Combined Linear and Nonlinear Residual Echo Suppression Using a Deficient Distortion Model — A Proof of Concept

Ingo Schalk-Schupp[1], Friedrich Faubel[1], Markus Buck[1], Andreas Wendemuth[2]

[1]Acoustic Speech Enhancement Research, Nuance Communications Deutschland GmbH, 89077 Ulm, Germany
[2]Otto von Guericke University Magdeburg, Chair of Cognitive Systems, 39016 Magdeburg, Germany
[1]Email: Ingo.Schalk-Schupp@nuance.com, [2]Email: andreas.wendemuth@ovgu.de

Abstract

Acoustic echo control is still challenged with finding computationally efficient and accurate methods for handling nonlinear echo components introduced by cheap loudspeaker components and high signal levels.

We propose a robust echo control scheme that utilizes a common linear acoustic echo canceler as well as a combined linear and nonlinear residual echo suppression unit. Linear and nonlinear residual echo powers are estimated separately making use of the same room impulse response estimate provided by the linear canceler. Frequency-dependent coupling factors between the observed and the respective estimated residual echo power are used to account for deviations due to varying convergence states of the linear filter or the nonlinear model. A weighted sum of the two power estimates is used in a sub-band-domain suppression filter.

This contribution compares the potential of the proposed concept to a Hammerstein-type nonlinear acoustic echo canceler under fair simulation conditions.

1 Introduction

Embedded devices with an audio interface usually have a loudspeaker (LS) unit comprising a digital-to-analog converter (DAC), followed by or integrated with an electric amplifier driving the physical loudspeaker. When this unit is utilized outside the bounds of its specification, or if cheap components are used, then the LS unit may expose nonlinear behavior. In this case, the digital signal sent to the unit – called the reference signal – is distorted before traversing the device's enclosure and being received by the device's microphone unit.

A commonly used model for this type of behavior is the Hammerstein model. It describes a device's echo path, i.e., the system between the digital reference signal output and the digital microphone signal input, as a cascade of a memoryless nonlinear system (the distortion function f) and a linear system.

1.1 Notational Note

In this document, we will refer to time-domain signals, both continuous and sampled, with lowercase letters, while their short-time Fourier transform (STFT) counterparts will be referred to with the corresponding capital letter. We will mostly omit dependencies on time t, sample index n, sub-band index k and frame index ℓ for the sake of notational simplicity.

1.2 Hammerstein Echo Path Model

From the processing unit's point of view, the echo path, or loudspeaker–enclosure–microphone system (LEMS), comprises the device's acoustic environment as well as the loudspeaker unit and the microphone unit as seen in figure 1.

It transforms the reference signal x into the microphone signal y, which is composed of echo (d), local noise (b) and local speech (s) components additively:

$$y = d + b + s. \tag{1}$$

To good approximation, the amplifier may distort the analog time-continuous reference signal values $x(t)$ by a sigmoid distortion function. This may be described by an odd-symmetric function f with which the distorted signal x_{dis} can be expressed as

$$x_{\text{dis}} = f(x). \tag{2}$$

The distorted signal traverses the environment and reaches the microphone. Also, signal portions are reflected and/or scattered back into the microphone as the echo signal d. This behavior is commonly modeled with a convolution of the signal with a transversal filter called the room impulse response (RIR) h:

$$d = h * x_{\text{dis}}. \tag{3}$$

A Hammerstein-type nonlinear acoustic echo cancellation (NLAEC) is constructed in a corresponding fashion with estimates for the distortion function and the RIR:

$$\hat{d} = \hat{h} * \hat{f}(x). \tag{4}$$

1.3 State of the Art

There are several approaches for nonlinear echo cancellation, which estimate the distortion function f, in addition to the RIR. This is usually achieved by assuming a parametric model for the estimated distortion function $\hat{f}_{\underline{\theta}}$ and then estimating the parameter vector $\underline{\theta}$. The estimated function is then applied to the time-domain reference signal and the result is convolved with the estimated RIR as in linear acoustic echo cancellation (AEC).

Hammerstein-group model [1, 2] approaches use a linear combination of basis functions to form the distortion function. This type of bases include power filter models [3], orthogonalized versions thereof [4], odd Fourier series [5], and filter combinations [6]. For the identification of the coefficients, each basis function branch contains a linear RIR estimator. Typical model orders range from 5 to 9 such branches.

Some nonlinear echo suppression schemes also use that type of model [7], while others use a harmonic model [8] or a neural network [9] to detect and suppress relevant nonlinear echo components. Another idea is to analyze correlations between basis functions applied on the reference signal in the STFT domain [10].

1.4 Motivation

While phase information is inherently important in echo cancellation, it is not used for echo suppression. This gives rise to the idea that distortion function estimation errors could be more detrimental to the results of echo cancellation than to echo suppression. On the other hand, suppression cannot attain perfect echo cancellation, as it constitutes a compromise between speech quality and the amount of suppression.

Nonlinearly distorting a signal introduces harmonics and intermodulation distortions. Two different odd-symmetric saturation-type distortion functions introduce the same frequency components, but with different powers. The idea in the proposed setup is hence to distort the reference signal with a low-order function, which we call deficient. Then we account for the power bias in a later processing stage by using a coupling factor for each sub-band.

Assuming a stationary nonlinear distortion function, it makes sense to accumulate long-term information to describe it. In contrast, adaptive algorithms as in [10] re-adapt on each onset and offset of nonlinear echo components.

The re-use of a single RIR estimate corresponds to the fact that all echo components traverse the same enclosure.

ISBN 978-3-8007-4275-2

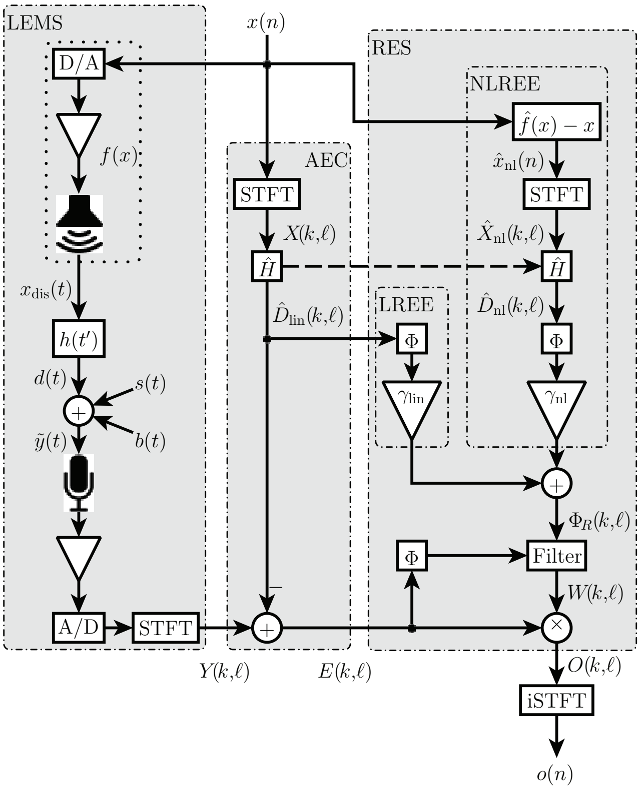

Figure 1: Signal flow. The digitized reference signal x is fed to the echo path, LEMS, a linear AEC, and the RES, which comprises a linear and a nonlinear residual echo power estimator (L-REE, NL-REE), as well as a suppression filter.

1.5 Overview

We first present a novel setup for the control of nonlinear echo in the following section. Section 3 explains the tests performed on the proposed system and how it is compared to a Hammerstein-type NLAEC. The results are described and discussed in section 4. A conclusion summarizes the findings.

2 Proposed Concept

We propose a concept comprising a regular linear AEC and a combined suppression of linear and nonlinear echo as shown in figure 1.

The output $\hat{D}_{\mathrm{lin}}$ of the linear AEC filter is subtracted from the microphone signal Y, which results in the error signal E.

If the distorted reference signal x_{dis} is formally regarded as a sum of linear and nonlinear reference:

$$x_{\mathrm{dis}} = x + x_{\mathrm{nl}} \overset{(2)}{=} x + \left(f(x) - x\right), \tag{5}$$

then the echo signal d can be expressed as a sum of linear and nonlinear echo:

$$\begin{aligned} d &\overset{(5)}{=} h * (x + x_{\mathrm{nl}}) && (6)\\ &= \underbrace{h * x}_{} + \underbrace{h * x_{\mathrm{nl}}}_{} && (7)\\ &=: d_{\mathrm{lin}} + \quad d_{\mathrm{nl}}. && (8) \end{aligned}$$

Note that this differs from Schüßler's method mentioned in [11, (4)]. It is motivated by the assumption that the distortion function exhibits a linear interval around zero and that the linear AEC is adapted only in the absence of higher excitation, i.e., nonlinear distortions.

We define the undisturbed residual echo R as the error signal E minus local speech and noise. It can be attributed in parts to linear and nonlinear echo:

$$\begin{aligned} R &:= E - S - B = D - \hat{D}_{\mathrm{lin}} && (9)\\ &\overset{(8)}{=} \underbrace{D_{\mathrm{nl}}}_{} + \underbrace{D_{\mathrm{lin}} - \hat{D}_{\mathrm{lin}}}_{} && (10)\\ &=: R_{\mathrm{nl}} + \quad R_{\mathrm{lin}}. && (11) \end{aligned}$$

While the linear residual component is expected to be small when the linear AEC has converged, the nonlinear echo fully remains. The residual linear echo depends on the convergence state of the AEC's estimated filter, $\hat{H}$, while the residual nonlinear echo does not. Therefore, residual linear and nonlinear echo power spectral density (PSD) estimates $\hat{\Phi}_{R,\mathrm{lin}}$ and $\hat{\Phi}_{R,\mathrm{nl}}$ must be produced independently of each other to form a combined estimate of the residual echo PSD $\hat{\Phi}_R$.

2.1 Nonlinear Echo PSD Estimation

Presuming an estimate of the distortion function $\hat{f}$ (a method will be published in [12]), this function is applied to the reference signal x. Since the linear echo is treated by the linear AEC, the reference signal is subtracted from the result to get the nonlinear estimate only:

$$\hat{x}_{\mathrm{nl}} = \hat{f}(x) - x. \tag{12}$$

Subsequently, the RIR estimate $\hat{H}$ provided by the linear AEC is convolved with the STFT of the nonlinear reference estimate to get the reverberated nonlinear echo estimate:

$$\hat{D}_{\mathrm{nl}}(k,\ell) = \sum_{\ell'=0}^{L'-1} \hat{H}\left(k,\ell,\ell'\right) \cdot \hat{X}_{\mathrm{nl}}\left(k,\ell-\ell'\right). \tag{13}$$

2.2 Coupling Factors

We obtain the residual echo PSD estimate $\hat{\Phi}_R$ by calculating the linear and nonlinear echo PSD estimates $\Phi := |\cdot|^2$ and combining them in a weighted sum:

$$\hat{\Phi}_R := \gamma_{\mathrm{nl}} \cdot \Phi_{\hat{D},\mathrm{nl}} + \gamma_{\mathrm{lin}} \cdot \Phi_{\hat{D},\mathrm{lin}}. \tag{14}$$

For the weighting, we use frequency-dependent coupling factors, i.e., an expected value or a long-term average between residual echo and estimated echo powers, respectively [13]:

$$\gamma_{\mathrm{lin}} := \frac{\mathscr{E}\left(|R_{\mathrm{lin}}|^2\right)}{\mathscr{E}\left(|\hat{D}_{\mathrm{lin}}|^2\right)}, \qquad \gamma_{\mathrm{nl}} := \frac{\mathscr{E}\left(|R_{\mathrm{nl}}|^2\right)}{\mathscr{E}\left(|\hat{D}_{\mathrm{nl}}|^2\right)}. \tag{15}$$

They can be estimated when no respective disturbance is present. For the linear coupling factor, this is:

$$\hat{\gamma}_{\mathrm{lin}} = \left.\frac{\Phi_E}{\Phi_{\hat{D},\mathrm{lin}}}\right|_{\substack{S(k,\ell)\approx 0,\\ D_{\mathrm{nl}}(k,\ell)\approx 0,}} \tag{16}$$

at times of local silence and when the reference stays undistorted $x(t) < a$ for some time (cf. figure 2). The nonlinear coupling factor must also be estimated. This can be done in local silence and when the linear AEC has sufficiently converged, i.e.:

$$\hat{\gamma}_{\mathrm{nl}} = \left.\frac{\Phi_E}{\Phi_{\hat{D},\mathrm{nl}}}\right|_{\substack{S(k,\ell)\approx 0,\\ \hat{D}_{\mathrm{lin}}(k,\ell)\approx D_{\mathrm{lin}}(k,\ell).}} \tag{17}$$

These estimators can, e.g., be implemented with a first-order infinite impulse response (IIR) smoothing filter. A control algorithm should freeze the smoothing appropriately.

Finally, the total residual echo power estimate is applied in a suitable filter rule, such as the well-known Wiener filter:

$$O := W \cdot E = \left(1 - \frac{\hat{\Phi}_R}{\Phi_E}\right) \cdot E \tag{18}$$

ISBN 978-3-8007-4275-2

3 Evaluation

We compare the potential performance of the proposed system against a Hammerstein-type NLAEC in a fair setup. Since signal generation and evaluation takes place off-line, the following measures can use the signals' total energies instead of their powers.

3.1 Condition Measures

Reference-to-nonlinear power ratio. We characterize the amount of distortion introduced by the distortion function f on the reference signal x using the reference-to-nonlinear power ratio (RNLR):

$$\mathrm{RNLR} := 10\log_{10}\frac{\sum_{n=0}^{N-1} x(n)^2}{\sum_{n=0}^{N-1}\left(x(n)-f(x(n))\right)^2}. \tag{19}$$

Higher values indicate less distortion.

Signal-to-echo power ratio. Modeling different local conditions including the local speaker's speech volume and distance from the microphone is achieved by varying the signal-to-echo power ratio (SER):

$$\mathrm{SER} := 10\log_{10}\frac{\sum_{n=0}^{N-1} s(n)^2}{\sum_{n=0}^{N-1} d(n)^2}. \tag{20}$$

Higher values indicate less echo interference.

3.2 Performance Measures

Echo return loss enhancement. The well-known echo return loss enhancement (ERLE) can in our case be directly calculated on the known echo signal d:

$$\mathrm{ERLE} := 10\log_{10}\frac{\sum_{n=0}^{N-1} d(n)^2}{\sum_{n=0}^{N-1} o(n)^2} \tag{21}$$

for segments without local speech activity. Higher values indicate better echo suppression.

Speech–to–speech-distortion power ratio. The multiplicative nature of any noise suppression algorithm raises the question how much the speech signal is degraded. We use the speech-to-speech-distortion power ratio (SSDR) [14]:

$$\mathrm{SSDR} := 10\log_{10}\frac{\sum_{n=0}^{N-1} s(n)^2}{\sum_{n=0}^{N-1}\left(s(n)-\tilde{s}(n)\right)^2}. \tag{22}$$

Higher values indicate better speech quality.

Nonlinear echo suppression power ratio. On the other hand, the same filter coefficients should provide good echo suppression, which we measure with the nonlinear-echo–to–suppressed-nonlinear-echo power ratio (NLSR):

$$\mathrm{NLSR} := 10\log_{10}\frac{\sum_{n=0}^{N-1} d_{\mathrm{nl}}(n)^2}{\sum_{n=0}^{N-1} \tilde{d}_{\mathrm{nl}}(n)^2}. \tag{23}$$

Higher values indicate better echo suppression.

3.3 Assumptions

Without loss of generality, we only consider odd-symmetric distortion functions with slope 1 at the origin: Any scalar post-gain can equivalently be attributed to the cascaded RIR.

Segmentation of local and far-end speech activity is determined off-line and on the pure signals. This idealization is motivated by the assumption that in a fully developed system, well-tried segmentation algorithms are available.

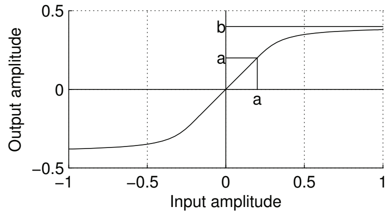

Figure 2: Echo path distortion function with $a = 0.2$, $b = 0.4$.

Some measures require access to the filter coefficients $W(k,\ell)$ internal to the algorithm. The saved filter coefficients are applied on a known signal in the STFT domain. This white-box evaluation method [15] is used in the SSDR, where $\tilde{s}$ is obtained from processing s instead of y, and the NLSR, where $\tilde{d}_{\mathrm{nl}}$ is obtained from processing d_{nl} instead of y. Black box performance measures, such as ERLE, are based on external signals [15].

3.4 Setup

Two evaluations are presented, differing in which conditions are fixed or variable.

All signals are sampled with $f_{\mathrm{S}} = 16\,\mathrm{kHz}$. For the STFT, a Hann analysis window of length $N_{\mathrm{STFT}} = 512$ samples is used with a frame shift of $F = 128$ samples and a corresponding synthesis window.

We call the conditions constituting the task that the compared systems are facing "task variables", whereas the systems' configurable parameters are named "system variables".

3.4.1 Task variables

In all simulations, we use a fixed triplet of far-end and local speech and noise signal. The nonlinear echo path is simulated by applying the distortion function described below on the reference signal with intermediate resampling to avoid aliasing [16] and convolving the result with a fixed sample based RIR filter obtained from measurements in a car cabin. The distortion function is taken from [17]:

$$f(x) := \begin{cases} a + c\cdot\arctan\frac{1}{c}\left(|x|-a\right), & x > +a \\ -a - c\cdot\arctan\frac{1}{c}\left(|x|-a\right), & x < -a \\ x, & \text{otherwise,} \end{cases} \tag{24}$$

where

$$c = \tfrac{2}{\pi}\left(b-a\right). \tag{25}$$

This function exhibits a linear interval $[-a,a]$, and a maximum absolute output value of b as depicted in figure 2.

The systems are provided with different SERs in steps of $10\,\mathrm{dB}$ from $-30\,\mathrm{dB}$ to $30\,\mathrm{dB}$, by scaling the local speech signal accordingly. A fixed distortion function f is used with a reference signal scaled to generate an RNLR of $12\,\mathrm{dB}$. A test speech sentence is used as reference signal. The echo-to-noise power ratio (ENR) is set to $\mathrm{ENR} = 40\,\mathrm{dB}$ by scaling the local noise signal accordingly to emulate low noise conditions.

3.4.2 System variables

The linear AEC's filter coefficient matrix $\hat{H}(k,\ell,\ell')$ with a number $L' = 12$ of filter taps is adapted off-line on an undistorted white noise signal. Adaptation is stopped once $\mathrm{ERLE}_{\mathrm{lin}} = 28\,\mathrm{dB}$ is achieved to simulate an adequate but imperfect convergence state as is typically attained in real-world automobile scenarios. Adaptation in the simulation is prohibited, and the same $\hat{H}$ is used for all conditions. The estimated distortion function model $\hat{f}_{\underline{\theta}}$ and its parameter vector $\underline{\theta}$ are varied in the evaluations. In place of using a segmentation, idealized coupling factor estimates are provided based on the known echo signal components.

Table 1: Distortion function models

name	$\hat{f}(x)$
ideal	$f(x)$
order-9 odd polynomial	$x+\sum_{b=1}^{4}\theta_{2b+1}x^{2b+1}$
clipping	$\min(\theta,\max(-\theta,x))$
Fu and Zhu [18]	$\frac{2\theta_2}{1+e^{-\theta_1 x}}-\theta_2$

The first evaluation compares the proposed system and a Hammerstein-type NLAEC using a clipping distortion model with different clipping thresholds θ at an SER of 0 dB.

The second evaluation uses a variety of distortion function models, compare table 1. Their parameter vectors $\underline{\theta}$ are each fitted to minimize the cost function:

$$C=\sum_{n=0}^{N-1}\left(f(x(nT))-\hat{f}_{\underline{\theta}}(x(nT))\right)^2. \qquad (26)$$

4 Discussion of the Results

For the proposed system, we investigate the NLSR and SSDR tradeoff for a hard clipping model with respect to different clipping thresholds. Further, we compare the proposed system's ERLE to that of an NLAEC, for which suppression-based measures are unavailable. Finally, said measures are evaluated as the SER is varied. The resulting ERLE values are averaged over all SER values, since they have shown to have no effect of relevant size.

4.1 Results

Varying the clipping threshold θ in the distortion function estimate reveals the sensitivity to parameter estimation errors.

The resulting tradeoff between NLSR and SSDR is shown in figure 3 A. A lower clipping threshold yields better nonlinear echo suppression at the cost of additional speech distortion.

Figure 3 B shows that the NLAEC trends towards a linear AEC for vanishing clipping. Also, the proposed system trends toward a linear AEC with a linear RES.

The NLAEC's ERLE reaches its maximum for the fitted parameter $\theta=\theta_{\text{fit}}$. However, the proposed system's maximum ERLE occurs at a lower threshold, indicating that even better performance is achieved when all echo path distortions are also represented by the estimated distortion function.

Under fitted parameterization, figure 4 A shows that both SSDR and NLSR are similar across different function models. The former deteriorates with worse SER (i.e., lower relative local speech power). As seen in figure 4 B, the proposed system performs better than the NLAEC with deficient distortion function models, and slightly worse for the correct (ideal) and the polynomial function model, where the NLAEC reaches the preset $\text{ERLE}_{\text{lin}}=28\,\text{dB}$. The linear AEC could perform better if allowed to converge to the Wiener solution, which minimizes (26) with $\hat{f}_{\underline{\theta}}(x)=\theta x$ but introduces errors in non-distorted segments [19].

4.2 Discussion

The proposed system is more robust against parameter errors than a pure NLAEC. It also shows advantages when the distortion function model is deficient with respect to the true distortion function. This comes at the cost of speech distortion.

Provided an operative segmentation for the signal-dependent occurrence of nonlinear distortion and a distortion function with a linear interval around zero, a common AEC can be used to identify the RIR, while the result can be used in the nonlinear residual echo estimation (NLREE) to reverberate the estimate for the distorted reference signal without the need for further, computationally expensive, adaptation. Still, the identification of the distortion function remains a challenge, but should become an easier task, since low-order parametric models become more feasible.

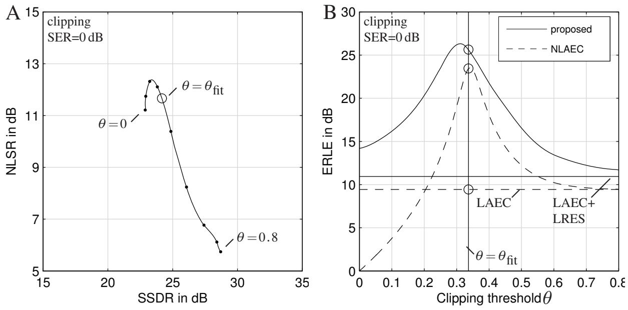

Figure 3: Performance for the clipping function model with different clipping thresholds θ. SER equals 0 dB; RNLR is 12 dB. A – White box evaluations of the proposed system. Clipping thresholds are marked with dots in steps of 0.1, and a circle indicates the fitted θ_{fit}. B – Black box (ERLE) evaluations comparing the proposed system to an NLAEC. The vertical line marks the clipping threshold θ_{fit} fitted optimally for the NLAEC. The dashed horizontal baseline represents the linear AEC's (LAEC) ERLE, while the solid one equals the ERLE of a combined linear AEC and linear RES (LAEC+LRES). Circles correspond to the clipping ERLE values in figure 4 B.

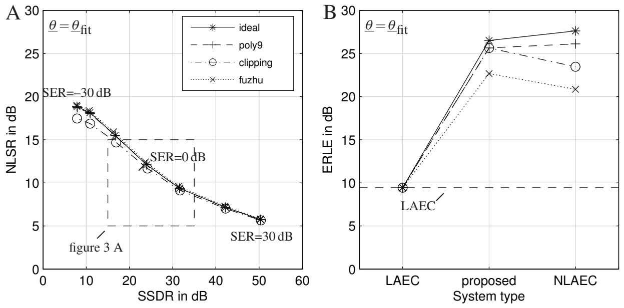

Figure 4: Performance for different distortion function models. RNLR is 12 dB. A – White box evaluations of the proposed system for different SERs. SSDR and NLSR are plotted against each other for different conditions to visualize the compromise between nonlinear echo suppression and speech distortion. Markers represent evaluations at SER values uniformly spaced 10 dB apart. Refer to table 1 for the different function models. B – Black box (ERLE) evaluations for different echo control algorithms. The markers are connected with lines to emphasize the change in ERLE between the systems. The linear AEC (LAEC) is presented as an ERLE baseline.

5 Conclusion

We introduced an efficient and effective concept for the control of nonlinear echo components. All non-linear signal components are represented by a single signal rather than a sum of higher-order signal components that would be derived from a set of basis functions. The computational complexity is reduced by re-using the existing RIR estimate from the linear AEC. However, the proposed concept relies on the presence and identifiability of non-distorted signal segments to adapt the linear RIR coefficients.

By using coupling factors to compensate for long-term deviations, the system becomes robust against a wide range of deficient distortion models and parameter inaccuracies. Further savings in computation power can be expected for distortion function adaptation, since low-order models and rough estimators thus become a viable option due to the increased robustness. Moreover, no significant amount of memory is required over a linear AEC.

Based on this contribution's results, future work should devise an approximation algorithm for a clipping function's threshold parameter. Also, concurrent or alternating adaptation of both the linear AEC and the distortion function parameter must be investigated, as they interfere with each other.

ISBN 978-3-8007-4275-2

References

[1] C. Hofmann, C. Hümmer, and W. Kellermann, "Significance-aware Hammerstein Group Models for Nonlinear Acoustic Echo Cancellation," in *Proc. of ICASSP*, 2014, pp. 5975–5979.

[2] A. Schwarz, C. Hofmann, and W. Kellermann, "Combined Nonlinear Echo Cancellation and Residual Echo Suppression," in *Proc. of Speech Communication; 11. ITG Symposium*, 2014.

[3] E. L. O. Batista and Rui Seara, "Improving the Convergence of Adaptive Hammerstein Filters," in *Proc. of EUSIPCO*, 2013.

[4] F. Küch, "Adaptive Polynomial Filters and their Application to Nonlinear Acoustic Echo Cancellation," PhD Thesis, Friedrich-Alexander-Universität Erlangen-Nürnberg, 2005.

[5] S. Malik and G. Enzner, "Fourier expansion of Hammerstein models for nonlinear acoustic system identification," in *Proc. of ICASSP*. IEEE, 2011, pp. 85–88.

[6] L. A. Azpicueta-Ruiz, M. Zeller, J. Arenas-García, and W. Kellermann, "Novel Schemes for Nonlinear Acoustic Echo Cancellation Based on Filter Combinations," in *Proc. of ICASSP*, no. 1, 2009, pp. 193–196.

[7] F. Küch and W. Kellermann, "Nonlinear Residual Echo Suppression Using a Power Filter Model of the Acoustic Echo Path," in *Proc. of ICASSP*, 2007, pp. 73–76.

[8] D. A. Bendersky, J. W. Stokes, and H. S. Malvar, "Nonlinear Residual Acoustic Echo Suppression for High Levels of Harmonic Distortion," in *Proc. of ICASSP*, 2008, pp. 261–264.

[9] A. Schwarz, C. Hofmann, and W. Kellermann, "Spectral Feature-based Nonlinear Residual Echo Suppression," in *Proc. of WASPAA*, 2013.

[10] K. Shi, X. Ma, and G. T. Zhou, "A residual echo suppression technique for systems with nonlinear acoustic echo paths," in *Proc. of ICASSP*, 2008, pp. 257–260.

[11] G. Enzner, "From Acoustic Nonlinearity to Adaptive Nonlinear System Identification," in *Proc. of Speech Communication; 10. ITG Symposium*, 2012.

[12] I. Schalk-Schupp, F. Faubel, M. Buck, and A. Wendemuth, "Approximation of a Nonlinear Distortion Function for Combined Linear and Nonlinear Residual Echo Suppression," in *Proc. of IWAENC*, 2016, to be published.

[13] E. Hänsler and G. Schmidt, *Acoustic Echo and Noise Control: A Practical Approach*, ser. Adaptive and Learning Systems for Signal Processing, Communications and Control Series. Wiley, 2004.

[14] T. Fingscheidt and Suhadi, "Data-Driven Speech Enhancement," in *Proc. of Speech Communication; ITG Symposium*, 2006.

[15] Suhadi, "Speech Enhancement Using Data-Driven Concepts," PhD thesis, Technische Universität Carolo-Wilhelmina zu Braunschweig, 2012.

[16] I. Schalk-Schupp, F. Faubel, and M. Buck, "Effects of Resampling in Acoustic Echo Cancellation With Static Nonlinear Loudspeaker Distortion," in *Proc. of Speech Communication; 11. ITG Symposium*. Erlangen: VDE, 2014.

[17] J. Seppänen, S. Kananoja, J. Yli-Hietanen, K. Koppinen, and J. Sjöberg, "Maximization of the Subjective Loudness of Speech with Constrained Amplitude," in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, 1999, pp. 139–142.

[18] J. Fu and W.-P. Zhu, "A Nonlinear Acoustic Echo Canceller Using Sigmoid Transform in Conjunction With RLS Algorithm," *IEEE Transactions on Circuits and Systems*, vol. 55, no. 10, pp. 1056–1060, 2008.

[19] H. W. Schüßler and Y. Dong, "A new method for measuring the performance of weakly nonlinear systems," in *Proc. of ICASSP*, 1989, pp. 2089–2092.

ISBN 978-3-8007-4275-2