APPROXIMATION OF A NONLINEAR DISTORTION FUNCTION FOR COMBINED LINEAR AND NONLINEAR RESIDUAL ECHO SUPPRESSION

Ingo Schalk-Schupp, Friedrich Faubel, Markus Buck

Nuance Communications Deutschland GmbH
Acoustic Speech Enhancement Research
89077 Ulm, Germany

Andreas Wendemuth

Otto von Guericke University
Chair of Cognitive Systems
39106 Magdeburg, Germany

ABSTRACT

This work shows a novel method to suppress linear and nonlinear residual echo components after application of a linear echo canceler.

The main idea is to separately treat linear and nonlinear residual echo components, as linear echo is reduced by AEC, while nonlinear echo passes the AEC unchanged. In particular, it is shown that a very simple model, such as a hard clipping function, is sufficient to approximate the nonlinear residual echo power; and that the clipping threshold can be estimated by comparing the broad-band predicted nonlinear residual echo power (produced with the current clipping threshold estimate) to the broad-band observed nonlinear residual echo power (obtained through linear AEC and subtraction of the linear residual echo power, as determined with linear coupling factors).

Experimental evaluations show ERLE improvements by up to $14.9\,\mathrm{dB}$ compared to linear echo cancellation and suppression at a negligible decrease in speech quality during double talk.

Index Terms— Nonlinear Echo, Acoustic Echo Control, Echo Suppression

1. INTRODUCTION

In hands-free speech communication, the microphone not only records the desired near-end signal, such as speech, but also picks up portions of the loudspeaker signal that are reflected or scattered back to the device. These delayed and attenuated versions of the loudspeaker signal are called echo and appear as an additive component d in the microphone signal y, alongside the local signal s and noise b:

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

In most use cases, only the local speech signal is desired while echo and noise are to be removed from the microphone. Practical systems typically treat echo with a combination of two different methods: An acoustic echo canceler (AEC) estimates the echo component in the microphone signal and then subtracts this estimate from the microphone. A residual echo suppressor (RES) estimates the power spectral density power spectral density (PSD) of the residual echo after cancellation and then suppresses it with a Wiener filter [1, 2]. While acoustic echo cancellation has the theoretical potential to perfectly remove the echo, it is, in practice, limited by the accuracy of the echo estimate. Hence, residual echo suppression is used to estimate the remaining residual echo power and to selectively suppress it by further multiplying the signal's time-frequency components with appropriate filter coefficients $W(k, \ell) \in [0, 1]$. The filter operation equally attenuates all of the additive signal components and necessarily leads to a compromise between suppression of the undesired parts and distortion of the desired parts.

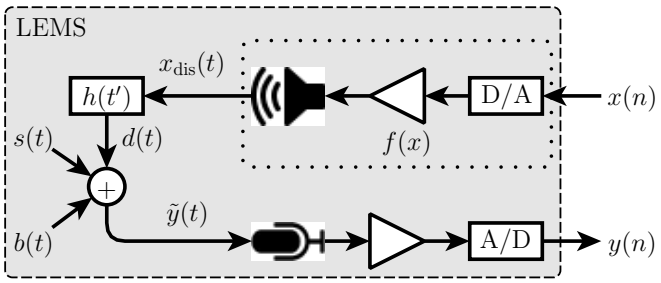

Fig. 1. Hammerstein echo path model. The distorting loudspeaker unit is modeled using the function $f(x)$, where $x(n)$ is called the reference signal. Its distorted output $x_{\mathrm{dis}}(t)$ is convolved with the room impulse response $h(t')$. The microphone unit receives a sum of the distorted echo $d(t)$, local speech $s(t)$, and local noise $b(t)$. The digitized result $y(n)$ is called the microphone signal.

The control of linear echo components with the above methods is a well-studied topic [1]. However, many current publications investigate echo treatment in the presence of nonlinear distortions [3, 4, 5, 6]. Such nonlinearities can, in general, appear at any point of the echo path (i.e., at the microphone, loudspeaker, amplifier, A/D or D/A converter). Volterra filters [7] provide a means of canceling a very broad class of nonlinearities, but the large number of required coefficients make it computationally expensive, even if careful considerations are taken to reduce the complexity [8]. Hence, the echo path is often approximated by a HAMMERSTEIN model [9, p. 283], which has successfully been used in most of the recent work [10, 4, 11]. The HAMMERSTEIN model is a special case of a Volterra-type system and consists of a memoryless (i.e., static) nonlinear component in cascade with a dynamic linear system, as shown in the loudspeaker–enclosure–microphone system (LEMS) in figure 1.

Notable approaches include [12], where a power filter model is applied and the linear branch is separated from all nonlinear polynomial powers. In [13], a linear AEC is combined with a neural-network trained nonlinear RES. The algorithm in [6] also combines a linear AEC with a nonlinear RES, but the latter estimates distorted echo based on a running correlation estimation between monomial power branches.

Our novel approach is explained in more detail in the following section. The experimental results show that the proposed clipping threshold estimate gives significant improvements when used for nonlinear residual echo suppression, with improvements of up to $23.0\,\mathrm{dB}$ time-averaged echo return loss enhancement (ERLE) compared to a baseline AEC system with linear residual echo suppression.

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

The remainder of this paper uses the following notation. The short-time Fourier transform of an arbitrary time-discrete signal x is denoted by $\mathrm{STFT}(x)(k,\ell)$, with sub-band index k and frame index ℓ. The inverse short-time Fourier transform of an arbitrary sub-band signal X is denoted by $\mathrm{iSTFT}(X)(n)$. The instantaneous PSD of an arbitrary sub-band signal X is given by $\Phi_X(k,\ell) = |X(k,\ell)|^2$, while a PSD estimated differently is denoted by $\hat{\Phi}_X(k,\ell)$.

2. ALGORITHM

This contribution introduces an algorithm that approximates the echo path's distortion function f using a clipping function $\hat{f}$ with one parameter, θ:

$$\hat{f}_\theta(x) := \begin{cases} \theta, & x > \theta, \\ -\theta, & x < -\theta, \\ x & \text{otherwise.} \end{cases} \tag{2}$$

This order-1 model allows for an efficient approximation rule while still allowing for high-quality echo suppression.

Assuming that the linear AEC has enough time to converge while no nonlinear distortion is present, that is, $x(n) \in [-\theta, \theta]$ for all $n < n^*$ the resulting filter coefficients $\hat{H}(k,\ell')$ will explain any scaling $c \cdot f$ in the echo path's true distortion function f, because the scaling can equivalently be attributed to the cascaded room impulse response (RIR): $h * (c \cdot f) = (c \cdot h) * f$. By freely choosing the scaling $c = 1/f'(0)$ to produce an effective unitary slope at the origin, we can model the distorted reference signal x_{dis} as a sum of the unscaled linear reference signal x and a remaining nonlinear summand x_{nl}:

$$x_{\text{dis}} := x + x_{\text{nl}}. \tag{3}$$

Even though saturation-type distortions reduce a signal's power ($\Phi_{x,\text{dis}} < \Phi_x$), the model from (3) dictates that there will be power in the nonlinear signal x_{nl} whenever any distortions occur. This allows for separate treatment of linear and nonlinear echo components and forms the basis for approximating θ.

The approximation algorithm is embedded in the echo suppression setup described in [14]. The signal flow is illustrated in figure 2 and described in the following.

Linear echo canceler (LAEC). A linear AEC convolves the short-time Fourier transform (STFT)-transformed reference signal x with its filter coefficients $\hat{H}$ to provide a linear echo estimate $\hat{D}_{\text{lin}}$:

$$\hat{D}_{\text{lin}}(k,\ell) = \sum_{\ell'=0}^{L'-1} \hat{H}\big(k,\ell',\ell\big) \cdot X\big(k,\ell-\ell'\big), \tag{4}$$

which is subtracted from the microphone signal transform Y:

$$E(k,\ell) = Y(k,\ell) - \hat{D}_{\text{lin}}(k,\ell) \tag{5}$$

resulting in the error signal E, which is typically used to adapt the coefficients $\hat{H}$. Depending on the canceler's convergence state, a residual echo signal R remains in E even in the case of a linear echo path:

$$E(k,\ell) = S(k,\ell) + B(k,\ell) + \underbrace{D(k,\ell) - \hat{D}_{\text{lin}}(k,\ell)}_{=R(k,\ell)}. \tag{6}$$

Corresponding to (3), the echo and residual echo signals naturally separate accordingly:

$$E(k,\ell) = S(k,\ell) + B(k,\ell) + \underbrace{D_{\text{lin}}(k,\ell) - \hat{D}_{\text{lin}}(k,\ell)}_{=R_{\text{lin}}(k,\ell)} + \underbrace{D_{\text{nl}}(k,\ell)}_{=R_{\text{nl}}(k,\ell)}. \tag{7}$$

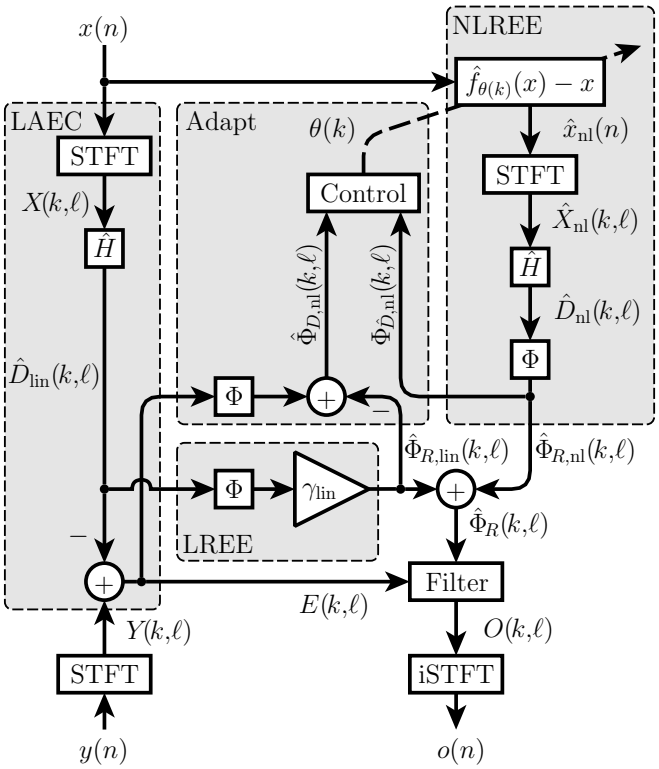

Fig. 2. Signal flow for the approximation algorithm and its application for echo suppression. Based on the reference signal x, a linear echo canceler (LAEC) estimates the linear echo $\hat{D}_{\text{lin}}$ and subtracts it from the microphone signal Y. The resulting error signal E passes a suppression filter providing the output signal O. The filter's coefficients result from the combined estimated residual echo PSD from the linear (LREE)and nonlinear (NLREE) echo estimation. An estimated distortion function $\hat{f}_\theta$ estimated the distorted reference signal. The parameter θ is controlled (Adapt) based on a power comparison.

Linear residual echo power estimation (LREE). Residual echo suppression (RES) requires a PSD estimate for the residual echo components. The resulting linear residual echo PSD estimate is determined as:

$$\hat{\Phi}_{R,\text{lin}}(k,\ell) = \gamma_{\text{lin}}(k,\ell) \cdot \Phi_{\hat{D},\text{lin}}(k,\ell) \tag{8}$$

and can be reduced by determining the linear coupling factor [1]:

$$\gamma_{\text{lin}}(k,\ell) = \frac{\overline{\Phi}_E(k,\ell)}{\overline{\Phi}_{\hat{D},\text{lin}}(k,\ell)}, \tag{9}$$

where $\overline{\Phi}_E$ and $\overline{\Phi}_{\hat{D},\text{lin}}$ are the PSDs of E and $\hat{D}_{\text{lin}}$ smoothed in time direction, respectively. In the latter, smoothing constants are chosen so as to fall faster than it rises. This ensures that sporadically occurring nonlinear distortions are not explained by the linear coupling factor alone. Coupling factor estimation is stopped whenever there is local speech activity.

Nonlinear residual echo power estimation (NLREE). According to (3), the nonlinear reference signal can be estimated as:

$$\hat{x}_{\text{nl}}(n) = \hat{x}_{\text{dis}}(n) - x(n) = \hat{f}_{\theta(k)}(x(n)) - x(n) \tag{10}$$

using the hard clipping from (2) with the clipping threshold θ. Its transform $\hat{X}_{\text{nl}}(k,\ell) = \mathrm{STFT}(\hat{x}_{\text{nl}})(k,\ell)$ is then convolved with the

LAEC's filter coefficients $\hat{H}$:

$$\hat{D}_{\text{nl}}(k,\ell) = \sum_{\ell'=0}^{L'-1} \hat{H}\big(k,\ell',\ell\big) \cdot \hat{X}_{\text{nl}}\big(k,\ell-\ell'\big) \tag{11}$$

to apply the linear AEC's RIR model to the nonlinear signal and thus predict the nonlinear echo component.

In contrast to [14], no coupling factor is determined for the nonlinear echo. The nonlinear echo PSD estimate is thus:

$$\hat{\Phi}_{R,\text{nl}}(k,\ell) = \Phi_{\hat{D},\text{nl}}(k,\ell). \tag{12}$$

Residual echo suppression (RES). Both the linear and nonlinear residual echo PSD estimates are combined:

$$\hat{\Phi}_R(k,\ell) = \hat{\Phi}_{R,\text{lin}}(k,\ell) + \hat{\Phi}_{R,\text{nl}}(k,\ell) \tag{13}$$

neglecting their cross-spectral density. The resulting echo PSD estimate is fed to a Wiener filter to produce the output signal O:

$$O(k,\ell) := W(k,\ell) \cdot E(k,\ell) = \left(1 - \frac{\hat{\Phi}_R(k,\ell)}{\Phi_E(k,\ell)}\right) \cdot E(k,\ell). \tag{14}$$

Distortion function approximation (Adapt). The main task for the proposed algorithm is the approximation of the distortion function f using a deficient model (2).

It is the basic idea that the PSD $\Phi_{\hat{D},\text{nl}}$ of the estimated nonlinear echo signal must be roughly equal to the observed residual echo $\hat{\Phi}_{D,\text{nl}}$ in a long-term average, given that the clipping function model roughly resembles the true distortion function. If the model produces too much power, then its clipping threshold θ is too low. Conversely, if there is more power than can be explained, θ must be too high. This is reflected in the PSD ratio:

$$P(\ell) = \frac{\sum_{i=0}^{\|\mathbb{K}\|-1} \hat{\Phi}_{D,\text{nl}}(k_i,\ell)}{\sum_{i=0}^{\|\mathbb{K}\|-1} \Phi_{\hat{D},\text{nl}}(k_i,\ell)}, \tag{15}$$

which is used to decide on the direction of the adapation only:

$$\delta(\ell) = \begin{cases} -1, & P(\ell) > 1 \\ 1, & \text{otherwise.} \end{cases} \tag{16}$$

The adaptation itself is performed using:

$$\theta(\ell+1) = \alpha_\theta(\ell) \cdot \theta(\ell) \; + \; (1-\alpha_\theta(\ell)) \cdot \delta(\ell) \tag{17}$$

with the adaptation speed α_θ, which changes dependent on the signal by first choosing a set $\mathbb{K}(\ell)$ of frequency bins in each frame ℓ that contain more power than is explained by the linear residual (8):

$$\mathbb{K}(\ell) = \left\{ k \,\middle|\, \frac{\max(\Phi_E(k,\ell), \Phi_{\hat{D},\text{nl}}(k,\ell))}{\max(c \cdot \hat{\Phi}_{R,\text{lin}}(k,\ell), \Phi_{\text{min}})} > 1 \right\}. \tag{18}$$

Also, it should exceed a minimum PSD threshold Φ_{min} to avoid adaptation to spurious remnants.

The fraction r of active bins:

$$r(\ell) = \frac{\|\mathbb{K}(\ell)\|}{K} \in [0,1] \tag{19}$$

is used to freeze adaptation when it is too small $r(\ell) < 0.1$. It is averaged over the last $L_r = 10$ active frames to control the adaptation rate α_θ:

$$\alpha_\theta(\ell) = \alpha_0^{\overline{r}(\ell)}. \tag{20}$$

3. EVALUATION

Novel algorithms should be tested and evaluated under a broad set of conditions. However, a concise presentation demands the identification of those with the most impact on performance. Moreover, performance measures must be chosen according to the type of algorithm.

All signals were sampled at a sample rate of $f_\text{s} = 16000\,\text{Hz}$. The STFTs were performed with a Hann analysis window of length $N_\text{FFT} = 512$ samples. For the frame shift of $R = 128$ samples per frame, an appropriate synthesis window was designed [15] for use in the inverse short-time Fourier transforms (iSTFTs). Simulations were performed using a RIR measured inside a car cabin. The number of AEC filter taps amounts to $n_{\text{tap,AEC}} = 12$ frames and accommodates the RIR length plus on- and offset due to sub-sampling.

3.1. Condition measures and signals

There are many variables that affect the suppression algorithm's performance. In order to provide reproducible results, we set some of them constant, while others are varied.

We assume the linear AEC to have converged to the LEMS sufficiently to provide an ERLE of $\text{ERLE}_\text{lin} = 28\,\text{dB}$ for non-distorted signals. This is achieved by running an normalized least mean squares (NLMS) algorithm on a white-noise reference signal with the corresponding response signal and aborting the adaptation once the target ERLE is reached. The resulting filter coefficients $H(k,\ell')$ are then fixed and provided to each algorithm under consideration.

Also, no actual voice activity detection (VAD) was implemented, and the segmentation is provided based on the clean speech signal s, which would be unknown under real-world conditions.

No local noise is present in any of the evaluation scenarios.

For both the far-end and the local signal, we use clean, mostly continuous speech signals. While far-end activity starts right away, local speech only begins after 12 seconds, when a double-talk segment of 8 seconds begins.

Two variables remain, which define the better part of a condition's challenge. First, the signal-to-echo power ratio (SER) reflects the balance between echo power and local signal power:

$$\text{SER} := 10\log_{10} \frac{N_d \sum_{\nu=0}^{N_s-1} s(n_\nu)^2}{N_s \sum_{\nu=0}^{N_d-1} d(n_\nu)^2}, \tag{21}$$

where N_x is the number of active samples. Higher values indicate less echo interference and an easier condition. Evaluations were performed for a range of SER from $-30\,\text{dB}$ (very quiet local speech) over $0\,\text{dB}$ (echo and local speech equally loud) to $30\,\text{dB}$ (very loud local speech) in steps of $10\,\text{dB}$.

Second, we assess the nonlinear distortion by using the reference-to-nonlinear power ratio (RNLR), which we define as

$$\text{RNLR} := 10\log_{10} \frac{\sum_{n=0}^{N-1} x(n)^2}{\sum_{n=0}^{N-1} \big(f(x(n)) - x(n)\big)^2}, \tag{22}$$

where N is the number of samples in x. Higher values indicate less distortion and an easier condition. We evaluated performances at RNLRs of $6\,\text{dB}$ (intense distortion), $12\,\text{dB}$ (medium distortion), and $18\,\text{dB}$ (slight distortion).

3.2. Performance measures

It is common to evaluate echo control algorithms in terms of the well-known ERLE measure. This is a good choice in black-box scenarios, where one has no access to an algorithm's internal states and

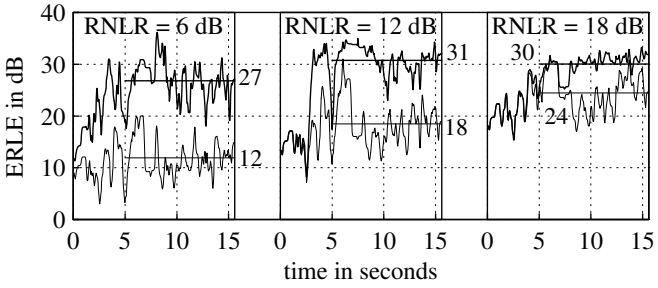

Fig. 3. Black box evaluation with double-talk segment removed. Thick lines represent the proposed algorithm's ERLE, while thin lines stand for purely linear echo suppression. Horizontal lines and values mark the mean log ERLE in decibels over the indicated time.

signals. However, since we evaluate a suppression-type algorithm's performance, we additionally use the following measures.

The speech–to–speech-distortion power ratio (SSDR) inversely reflects how much the local speech signal is distorted, where higher values indicate better speech quality [16]:

$$\text{SSDR} := 10\log_{10}\frac{\sum_{\nu=0}^{N_{\text{DT}}-1} s(n_\nu)^2}{\sum_{\nu=0}^{N_{\text{DT}}-1}\left(s(n_\nu)-\tilde{s}(n_\nu)\right)^2}, \tag{23}$$

where N_{DT} is the number of samples in the double talk segment and $\tilde{s}$ is the speech signal s processed with the filter coefficients $W(k,\ell)$.

Higher values of the disturbance–to–suppressed-disturbance power ratio (DSR) indicate better echo and noise suppression:

$$\text{DSR} := 10\log_{10}\frac{\sum_{\nu=0}^{N_{\text{DT}}-1} i(n_\nu)^2}{\sum_{\nu=0}^{N_{\text{DT}}-1}\tilde{i}(n_\nu)^2}, \tag{24}$$

where the disturbance signal i is the sum of all unwanted signal components in the AEC output e:

$$i(n) = \text{iSTFT}(E-S)(n) = \text{iSTFT}(R+B)(n), \tag{25}$$

and $\tilde{i}$ is i processed with the filter coefficients $W(k,\ell)$.

3.3. Baseline and reference

We compare our algorithm to two other suppression-type algorithms using the same signals, conditions, and performance measures.

For the baseline, we use a combined linear echo cancellation and coupling factor based suppression [17], since it is appropriate for embedded real time applications, where computation resources are limited even today. As an idealized reference, we use the optimal Wiener filter coefficients resulting from perfect PSD estimation of the disturbance signal i.

4. RESULTS

As expected, the proposed algorithm outperforms linear echo suppression for all distortion levels (RNLR) in terms of ERLE, which is independent of the SER and only evaluated for far end single talk segments. See figure 3. After convergence of the proposed algorithm, a log mean ERLE between 27 dB and 31 dB is achieved.

White box measures are only evaluated in the double-talk segment. Figure 4 shows that the proposed algorithm outperforms the linear baseline in terms of echo suppression but at the cost of -2.5 dB of additional speech distortion on average. The DSR improvement

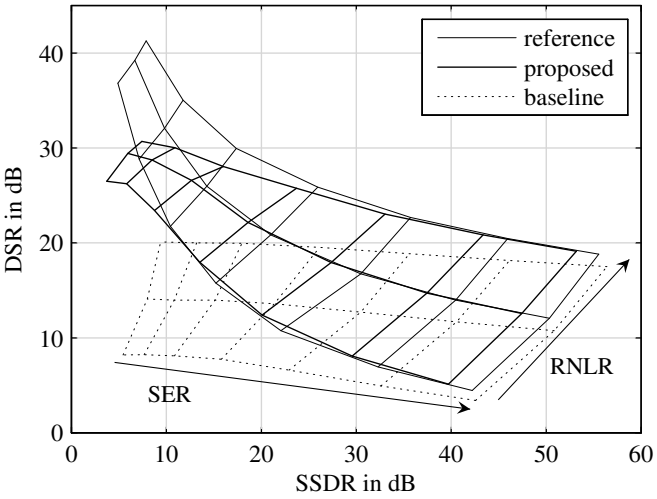

Fig. 4. White box evaluation for double talk segment. Each grid represents one algorithm and is parameterized by two condition measures: SER from -30 dB to 30 dB and RNLR from 6 dB to 18 dB. Both make an easier task the higher they are. Echo suppression (DSR) and speech quality (SSDR) reflect the algorithms' performance for each condition. Higher values indicate better performance for both.

ranges from 1.6 dB at high SER to 18.3 dB at low SER and low RNLR.

Compared to the reference, proposed echo attenuation ranges from 2.2 dB improvement to -10.6 dB below reference at low SER.

Note that the tradeoff in all algorithms can be adjusted to different requirements by applying an overestimation factor to the disturbance PSD estimate. This way, up to 35 dB DSR can be achieved at the cost of additional speech distortion of -2.5 dB compared to the proposed one without overestimation. Alternatively, speech distortion can be reduced to reference level, resulting in echo suppression performance roughly in the middle between baseline and reference. The illustration depicts the algorithms' "natural" behavior without overestimation.

5. CONCLUSION

We introduced a novel algorithm for the suppression of nonlinear echo components resulting from saturation-type Hammerstein distortions.

The clipping threshold of a hard-clipping distortion function model is approximated by comparing the model's output power to the observed residual echo power after a linear echo cancellation and a suppression stage.

Local background noise was not yet considered in this contribution but is the subject of follow-up research.

We exploited the suppression-type approach by assuming a deficient distortion model that would not perform well for a cancellation-type approach. Still, we have shown that this model is able to provide the information necessary for suppressing most nonlinear echo components while preserving speech quality during double talk.

The computational complexity, compared with linear echo cancellation and suppression, is dominated by just one additional STFT and a sub-band convolution. Depending on the number of branches used in a power filter model, the proposed algorithm is thus by far less complex.

6. REFERENCES

[1] Eberhard Hänsler and Gerhard Schmidt, *Acoustic Echo and Noise Control: A Practical Approach*, Adaptive and Learning Systems for Signal Processing, Communications and Control Series. Wiley, 2004.

[2] Gerald Enzner, Rainer Martin, and Peter Vary, "Unbiased Residual Echo Power Estimation for Hands-free Telephony," in *Proc. ICASSP*, 2002, pp. 1893–1896.

[3] Diego A. Bendersky, Jack W. Stokes, and Henrique S. Malvar, "Nonlinear Residual Acoustic Echo Suppression for High Levels of Harmonic Distortion," in *Proc. ICASSP*, Buenos Aires, Argentina, 2008, pp. 261–264.

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

[5] Andreas Schwarz, Christian Hofmann, and Walter Kellermann, "Spectral Feature-based Nonlinear Residual Echo Suppression," in *Proc. of WASPAA*, 2013.

[6] Kun Shi, Xiaoli Ma, and G. Tong Zhou, "A residual echo suppression technique for systems with nonlinear acoustic echo paths," in *Proc. of ICASSP*, 2008, pp. 257–260.

[7] Martin Schetzen, *The Volterra and Wiener Theories of Nonlinear Systems*, Wiley, 1980.

[8] Luis Antonio Azpicueta-Ruiz, Marcus Zeller, Aníbal Ramón Figueiras-Vidal, Jerónimo Arenas-Garcia, and Walter Kellermann, "Adaptive Combination of Volterra Kernels and Its Application to Nonlinear Acoustic Echo Cancellation," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 19, pp. 97–110, 2011.

[9] Robert Haber and László Keviczky, *Nonlinear System Identification. 1. Nonlinear system parameter identification*, Springer Science & Business Media, 1999.

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

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

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

[13] Andreas Schwarz, Christian Hofmann, and Walter Kellermann, "Combined Nonlinear Echo Cancellation and Residual Echo Suppression," in *Proc. of Speech Communication; 11. ITG Symposium*, 2014.

[14] Ingo Schalk-Schupp, Friedrich Faubel, and Markus Buck, "Combined Linear and Nonlinear Residual Echo Suppression Using a Deficient Distortion Model," to be published in *Speech Communication; 12. ITG Symposium*, 2016.

[15] Patrick Hannon, Mohamed Krini, Gerhard Schmidt, and Arthur Wolf, "Reducing the Complexity or the Delay of Adaptive Subband Filtering," in *Proc. ESSV*, 2010, pp. 158–165.

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

[17] Andreas Mader, Henning Puder, and Gerhard Uwe Schmidt, "Step-Size Control for Acoustic Echo Cancellation Filters – An Overview," *Signal Processing*, vol. 80, no. 9, pp. 1697–1719, 2000.