Nonlinear Echo Cancellation using Generalized Power Filters

Jiri Malek and Zbyněk Koldovský
Faculty of Mechatronics, Informatics, and Interdisciplinary Studies,
Technical University of Liberec,
Studentská 2, 461 17 Liberec, Czech Republic.
E-mail: jiri.malek@tul.cz, zbynek.koldovsky@tul.cz

***Abstract*—This paper deals with the problem of nonlinear echo cancellation. One of the established approaches to this task is that of adaptive filtering using power filters. These filters model the unknown nonlinear system as a set of monomials, followed by a set of linear systems. The monomials are monotonously increasing functions useful for modeling nonlinearities such as soft-clipping, but only to a limited extent. We propose to replacing the monomials by a set of basis functions which resemble known real-world nonlinear distortions. Next, we propose performing the identification using the Multichannel Recursive Least Squares algorithm instead of the single-channel Normalized Least Mean Squares. In this way, the method can avoid the orthogonalization procedure used in the conventional power filter, thereby improving the reliability of the algorithm. We demonstrate the performance of the proposed algorithm in experiments with artificial and real-world data and compare it to several state-of-the-art methods.**

I. Introduction

The acoustic echo often arises in the context of duplex communication, for example, during phone calls. Here, speech originating from the far-end is played by loudspeaker(s). Subsequently, the emitted signal is convolved with the acoustic room impulse response and the modified signal, called echo, arrives at the microphone(s). The echo is transmitted back to the far-end speaker, along with the near-end speech [1]. This phenomenon deteriorates the intelligibility of the transmitted speech.

The process of removal of the acoustic echo is called Echo Cancellation [1]. Most often, the influence of the acoustic environment is modeled through a linear filter, which needs to be identified using the known far-end signal as a reference. The output of the acoustic echo canceler (AEC), i.e., filtered far-end signal, is then subtracted from the microphone signal prior to its transmission back to the far-end user. The acoustic echo paths can change in time. The filter should, therefore, be adapted to those changes.

One of the most important parameters of modern communication devices is their price. To keep the prices low, the utilization of low-cost components, which exhibit nonlinear characteristics, is necessary. For example, amplifiers operating near saturation or miniaturized loudspeakers are potential sources of nonlinear distortions. This complicates the echo cancellation, because the combination of electronic and acoustic echo paths needs to be modeled as a nonlinear system [2].

Nonlinear systems can either have memory or be memoryless (instantaneous). A memoryless system is encountered, e.g., in the context of low-cost power amplifiers or loudspeakers [3]. It is usually described as a nonlinear function applied sample-wise on the processed signal. Systems with memory are usually described through the nonlinear auto-regressive moving average (NARMA) model or as the Voltera filters; see, e.g., [5]. The identification of a full Voltera system is complicated from computational and algorithmic point of view. Therefore, many simplified models are used in practice, such as the Hammerstein model [6] (an instantaneous nonlinearity followed by an all-pole linear system) or the LTI-ZMNL-LTI model [7] (Linear Time Invariant - Zero Memory Nonlinear, i.e., two linear subsystems separated by an instantaneous nonlinearity).

There are three main groups of methods for removal of nonlinear acoustic echo. The first group is called the predistortion linearization [2]. Here, the assumed nonlinear distortion is compensated in the electric domain, i.e., before the far-end signal is emitted through loudspeakers. An advantage of this approach is that the compensation of the nonlinear system, caused by the electronic components, and the linear acoustic system, caused by the environment, are solved separately. The latter can be performed using a conventional linear echo canceler. A disadvantage is the necessity to modify (and distort) the far-end signal transmitted through the loudspeaker, while it remains unchanged within the other approaches.

The algorithms of the second group attempt to identify the electric and acoustic parts simultaneously, as a single nonlinear system with memory (using models described above). No modifications are required for the far-end signal played through the loudspeaker. The methods are referred to as Nonlinear Acoustic Echo Cancelers (NAEC). These adaptive cancelers can be derived using a deterministic filter design (see methods in [3], [8], [9], [10]) as well as a stochastic filter design (see, e.g., [11]). The advantage of these approaches is their efficiency, which is achieved without any distortion to the near-end speech. The disadvantage is high computational complexity and possible divergence of the adaptive filters.

The methods of the latter group are referred to as Nonlinear Residual Echo Suppressors (NRES). These approaches use conventional linear echo cancelers to suppress the linear part of the echo. The residual nonlinear echo is then suppressed using various types of frequency masking; an accurate identification of the nonlinear system is not necessary [12], [13]. The NRES methods are computationally inexpensive and achieve very good levels of echo removal. However, they often cause distortions of the near-end signal to be transmitted.

A specific type of NAEC is called power filter [3], [4]. It consists of a memoryless nonlinearity and a set of linear systems. The nonlinearity is approximated as a polynomial (usually of order three). Each monomial from the polynomial is followed by a Finite Impulse Response (FIR) filter. The identification of the FIR filters is performed separately, using the single-channel Normalized Least Mean Square (NLMS, [14]) adaptive algorithm. To improve the convergence, the channels are mutually orthogonalized using the Gram-Schmidt process [15].

In this paper, a modified version of the power filter operating in the time-domain is proposed. We express the memoryless nonlinearity as a sum of basis functions (a similar model of memoryless nonlinearity is utilized in [11]). We select the basis functions as models of known nonlinear distortions encountered in low-cost electronic components, such as soft-clipping [16], [17]. The proposed model leads to an improved echo cancellation performance in cases where the nonlinearity corresponds to a certain type of saturation. This saturation is difficult to model using monomials. To improve the robustness and convergence speed of the algorithm, we propose performing the identification using multichannel Recursive Least Square algorithm (MRLS, [18]). In this manner, we avoid the Gram-Schmidt orthogonalization used within the conventional power filters.

II. Proposed power filters in the time-domain

A. Problem formulation

In the context of acoustic echo cancellation (see Fig. 1), the AEC 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 the signal $d(n)$ observed on the microphone. The echo signal $y(n)$ is a version of the far-end speech signal $x(n)$ modified by electronic components, such as amplifier or loudspeaker, and by the acoustic path. The measured signal $d(n)$ consists of echo $y(k)$, background noise $v(n)$ and near-end speech signal $z(n)$.

The system identification is performed only during intervals where the near-end signal is not active, i.e., $z(n) = 0$ and $d(n) = v(n) + y(n)$. Note that the presence of near-end speech significantly complicates the identification. The only way to perform the identification during activity of the near-end signal is by using Blind Source Separation (BSS) methods; see, for example, [19] and [20].

B. Algorithm description

In this paper, we consider the nonlinear echo path model depicted in Fig. 2, i.e., the LTI-ZMNL-LTI model, which consists of a memoryless nonlinearity $g(\cdot)$ that is applied in between two linear systems $w(n)$ and $c(n)$. The system $w(n)$ and the function $g(\cdot)$, respectively, model the linear and nonlinear effects of electronic devices, such as the D/A converter, amplifier and loudspeaker. The system $c(n)$ represents the influence of the acoustic environment.

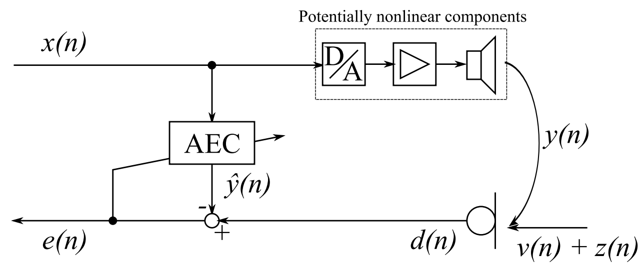

Fig. 1. The process of Acoustic Echo Cancellation

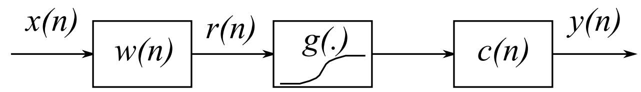

Fig. 2. The considered nonlinear echo path

To estimate the output $y(n)$ of the cascade in Fig. 2, we propose the structure of the power filter, which is shown in Fig. 3. The output relation using the reference signal $x(n)$ as its input is

$$\hat{y}(n) = \sum_{p=1}^{P} \sum_{k=0}^{L} h_p(k) x_p(n-k), \tag{1}$$

where $h_p(k)$ are impulse responses of P linear filters of length L and

$$x_p(n) = b_p(x(n)), \qquad p = 1, \ldots, P, \tag{2}$$

where $b_p(\cdot)$ are suitable basis functions, which should approximate the unknown memoryless nonlinearity $g(\cdot)$ according to

$$g(\cdot) \approx \sum_{p=1}^{P} b_p(\cdot). \tag{3}$$

The selection of the basis function will be discussed in the next section.

The nonlinear model with memory described by (1) is an approximation of the echo path considered in Fig. 2. The equivalence happens only if $w(n)$ is a pure delay or $g(\cdot)$ is a linear function. For details on the approximation see [3]. If the basis functions $b_p(\cdot)$ are polynomials, the power filter corresponds to the diagonal form of the Voltera filter; see [21].

In matrix form, (1) can be rewritten as

$$\hat{y}(n) = \mathbf{h}^T \mathbf{x}(n), \tag{4}$$

where

$$\mathbf{h} = [h_1(0), \ldots, h_1(L-1), \ldots, h_P(L-1)]^T, \tag{5}$$

$$\mathbf{x}(n) = [\mathbf{x}_1(n), \ldots, \mathbf{x}_1(n-L+1), \ldots, \mathbf{x}_P(n-L+1)]^T. \tag{6}$$

Now, the task is to identify the unknown filters $h_p(n)$ so that $\hat{y}(n)$ is as similar to $y(n)$ as possible in the least squares sense. Let $\widehat{h}_p(k,n)$ denote the taps of the estimate of the pth filter at time index n, which means that

$$\widehat{\mathbf{h}}(n) = [\widehat{h}_1(0,n), \ldots, \widehat{h}_1(L-1,n), \widehat{h}_P(L-1,n)]. \tag{7}$$

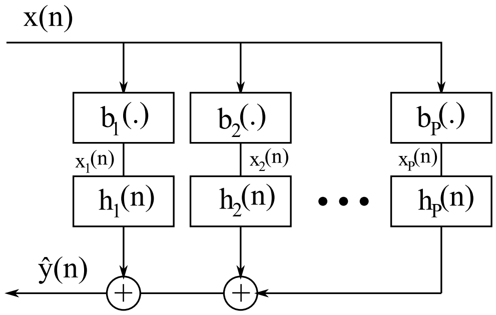

Fig. 3. The structure of the proposed power filter

The identification is performed through minimization of the energy of the error signal given by

$$e(n) = d(n) - \widehat{\mathbf{h}}^T(n)\mathbf{x}(n). \tag{8}$$

We apply the Multichannel Recursive Least Squares (MRLS) algorithm [18]. The filter update is given by

$$\widehat{\mathbf{h}}(n) = \widehat{\mathbf{h}}(n-1) + \mathbf{R}^{-1}(n)\mathbf{x}(n)e(n), \tag{9}$$

where $\mathbf{R}$ is the estimate of the cross-channel covariance matrix. The inverse of $\mathbf{R}$ can be computed recursively as

$$\begin{aligned}\mathbf{R}^{-1}(n) &= \lambda^{-1}\mathbf{R}^{-1}(n-1) \\ &\quad - \frac{\lambda^{-2}\mathbf{R}^{-1}(n-1)\mathbf{x}(n)\mathbf{x}(n)^T\mathbf{R}^{-1}(n-1)}{1+\lambda^{-1}\mathbf{x}(n)^T\mathbf{R}^{-1}(n-1)\mathbf{x}(n)}.\end{aligned} \tag{10}$$

C. Choice of basis functions

In [3], the basis functions $b_p(\cdot)$ are selected as monomials, which, in fact, means that $g(\cdot)$ is being approximated through a truncated Taylor series (typical order is $P = 3$).

We propose to choose such functions that model certain known nonlinearities encountered in loudspeaker systems. In this paper, the following set is used

$$b_p(x(n)) = \begin{cases} x^p(n) & \text{if } p = 1..3 \\ \tanh(5x(n)) & \text{if } p = 4. \end{cases} \tag{11}$$

Note that $b_4(\cdot)$ simulates the soft-clipping; for other parametric models see, e.g., [17].

Our motivation behind this approach is the fact that the monomial basis functions approximate memoryless nonlinearities containing saturation only to a limited extent, because monomials are monotonically increasing functions.

D. Adaptive algorithm

The conventional power filters perform the adaptation of $h_p(n)$ through single-channel Normalized Least Mean Squares (NLMS) updates [4], [14]. The updates are performed one filter at a time, with shared error $e(n)$. The signals $x_p(n)$ are correlated, which would lead to a slow convergence to an optimum solution. To mitigate this drawback, approximate mutual orthogonalization of the channels is performed [4]. This procedure is based on Gram-Schmidt [15] method, however, the transform changes in time because of the nonstationarity of the speech signal.

The proposed optimization via MRLS allows to omit the approximate orthogonalization because the algorithm takes the full channel-cross-correlation into account via the utilization of the matrix $\mathbf{R}$ in (9). Moreover, due to the simultaneous update of $h_p(n)$, the MRLS algorithm converges more quickly and precisely to an optimum solution than the single-channel NMLS does, as will be seen in our experiments and as is described in [23].

The experiments shown in this paper use the original definition of the MRLS algorithm, i.e., equations (8)-(10). The algorithm has high complexity of $O(P^2L^2)$. Several fast calculation schemes exist which significantly reduce the computational burden and are thus suitable to be implemented for hand-held devices. For example, the Fast MRLS algorithm, as described in [22], exhibits the complexity $O(6P^2L+2PL)$. In theory, the fast algorithm is mathematically equivalent to MRLS. In practice though, this property is lost because of finite precision effects in the implementation, which could lead to worse performance compared to the original MRLS.

E. Batch version of the power filters

In addition to the adaptive version of the power filters, we also implement a batch version which analyzes the signal $y(n)$ as a whole, not sample by sample as within the adaptive methods. Comparison with the performance of the batch version will verify whether the adaptive power filters converge to the optimal filter in the least square sense.

The filter estimated by the batch version of the power filter is computed as

$$\widehat{\mathbf{h}}_b = \arg\min_{\mathbf{h}} \|\mathbf{X}\mathbf{h} - \mathbf{D}\|_2^2, \tag{12}$$

where

$$\mathbf{D} = [d(1), \ldots, d(N)], \tag{13}$$

and N is the length of $d(n)$. Further, $\mathbf{X}$ is an $N \times PL$ matrix given by

$$\mathbf{X} = [\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_P], \tag{14}$$

where $\mathbf{X}_p$ is an $N \times L$ Toeplitz matrix whose first column is $[x_p(1), x_p(2), \ldots, x_p(N)]^T$ and the first row is $[x_p(1), 0, \ldots, 0]$. The solution of (12) is

$$\widehat{\mathbf{h}}_b = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{D}. \tag{15}$$

III. Experimental Evaluation

The following section describes our experimental comparison of the echo cancellation performed by the proposed power filters and by several other conventional methods. The competing algorithms are summarized in Table I. Note that the echo suppressors need the output of the linear echo canceler as their input. To this end, we utilize the output of the NLMS algorithm.

The free parameters of the algorithms are set to the values summarized in Table II. The variable names correspond to those in the respective papers; see the references for details. Several suitable parameter settings were experimentally tested in order to ensure the best functionality of these algorithms.

The speech signals used for experimental purposes originate from the TIMIT database [24]. The far-end speech is

TABLE I. ECHO CANCELLATION METHODS COMPETING IN OUR EXPERIMENTS

Algorithm	Abbreviation
Linear Normalized Least Mean Square [14]	NLMS
Nonlinear Echo Suppressor (Osamu, [12])	NAESO
Nonlinear Echo Suppressor (Shi, [13])	NAESS
Original Power Filter [3]	OPF
Proposed Power Filter, Monomial $b_p(\cdot), p = 1 \ldots 3$, Batch version	PPF_MB
Proposed Power Filter, Generalized $b_p(\cdot)$ given in (11), Batch version	PPF_GB
Proposed Power Filter, Monomial $b_p(\cdot), p = 1 \ldots 3$, Adaptive version	PPF_MA
Proposed Power Filter, Generalized $b_p(\cdot)$ given in (11), Adaptive version	PPF_GA

TABLE II. VALUES OF THE FREE PARAMETERS USED IN THE COMPETING METHODS.

Algorithm	Meaning	Parameter value
NLMS	Adaptation step-size	$\nu = 0.2$
NAESO	FFT length	2048 samples
NAESO	FFT shift	128 samples
NAESO	Other parameters	Default values as in [12]
NAESS	FFT length	512 samples
NAESO	FFT shift	256 samples
NAESO	Forgetting factor	$\lambda = 0.95$
NAESO	Degree of polynomial nonlinearity	$K = 6$
OPF	Forgetting factor - estimation of moments	$\lambda = 0.9999$
OPF	Adaptation step-sizes	$\nu_p = [0.1, 0.15, 0.15],\ p = 1 \ldots 3$
PPF	Forgetting factor in (10)	$\lambda = 0.9999$

a female utterance; the near-end speech is a male utterance; both signals are 9 s in length. To create the echo, the utterance is either processed through an artificial nonlinear system, as described in Section III-A, or played by low-cost loudspeakers and recorded, as reported in Section III-B. First, three seconds of the signal are used for identification of the filters $h_p(n)$; the near-end speech is silent. The remaining six seconds are used to evaluate the cancellation when the near-end speech is active and the adaptation of the filters is stopped.

The echo cancellation performance is measured through Echo Return Loss Enhancement (ERLE) defined as

$$\text{ERLE} = \frac{\sum y(n)^2}{\sum [y(n) - \hat{y}(n)]^2}. \tag{16}$$

The nonlinear echo suppressors cause distortion within the near-end signal. To quantify this effect we compute the Signal-to-Distortion-Ratio given by

$$\text{SDR} = \frac{\sum z^2(n)}{\min_\alpha [\sum (z(n) - \alpha \widehat{z}(n))^2]}, \tag{17}$$

where $\widehat{z}(n)$ is the near-end signal at the output of the corresponding processing technique, and α is a scaling parameter to eliminate the difference in magnitude between the input and output. The SDR is presented only in the real-world scenario.

A. Experiments with artificial data

We consider two different experimental scenarios in this section. 1) The linear filter is $w(n) = \delta(n)$ (the unit impulse) in the echo path model in Fig. 1, i.e., the model of the power filter can match the true echo path. 2) The linear filter $w(n)$ is a random FIR filter of the length $L_w = 64$, i.e., the model of the power filter approximates the true echo path only. In both scenarios, the linear filter $c(n)$ is generated by the technique presented by Lehmann in [25] for generation of Room Impulse Responses. The source-microphone distance is 1 m. The filter corresponds to a reverberation time of 200 ms and is truncated to a length of $L_c = 128$ taps. The nonlinearity $g(r(n)) = \tanh(3r(n))$, i.e., a function resembling the soft-clipping. This nonlinearity is similar to the basis function $b_4(\cdot)$ used in the proposed method PPF_GA but it is not the same. For both scenarios we consider a noise-free case, i.e., $v(n) = 0$, or a noisy scenario where the Signal-to-Noise Ratio (SNR) between $y(n)$ and $v(n)$ is 35 dB. The length of estimated filters is $L = 128$ taps for the matching model scenario and $L = 192$ for the approximate model scenario.

Matching model scenario: The results of the experiments are presented in Fig. 4. It can be seen that all nonlinear techniques for echo removal perform better than the conventional linear echo canceler. The performance of the echo suppressing techniques is comparable to or worse than that of the proposed power filters. They benefit from the fact that their model match the true echo path. Significant distortions are audible in the near-end speech at the output of the echo suppressors, especially in the case of NAESS. The results indicate that the MRLS algorithm used within the PPF methods converges more precisely to the desired filters $h_p(n)$ than the NLMS algorithm along with channel orthogonalization used in the OPF. The presence of the basis function $b_4(n)$ in the PPG_GA method improves the performance significantly compared to the PPF_MA method (by 15 dB). It can be also seen that the performance of the adaptive algorithms is very close to the batch implementation, which indicates that the convergence of the adaptive filters is good. The presence of noise deteriorates the convergence of the NLMS-based algorithms, whereas the RLS-based algorithms are robust with respect to the given noise level.

Approximate model scenario: The results of the experiments are shown in Fig. 5. As expected, the performance of the echo cancelers is worse than in the previous experiments, due to the mismatched model. The performance of the PPF_GA method is now comparable to the results of the suppressors (which, however, distort the near-end speech). The presence of the $b_4(\cdot)$ basis function improves the performance of the PPF_GA method compared to the conventional PPG_MA algorithm by almost 4 dB.

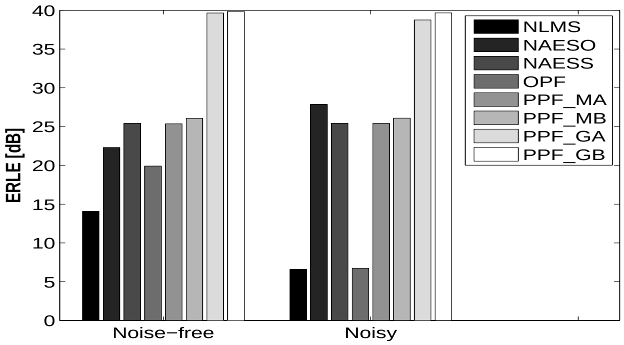

Fig. 4. Experimental results for the Matching Model Scenario.

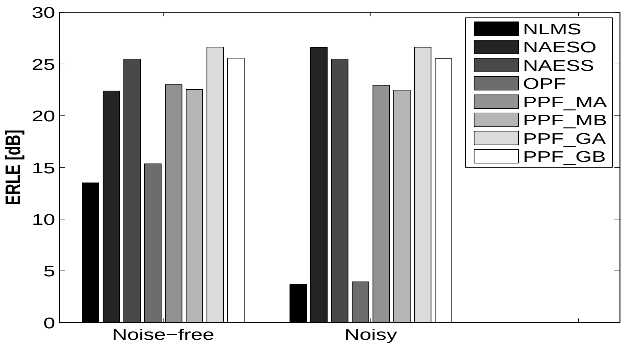

Fig. 5. Experimental results for the Approximate Model Scenario.

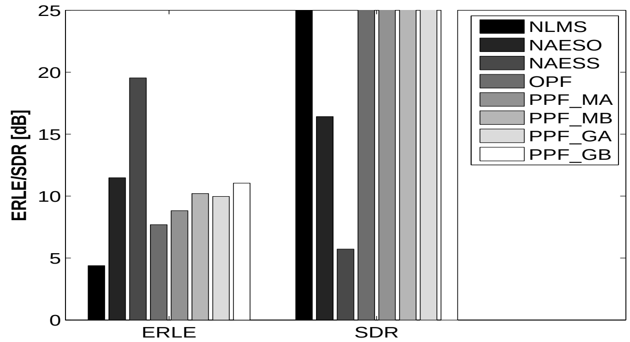

Fig. 6. Experimetal results for the real-world data.

B. Experiments with real-world data

In this setting, the far-end speech was emitted through a low-cost loudspeaker and recorded on a microphone. The microphone-speaker distance was about 15 cm. The recording was performed in a room with a reverberation time of $T_{60} = 490$ ms. The results of the experiment are shown in Fig. 6.

The PPF_GA algorithm yields a 1.5 dB higher performance than the PPF_MA. This indicates that the inclusion of basis function $b_4(\cdot)$ allows more precise modeling of real-world nonlinearities compared to a purely monomial set of basis functions. The proposed version of power filters achieves an ERLE value by 2.5 dB higher than the conventional original version OPF. The echo suppressor NAESS outperforms the cancelers by at least 9 dB. However, the output near-end speech of NAEES is significantly distorted, as is reflected by the low SDR value. None of the cancelers (LMS, all power filters) distorts the near-end speech (the computed SDR value is 250 dB, which corresponds to the numerical precision).

IV. Conclusions

We presented an adaptive method for nonlinear echo cancellation. The method exhibits an echo cancelling performance comparable to state-of-the-art echo suppressors and does not cause distortion in the near-end speech. The convergence of the proposed method is robust with respect to background noise present in the recorded signal. The proposed method yields improved performance in the sense of ERLE, compared to original power filters on artificial as well as real-world signals. Future work should focus on the alleviation of high computational demands, via reliable incorporation of the Fast MRLS algorithm and investigation of suitable basis functions based on the real-world speech signals recorded from low-cost loudspeakers.

Acknowledgment

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

References

[1] E. Hänsler, G. Schmidt, *Acoustic echo and noise control: a practical approach.*, Vol. 40., John Wiley & Sons, 2005.

[2] K. Shi, *Nonlinear acoustic echo cancellation*, Dissertation Thesis, Georgia Institute of Technology, 2008.

[3] F. Kuech, W. Kellermann, "Orthogonalized power filters for nonlinear acoustic echo cancellation.," *Signal Processing*, vol. 86, no. 6, pp. 1168-1181, 2006.

[4] F. Kuech, A. Mitnacht, W. Kellermann, "Nonlinear acoustic echo cancellation using adaptive orthogonalized power filters." in *Proc. ICASSP 2005*, Phyladelphia, USA, 2005.

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

[6] G. Picard and O. Cappe, Blind identification of hammerstein nonlinear distortion models, in Proc. WASPAA 2003, pp. 17-20, 2003.

[7] S. Prakriya and D. Hatzinakos, Blind identification of nonlinear models using higher order spectral analysis, in *Proc. ICASSP 1995*, pp. 1601-1604, 1995.

[8] G. Alexandre, G. Faucon, and L. B. Jeannes. "Nonlinear acoustic echo cancellation based on Volterra filters," *Speech and Audio Processing, IEEE Transactions on* vol. 11, no. 6, pp. 672-683, 2003.

[9] Ch. Hofmann, Ch. Huemmer, W. Kellermann. "Significance-aware Hammerstein group models for nonlinear acoustic echo cancellation," in *Proc. ICASSP 2014*, Florence, Italy, 2014.

[10] S. V. Vaerenbergh, L. A. Azpicueta-Ruiz. "Kernel-based identification of Hammerstein systems for nonlinear acoustic echo-cancellation," in *Proc. ICASSP 2014*, Florence, Italy, 2014.

[11] S. Malik, G. Enzner, "State-space frequency-domain adaptive filtering for nonlinear acoustic echo cancellation." *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 20, no. 7, pp 2065-2079, 2012.

[12] H. Osamu, A. Sugiyama, "An acoustic echo suppressor based on a frequency-domain model of highly nonlinear residual echo," in *Proc. ICASSP 2006*, Toulouse, France, 2006.

[13] K. Shi, M. Xiaoli, and G. T. Zhou, "A residual echo suppression technique for systems with nonlinear acoustic echo paths," in *Proc. ICASSP 2008*, Las Vegas, USA, 2008.

[14] S. S. Haykin, *Adaptive filter theory, 4th edition*, Pearson Education India, 2007.

[15] A. Papoulis, S. U. Pillai, *Probability, random variables, and stochastic processes*, Tata McGraw-Hill Education, 2002.

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

[17] S. B. Nollett, S. Bryan, L. D. Jones. "Nonlinear echo cancellation for hands-free speakerphones," *Urbana*, vol. 51, 1997.

[18] J. Benesty et al., *Adaptive algorithms for MIMO acoustic echo cancellation*, Springer US, 2004.

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

[20] Z. Koldovský, J. Málek, M. Müller, and P. Tichavský, "On Semi-Blind Estimation of Echo Paths During Double-Talk Based on Nonstationarity," Proc. IWAENC 2014, pp. 199-203, Antibes Juan les Pins, France, Sept. 2014.

[21] G. M. Raz, B. D. Van Veen, "Baseband Volterra filters for implementing carrier based nonlinearities," *Signal Processing, IEEE Transactions on*, vol. 46, no. 1, pp. 103-114, 1998.

[22] M. G. Bellanger, Adaptive Digital Filters and Signal Analysis. Marcel Dekker, New York, 1987.

[23] J. Benesty, et al. "Adaptive filtering algorithms for stereophonic acoustic echo cancellation, " in *Proc. ICASSP 1995*, Detroit, USA, 1995.

[24] J. S. Garofolo, et al., "TIMIT Acoustic-Phonetic Continuous Speech Corpus," Linguistic Data Consortium, Philadelphia, 1993.

[25] E. Lehmann and A. Johansson, "Prediction of energy decay in room impulse responses simulated with an image-source model," *Journal of the Acoustical Society of America*, vol. 124, no. 1, pp. 269-277, 2008.