NONLINEAR RESIDUAL ECHO SUPPRESSION USING A POWER FILTER MODEL OF THE ACOUSTIC ECHO PATH

Fabian Kuech

Fraunhofer Institute for Integrated Circuits
Am Wolfsmantel 33,
91058 Erlangen, Germany
fabian.kuech@iis.fraunhofer.de

Walter Kellermann

Multimedia Communications and Signal Proc.
University of Erlangen-Nuremberg
Cauerstr. 7, 91058 Erlangen, Germany
wk@LNT.de

ABSTRACT

Loudspeakers and amplifiers of mobile communication receivers may cause significant nonlinear distortion in the acoustic echo path, resulting in a limitation of the performance of linear echo cancelers. In this contribution, we present a nonlinear acoustic echo suppressor in order to increase the attenuation of the nonlinearly distorted residual echo. The proposed approach is based on a power filter model of the acoustic echo path which is applied to the estimation of the power spectral density of the nonlinear residual echo. These time-variant estimates are used to appropriately adjust the frequency-dependent gain values of the echo suppressor. The performance of the proposed approach is evaluated in realistic experimental set-ups.

Index Terms— Echo suppression, nonlinear filters

1. INTRODUCTION

In the acoustic echo cancellation problem, illustrated in Fig. 1, the acoustic echo canceler (AEC) seeks to minimize the contribution of the echo signal $y(k)$ to the power of the error signal $e(k)$ by subtracting an estimate of the echo signal $\hat{y}(k)$ from the microphone signal $d(k)$. Besides the acoustic echo, the microphone also picks up the signal $b(k)$ which is composed of background noise and the speech signal of the near-end talker. The performance of standard

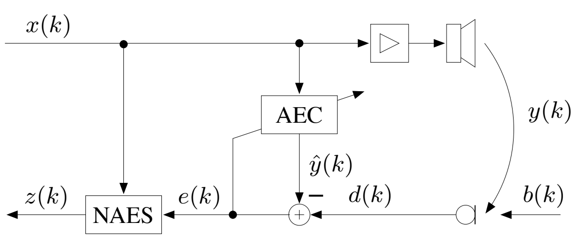

Fig. 1. Set-up of the nonlinear acoustic echo suppression problem.

approaches for the cancellation of acoustic echoes strongly rely on the assumption of a linear echo path. However, in applications such as echo cancellation for mobile phones, non-negligible nonlinear distortion is introduced by loudspeakers and their amplifiers [1, 2, 3]. With these nonlinear distortions, purely linear echo cancelers are not able to provide a sufficient echo attenuation.

A way to improve the achievable echo attenuation in case of nonlinear echo paths is to realize the AEC as a nonlinear adaptive structure [1, 2]. Unfortunately, the convergence of the nonlinear portions of known approaches is often too slow. Consequently, their application is also limited to rather slowly time varying scenarios which renders application often inattractive, especially in mobile phones. As shown in Fig. 1, a way to overcome these drawbacks of nonlinear acoustic echo *cancellation* is to apply nonlinear acoustic echo *suppression* to further reduce the residual echo that remains after a purely linear AEC. Such post-filtering of the residual echo represents a well established technique in the context of controlling linear echoes [4]. However, these methods are also based on the assumption of a linear echo path and, thus, not applicable in case of strong nonlinear distortions. The approach using a nonlinear acoustic echo suppressor (NAES) presented in [3] requires a frequency-domain model of the nonlinear residual echo that has to be determined in advance. Since this model depends on the hardware components actually included in the echo path, it has to be acquired for each hardware set-up separately. The NAES proposed in this paper uses power filters as a very general model for the nonlinear acoustic echo path. Such a model has already been proposed in [2] and it is briefly reviewed in Sec. 2. As described in Sec. 3, the power spectral density of the nonlinear residual echo is estimated using this nonlinear model of the echo path. Analogously to the linear case [4], this estimate is then used for determining the frequency-variable gains values of the NAES. The simulation results presented in Sec. 4 confirm the suitability of the proposed method to cope with strong nonlinearly distorted acoustic echoes.

2. MODEL OF THE NONLINEAR ECHO PATH

As shown in [2, 5], power filters represent an efficient parallelized model of the nonlinear acoustic echo path to be considered in mobile phones. In the following we briefly recall this concept for modelling the nonlinear echo paths.

2.1. Power filters

As depicted on the right hand side of Fig. 2, the input/output relation of a Pth-order power filter is given by

$$y(k) = \sum_{p=1}^{P} \sum_{n=0}^{N-1} h_{p,n}\, x^p(k-n), \tag{1}$$

where $h_{p,n}$ denotes the filter coefficients of the pth channel having input $x^p(k)$. Aiming at a frequency-domain implementation of power filters in Section 3, we further give the short-time Fourier transform (STFT) representation of (1):

$$Y(\kappa, m) = \sum_{p=1}^{P} H_{p,m}\, X_p(\kappa, m). \tag{2}$$

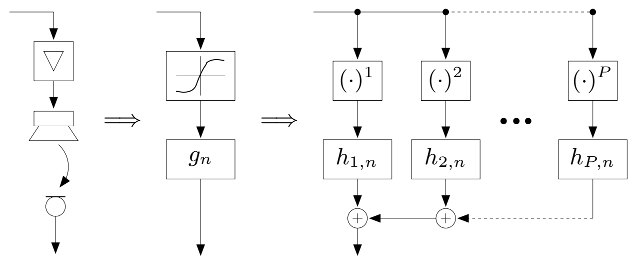

Fig. 2. Illustration of the nonlinear acoustic echo path (left), its cascaded model (center), and a corresponding parallelized implementation with power filters (right).

Here, $X_p(\kappa, m)$ denotes the STFT of $x^p(k)$ of length M, and $H_{p,m}$ is the coefficient of the pth power filter channel corresponding to the mth frequency bin. Accounting for the block processing of frequency-domain methods, κ represents the block time index with $k = \kappa R$, where R is the frame length. Commonly, the length of the STFT M is an integer multiple of the frame length R.

2.2. Power filters as model for the nonlinear echo path

As can be seen on the left hand side of Fig. 2, the echo path to be modeled by the echo canceler is a cascade of amplifier, loudspeaker, microphone, and the sound propagation path between the loudspeaker and the microphone. In this paper we consider two sources for nonlinear distortion: the loudspeaker and its amplifier. In mobile phones, high signal levels must be provided despite low battery voltages and very small loudspeakers. Typically, users prefer overloaded audio components over reduced output levels. The nonlinear behaviour of the amplifiers and the miniaturized loudspeakers of mobile phones can then be described as a memoryless saturation characteristic, representing the soft clipping of large amplitude values [1, 2]. As illustrated in the center of Fig. 2, the nonlinear echo path to be expected for the considered application can therefore be modeled by the cascade of a saturation characteristic followed by a linear filter, representing the remaining linear propagation path of the echo signal. It has been shown in [2] that power filters represent a parallelized realization of such nonlinear cascaded systems if the memoryless nonlinearity can sufficiently be approximated by a truncated Taylor series expansion. Fig. 2 illustrates the mapping of the nonlinear echo path to the corresponding representation using power filters of order P. It is straightforward to show that the coefficients of the power filter have to be chosen according to $h_{p,n} = a_p g_n$, where a_p denote the coefficients of the corresponding Taylor series expansion and g_n are the coefficients of the subsequent linear filter as depicted in Fig. 2.

3. NONLINEAR RESIDUAL ECHO SUPPRESSION

Standard methods for designing a so-called post-filter for the suppression of *linear* residual echoes rely on the assumption of a linear echo path and, moreover, exploit knowledge of certain states of the linear echo canceler [4]. In this contribution we discard these common constraints and consider *nonlinear* residual echo suppression based on the practical assumptions that the nonlinear echo path can sufficiently be modeled by a Pth-order power filter. In the following we do not require any specific information with respect to the preceding linear AEC, but base our discussion on the the residual echo $e(k)$ and the nonlinear model of the acoustic echo path only. Therefore, any linear adaptive algorithm can be chosen for updating the linear AEC [6]. From a computational complexity point of view it is, however, beneficial to apply frequency-domain adaptive filtering [7], since the subsequent echo suppression is also performed in the STFT domain.

Analogously to the post-filter commonly used in linear echo control, the AES is realized as a frequency-dependent, real-valued gain filter $G(\kappa, m)$ [4]. Accordingly, the output $z(k)$ of the AES with input $e(k)$ reads in the STFT domain

$$Z(\kappa, m) = G(\kappa, m)\, E(\kappa, m). \tag{3}$$

A suitable way to implement the gain function $G(\kappa, m)$ is presented in the sequel.

For the following considerations, it is beneficial, to express the power filter model of the nonlinear echo path by its so-called equivalent orthogonalized structure (EOS). As discussed in [5], the EOS implies a mutual orthogonalization of the channel inputs. Its frequency-domain representation is defined according to

$$X_{o,1}(\kappa, m) \quad = \quad X_1(\kappa, m), \tag{4}$$

$$X_{o,p}(\kappa, m) \quad = \quad X_p(\kappa, m) + \sum_{i=1}^{p-1} q_{p,i}(\kappa)\, X_i(\kappa, m), \tag{5}$$

for $1 < p \leq P$. The orthogonalization coefficients $q_{p,i}(\kappa)$ are chosen such that the input signals of different channels are mutually orthogonal, i.e.,

$$\mathrm{E}\{X_{o,i}(\kappa, m) X_{o,j}(\kappa, m)\} = 0, \quad \text{for } i \neq j. \tag{6}$$

Analogously to (2), the output of the power filter using its EOS reads

$$Y(\kappa, m) = \sum_{p=1}^{P} H_{o,p,m}\, X_{o,p}(\kappa, m), \tag{7}$$

where $H_{o,p,m}$ are the coefficients of the frequency-domain EOS. Furthermore, we explicitly give the output of the pth channel of the EOS according to

$$Y_{o,p}(\kappa, m) = H_{o,p,m}\, X_{o,p}(\kappa, m). \tag{8}$$

If we assume that the linear echo canceler attains to its Wiener solution, the error signal $e(k)$ contains only echo components that are orthogonal to the linear input $x(k)$ [6]. Regarding the results presented in [5], the error signal can then be expressed by

$$E(\kappa, m) = E_{\mathrm{nl}}(\kappa, m) + B(\kappa, m), \tag{9}$$

where $B(\kappa, m)$ is the STFT-domain representation of the near-end signal $b(k)$. The nonlinear residual echo signal $E_{\mathrm{nl}}(\kappa, m)$, is composed of the nonlinear channels ($p \geq 2$) of the EOS corresponding to the echo path and can be written as

$$E_{\mathrm{nl}}(\kappa, m) = \sum_{p=2}^{P} Y_{o,p}(\kappa, m). \tag{10}$$

The optimum gain values $G_{\mathrm{opt}}(\kappa, m)$ can then be derived by minimizing the contribution of the nonlinear residual echo $E_{\mathrm{nl}}(\kappa, m)$ to the output signal $Z(\kappa, m)$ in the mean square error (MSE) sense. Regarding the results known from the linear case [4, 6], we obtain

$$G_{\mathrm{opt}}(\kappa, m) = \frac{S_E(\kappa, m) - S_{E_{\mathrm{nl}}}(\kappa, m)}{S_E(\kappa, m)} \tag{11}$$

where $S_E(\kappa, m)$ and $S_{E_{\rm nl}}(\kappa, m)$ denote the power spectral densities (PSDs) of $E(\kappa, m)$ and $E_{\rm nl}(\kappa, m)$, respectively. Due to the orthogonality of the different channels of the EOS, it follows from (10) that the PSD of the nonlinear residual echo can be written as

$$S_{E_{\rm nl}}(\kappa, m) = \sum_{p=2}^{P} S_{Y_{{\rm o},p}}(\kappa, m), \tag{12}$$

where $S_{Y_{{\rm o},p}}(\kappa, m)$ denotes the PSD of the output of the pth channel of the EOS.

Let us now look at the estimation of the PSDs required in (11). An estimate of $S_E(\kappa, m)$ can, e.g., easily be obtained by recursive smoothing of $|E(\kappa, m)|^2$:

$$\widehat{S}_E(\kappa, m) = \lambda \widehat{S}_E(\kappa - 1, m) + (1 - \lambda)|E(\kappa, m)|^2, \tag{13}$$

where λ is a forgetting factor. Unfortunately, such a direct approach can not be applied for determining $S_{E_{\rm nl}}(\kappa, m)$ since $E_{\rm nl}(\kappa, m)$ is not accessible. As a solution, we extend the approach presented in [8] for linear AES to the case of nonlinear echo paths. Thereby, we apply adaptive power filters to obtain suitable estimates of the PSDs of the channel outputs $S_{Y_{{\rm o},p}}(\kappa, m)$ as shown below. The estimation of the PSD of the nonlinear residual echo can then be performed based on (12), i.e., by exploiting the orthogonality of the different channels.

Analogously to (8), the output of the pth channel of the adaptive EOS of a power filter reads

$$\widehat{Y}_{{\rm o},p}(\kappa, m) = \widehat{H}_{{\rm o},p,m}(\kappa)\, X_{{\rm o},p}(\kappa, m). \tag{14}$$

The adaptation of the coefficients $\widehat{H}_{{\rm o},p,m}(\kappa)$ is performed independently for each nonlinear channel ($p \geq 2$) with respect to the channel-dependent error signal

$$E_{{\rm o},p}(\kappa, m) = E(\kappa, m) - \widehat{Y}_{{\rm o},p}(\kappa, m). \tag{15}$$

In contrast to the application of adaptive power filters to nonlinear echo cancellation [2, 5], we are not interested in an exact identification of the nonlinear components of the echo path $H_{{\rm o},p,m}(\kappa)$ by their adaptive counterparts $\widehat{H}_{{\rm o},p,m}(\kappa)$. We are rather aiming at estimates of $S_{Y_{{\rm o},p}}(\kappa, m)$. In order to reduce the computational complexity we, therefore, discard all windowing and time constraints commonly applied in frequency-domain adaptive filtering [7]. In other words, the adaptation is based on circular convolutions implemented in the STFT domain instead of corresponding linear convolutions [7]. The update equation for the adaptive coefficients is then given by

$$\widehat{H}_{{\rm o},p,m}(\kappa + 1) = \widehat{H}_{{\rm o},p,m}(\kappa) + \mu_p\, \frac{E_{{\rm o},p}(\kappa, m)\, X^*_{{\rm o},p}(\kappa, m)}{\widehat{S}_{X_{{\rm o},p}}(\kappa, m)} \tag{16}$$

for $p \geq 2$. Here, X^* denotes the complex conjugate of X. The time-variant normalization factor $\widehat{S}_{X_{{\rm o},p}}(\kappa, m)$ represents an estimate of the PSD of $X_{{\rm o},p}(\kappa, m)$.

Note that in contrast to the adaptive power filter approach in [5], the adaptation (16) is performed independently for each channel p, implying that the step-size μ_p does not have to be controlled with respect to the adaptation of the other channels. Thus, μ_p can be controlled in each channel separately by using approaches known from linear adaptive filtering, e.g., such as [9].

The desired estimate of the PSD of the nonlinear residual echo is finally obtained by the approximation

$$\widehat{S}_{E_{\rm nl}}(\kappa, m) \approx \sum_{p=2}^{P} \widehat{S}_{\widehat{Y}_{{\rm o},p}}(\kappa, m), \tag{17}$$

where, analogously to (13), we apply recursive smoothing for the computation of $\widehat{S}_{\widehat{Y}_{{\rm o},p}}(\kappa, m)$:

$$\widehat{S}_{\widehat{Y}_{{\rm o},p}}(\kappa, m) = \lambda \widehat{S}_{\widehat{Y}_{{\rm o},p}}(\kappa - 1, m) + (1 - \lambda)|\widehat{Y}_{{\rm o},p}(\kappa, m)|^2. \tag{18}$$

The proposed gain values $G(\kappa, m)$ are obtained by introducing (13) and (17) into (11):

$$G(\kappa, m) = \frac{\widehat{S}_E(\kappa, m) - \beta \widehat{S}_{E_{\rm nl}}(\kappa, m)}{\widehat{S}_E(\kappa, m)}. \tag{19}$$

The overestimation factor β can be used to account for systematic estimation errors and/or to tune the 'aggressiveness' of the AES to any given preferences. The actual values of $G(\kappa, m)$ are usually limited to a desired level of attenuation $G_{\rm min}$ in order to avoid severe degradation of near-end speech signals during double-talk situations. Then, (19) finally becomes

$$G_{\rm limited}(\kappa, m) = \max\left\{G(\kappa, m),\, G_{\rm min}\right\}. \tag{20}$$

In practice, the parameters $G_{\rm min}$ and β can be chosen as a compromise between attenuation of the nonlinear residual echo and tolerable distortion of the near-end signals.

4. SIMULATION RESULTS

To evaluate the performance of the proposed approach we present simulation results obtained for nonlinear acoustic echo control. In the first experiment, the unknown echo path has been modeled by the cascade of a fifth-order memoryless polynomial, followed by a linear filter g_n of length $N = 512$. In order to verify the robustness of the NAES with respect to echo path changes, the linear filter g_n of the echo path model has been shifted by ten samples after about 4 seconds of the input signal has been processed. The input signal has been speech, recorded at a sampling rate of 8 kHz. A white noise signal has been added to the microphone signal, where a signal-to-noise ratio (SNR) of 30 dB has been preset. Here, we consider a linear frequency-domain AEC with and without NAES, and a nonlinear AEC (NAEC) based on a frequency-domain adaptive power filter [5]. For both, the NAES and the NAEC, only the linear and the cubical ($p = 3$) channel has been used, where a fixed orthogonalization of the channel inputs has been applied. The considered structures use a length of $N = 512$ for all required filters. The gain factors of the NAES has been realized with an overestimation factor $\beta = 2$ and the gain values has been limited to $G_{\rm min} = 0.25$, implying a maximum attenuation of about 12 dB. As evaluation criterion we use the echo return loss enhancement (ERLE) defined as

$$\text{ERLE} = 10 \log \frac{\text{E}\{d^2(k)\}}{\text{E}\{z^2(k)\}}\, [\text{dB}]. \tag{21}$$

The ERLE graphs obtained for the different approaches are shown in Fig. 3. As can be seen, both nonlinear approaches remarkably improve the echo attenuation compared to the purely linear AEC. Since the NAES does not require an exact identification of the nonlinear echo path, its initial convergence is faster than in case of the NAEC. Moreover, the achievable echo attenuation is higher when applying the NAES. As can be clearly noticed, the major advantage of the proposed method is its robustness against echo path changes. While the NAEC only slowly readapts to the new echo path, the NAES is able to provide high attenuation values almost immediately after the echo path change.

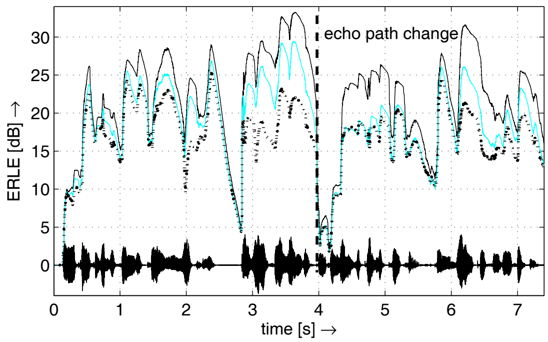

Fig. 3. Comparison of the proposed approach (black), a nonlinear AEC (gray), and a linear AEC (dotted). An echo path change takes place after approximately 4 seconds

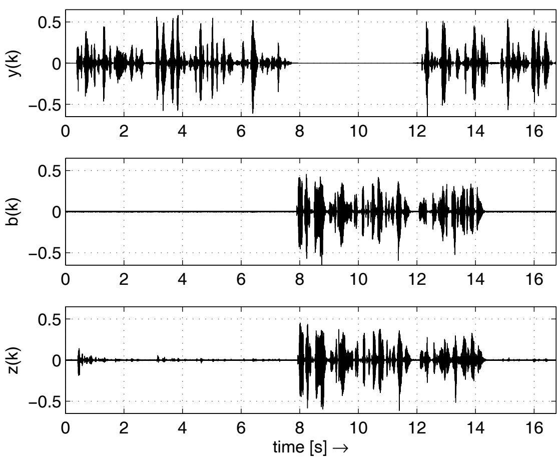

Fig. 4. Different signals for a nonlinear acoustic echo scenario: echo signal $y(k)$, near-end signal $b(k)$, and output of the NAES $z(k)$

In the next experiment we evaluate the performance of the proposed approach in double-talk situations. For the simulations, the same echo path has been used as in the previous set up. However, no echo path change has been simulated. The resulting echo signal $y(k)$ is shown at the top of Fig. 4. The near-end speech signal used for the simulations is depicted in the center of Fig. 4. The implementation of the linear echo canceler followed by the NAES has been the same as in the previous simulation. However, to account for the double-talk situations, the step-size for both, the linear AEC, and the NAES has been controlled using the approach presented in [9]. In Fig. 4 (bottom) the resulting residual signal $z(k)$ after processing by the NAES is shown. A visual comparison of the different signal waveforms in Fig. 4 indicates the suitability of the proposed approach: the near-end signal $b(k)$ is hardly distorted, while the echo signal $y(k)$ has been sufficiently suppressed. The audible distortions of the near-end signal during severe double-talk situations could be reduced by including a double-talk detection algorithm and a corrsponding control of the spectral subtraction approach.

The last experiment is based on recorded speech data from a loudspeaker of a mobile phone placed in an enclosure with low reverberation. For the recording, the loudspeaker has been mounted in the handset, while the microphone has been separated from it to avoid undesired vibration effects. As before, we use the parameters $N = 512$, $\beta = 2$, and $G_{\min} = 0.25$ for the implementation of the echo control, where the NAES only includes the cubical channel. In Fig. 5, the ERLE graphs obtained with NAES (black) and without NAES (gray) are shown. We notice that the nonlinear approach provides a persistent increase in echo attenuation throughout the whole simulation period.

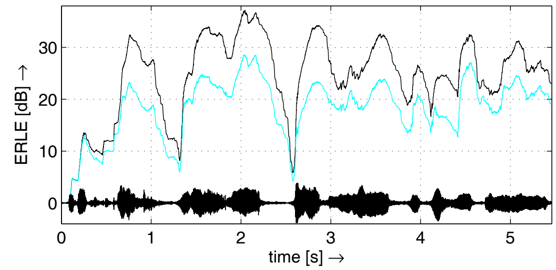

Fig. 5. Comparison of a linear AEC with and without NAES for recorded speech data.

5. CONCLUSION

We discussed the suppression of nonlinear acoustic residual echoes that remain when applying a linear AEC in the presence of nonlinear distortions. The proposed approach is based on a simple power filter model of the nonlinear echo path which is used to adjust the gain values of the NAES. The simulation results confirm the suitability of the proposed method for practical situations where the loudspeaker causes non-negligible nonlinear distortion.

6. REFERENCES

[1] B. S. Nollett and D. L. Jones, "Nonlinear echo cancellation for hands-free speakerphones," in *Proc. IEEE Workshop on Nonlinear Signal and Image Proc. (NSIP)*, Michigan, Sept. 1997.

[2] F. Kuech, A. Mitnacht, and W. Kellermann, "Nonlinear acoustic echo cancellation using adaptive orthogonalized power filters," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, Philadelphia, March 2005.

[3] O. Hoshuyama and A. Sugiyama, "An acoustic echo suppressor based on a frequency-domain model of highly nonlinear residual echo," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, Toulouse, May 2006.

[4] S. Gustafsson, R. Martin, and P. Vary, "Combined acoustic echo control and noise reduction for hands-free telephony," *Signal Processing*, vol. 64, no. 1, pp. 21–32, 1998.

[5] F. Kuech and W. Kellermann, "Orthogonalized power filters for nonlinear acoustic echo cancellation," *Signal Processing*, vol. 86, pp. 1168–1181, June 2006.

[6] S. Haykin, *Adaptive Filter Theory*, New Jersey: Prentice Hall, 1996.

[7] J. J. Shynk, "Frequency domain and multirate adaptive filtering," *IEEE Signal Processing Magazine*, pp. 14–37, Jan. 1992.

[8] C. Avendano, "Acoustic echo suppression in the STFT domain," in *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, New Paltz, Oct. 2001.

[9] G. Enzner and P. Vary, "Robust and elegant, purely statistical adaptation of acoustic echo canceler and postfilter," in *Proc. Int. Workshop on Acoustic Echo and Noise Control (IWAENC)*, Kyoto, Sept. 2003.