratio (SNR) measured in $d(n)$. The proposed method appears to be robust to the presence of the noise until SNR$>$ 10 dB. MPF/GPF are optimal for the scenario $g(x(n)) = x(n)$, achieving higher ERLE than the proposed method. The slower convergence (and thus lower ERLE) of NLMS can be mitigated by selecting higher step size μ (e.g., $\mu = 0.25$ for NLMS within the parallel branch of PPM:RLS, see in Fig. 5). However, unsuitably large values of μ can cause problems with the convergence of the adaptive filter. The regularization is essential for NLMS/APA algorithms when speech input is considered; see the performance drop of the PM:APA and $\delta = 0$ in Fig. 5.

4. REAL-WORLD EXPERIMENT

In this experiment, far-end speech was played by a loudspeaker in a room and recorded by a microphone at a distance of about 15 cm. The reverberation time of the room is $T_{60} \approx 490$ ms; the length of the recording is 8 s. After applying the compared methods, using the same settings as in Section 3, the ERLE quantity was measured directly at the output of the NAEC during the adaptation process. The average ERLE is shown in Table 2.

The nonlinear cancelers achieve higher ERLE compared to NLMS; the PM:APA achieves the best result. The lower performance of algorithms based on RLS (GPF and PM:RLS) could be caused by the high value of the forgetting factor $\lambda = 0.9999$, which yields stable convergence at the cost of a slow adaptation.

5. CONCLUSIONS

Compared to robust generalized power filter cancelers, the proposed method achieves a higher ERLE on white noise inputs and comparable performance with speech signals. The proposed technique is robust with respect to noise up to an SNR level 10 dB, provided that suitable regularization is applied to the adaptive filter identifying the parameters of the linear subsystem.

REFERENCES

[1] E. Hänsler, G. Schmidt, *Acoustic Echo and Noise Control: A Practical Approach.*, Vol. 40., John Wiley & Sons, 2005.

[2] T. W. S. Chow, H. Z. Tan, Y. Fang, *Nonlinear System Representation*, John Wiley and Sons, Inc., 2001.

[3] G. M. Raz and B. D. Van Veen. "Baseband Volterra Filters for Implementing Carrier based Nonlinearities," *IEEE Transactions on Signal Processing*, vol. 46, no. 1, pp. 103-114, 1998.

[4] G. Picard and O. Cappe, Blind Identification of Hammerstein Nonlinear Distortion Models, in *Proc. WASPAA 2003*, pp. 17-20, 2003.

[5] R. Rojas, *Neural Networks: A Systematic Introduction*, Springer Science & Business Media, 2013.

[6] G. Alexandre, G. Faucon, and L. B. Jeannes. "Nonlinear Acoustic Echo Cancellation based on Volterra Filters," *IEEE Transactions on Speech and Audio Processing* vol. 11, no. 6, pp. 672-683, 2003.

[7] F. Kuech, W. Kellermann, "Orthogonalized Power Filters for Nonlinear Acoustic Echo Cancellation.," *Signal Processing*, vol. 86, no. 6, pp. 1168-1181, 2006.

[8] A. N. Birkett and R. A. Goubran., "Acoustic Echo Cancellation Using NLMS-neural Network Structures," in *Proc. ICASSP 1995*, pp. 3305-3308, Detroit, USA, 1995.

[9] L. S. Ngia, and J. Sjobert, "Nonlinear Acoustic Echo cancellation Using a Hammerstein Model," in *Proc. ICASSP 1998*, pp.1229-1232, Seattle, USA, 1998.

[10] A. Janczak, Identification of Nonlinear Systems Using Neural Networks and Polynomial Models: A Block-Oriented Approach, *Springer Science & Business Media*, Vol. 310, 2004.

[11] D. Comminiello, "Functional Link Adaptive Filters for Nonlinear Acoustic Echo Cancellation." *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 21, no.7, pp 1502-1512, 2013.

[12] S. S. Haykin, *Adaptive Filter Theory*, 4th edition, Pearson Education India, 2007.

[13] S. L. Gay, *The Fast Affine Projection Algorithm*, Springer US, 2000.

[14] J. Gunther, Learning Echo Paths During Continuous Double-Talk Using Semi-Blind Source Separation, *IEEE Trans. on Audio, Speech, and Language Processing*, vol. 20, no. 2, Feb. 2012.

[15] Z. Koldovský, J. Málek, M. Müller, and P. Tichavský, "On Semi-Blind Estimation of Echo Paths During Double-Talk Based on Nonstationarity," *Proc. of the 14th International Workshop on Acoustic Signal Enhancement (IWAENC 2014)*, pp. 199–203, Antibes, France, Sept. 2014.

[16] M. F. Moller, "A Scaled Conjugate Gradient Algorithm for Fast Supervised Learning." *Neural networks,* vol. 6, no. 4, pp. 525-533, 1993.

[17] A. H. Sayed and Thomas Kailath. "Recursive Least-Squares Adaptive Filters." The Digital Signal Processing Handbook, vol. 21, no. 1, 1998.

[18] M. Stewart, "A Superfast Toeplitz Solver with Improved Numerical Stability," *SIAM Journal on Matrix Analysis and Applications*, vol. 25, no. 3, pp. 669-693, 2003.

[19] J. Benesty, C. Paleologu, and S. Ciochina, "On Regularization in Adaptive Filtering," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 19, no. 6, pp. 1734-1742, 2011.

[20] J. Malek, and Z. Koldovský. "Nonlinear Echo Cancellation Using Generalized Power Filters." *in Proc. IEEE ECMSM 2015*, pp. 1-6, Liberec, Czech Republic, 2015.