AN ACOUSTIC ECHO CANCELLER WITH COMPENSATION OF NONLINEARITIES

Alexander Stenger and Rudolf Rabenstein
Telecommunications Institute
University of Erlangen-Nuremberg
Cauerstrasse 7, 91058 Erlangen, Germany
e-mail: {stenger, rabe}@nt.e-technik.uni-erlangen.de

ABSTRACT

Traditional adaptation algorithms for acoustic echo cancellers are sufficiently robust to give satisfying performance even in the presence of nonlinear distortions, e.g. those induced by low-cost audio equipment in speech communication products for the consumer market. However, the performance of echo compensation can be improved if special measures are taken to acount for nonlinear behaviour in the electro-acoustic transmission system. To this end, a compensation method for memoryless nonlinearities, as encountered in overdriven amplifiers, is presented. Measurements demonstrate that the echo return loss enhancement can be improved without slowing down convergence. It is thus posssible to compensate the nonlinear effects of cheap audio components with few additional operations in the digital part of the device.

1 Introduction

The compensation of acoustic echos with adaptive filters is the state of the art in hands-free telephones and video conferencing. When high-quality audio equipment is used, then nonlinearities in the loudspeaker-enclosure-microphone system (LEMS) can be neglected. Under this assumption, the design of the adaptive filter and the adaptation algorithm can be carried out by methods from linear systems theory.

However, with decreasing prices for signal processing hardware, the above mentioned communication applications are penetrating the consumer market. Competition demands cutting prices by sacrificing sound quality of the analog equipment. The assumption of linearity no longer holds, due to distortions in the amplifier and speaker, when low quality equipment is used. On the one hand, this does not constitute a major problem, since the adaptation algorithms are sufficiently robust to handle also nonlinear distortions. On the other hand, their performance is impaired with respect to the achieveable echo reduction [1]. A method for compensating echoes in a commercial speakerphone using a time delay neural network has been proposed in [2], but it is numerically expensive and works only at modest loudspeaker power.

This paper presents a compensation method for the kind of nonlinearities encountered in low-cost audio equipment operated at high output power. With few additional complexity it enhances the performance of traditional adaptation algorithms when impaired by the distortions of an overdriven amplifier. After an analysis of the problem, we present the solution method and show its capability by measurements with low-cost equipment.

2 Problem Description

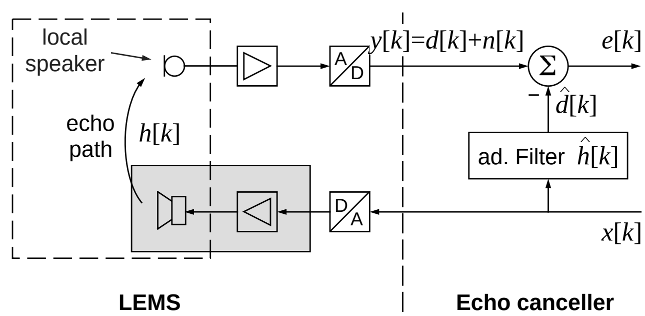

Figure 1: Nonlinearities in the electro-acoustic echo path (shaded area)

Fig. 1 shows the usual configuration of an adaptive filter for acoustic echo cancelling (see e.g. [3, 4]). $d[k]$ denotes the response of the LEMS to the far end speaker signal $x[k]$ under the assumption that this response can be modelled by a convolution with the room impulse response $h[k]$. When this impulse response is approximated by a digital counterpart $\hat{h}[k]$, then an estimate $\hat{d}[k]$ of the response $d[k]$ can be used to cancel unwanted echos. However, the microphone signal $y[k]$ not only contains the far end speaker response $d[k]$ but also additional signal components of different nature, denoted by $n[k]$. We discuss two kinds of additional signals $n[k]$: The local speaker in the enclosure and the contribution of nonlinear distortions, produced by low-cost audio equipment.

We use two figures to characterize the system performance in these cases. They are the signal-to-noise ratio (SNR) between $d[k]$ and $n[k]$ and the echo return loss enhancement (ERLE) between $y[k]$ and $e[k]$

$$\mathrm{SNR} = \frac{E\{d^2\}}{E\{n^2\}}, \qquad \mathrm{ERLE} = \frac{E\{y^2\}}{E\{e^2\}}. \tag{1}$$

Note that in our definition of ERLE the noise is included.

The case where $n[k]$ represents the local speaker signal is well studied, because it is the purpose of the system to transmit the local speaker to the far end. Assuming, that the LEMS is linear, the output of the adaptive filter $\hat{d}[k]$ can model the linear system response $d[k]$ arbitrarily well, if the adaption stepsize is appropriately chosen [3], [4], and enough filter taps for $\hat{h}[k]$ are used. Then the error $e[k]$ is determined only by $n[k]$ and thus the ERLE appoximates the SNR, if $E\{d^2\} \gg E\{n^2\}$

$$\mathrm{ERLE} = \frac{E\{y^2\}}{E\{e^2\}} = \frac{E\{d^2\} + E\{n^2\}}{E\{n^2\}} \approx \frac{E\{d^2\}}{E\{n^2\}} = \mathrm{SNR}. \tag{2}$$

Next, we consider the case where $n[k]$ represents nonlinear distortions of the analog equipment. This will mainly happen in the amplifier and loudspeaker for the far end speaker (shaded area in Fig. 1), since here the signal levels are higher than in the microphone path for the local speaker. In contrast to the case of local speaker activity considered above, we do not want to transmit the nonlinear distortions to the far end. Instead, we would like to suppress this signal by the echo canceller, as it is part of the echo.

Nonlinear behavior can be classified into nonlinearities with and without memory. Loudspeaker nonlinearities have to be modelled with memory [5]. We assume that they are neglectable compared to amplifier distortions, which we model without memory. Our treatment of nonlinear behaviour is based on this and two other assumptions compiled below:

- The nonlinearities are memoryless.
- The nonlinearities occur before the acoustic path, i.e. the nonlinear signal is convolved with the impulse response of a linear system.
- The nonlinearities are time-invariant, which excludes effects like temperature drift.

It has been shown [6, 7], that nonlinear behaviour can be characterized by the superposition of a linear system (here given by $h[k]$) and a nonlinear system, where the output signals of both systems are uncorrelated. Both systems can be obtained by measurement. Describing a nonlinear amplifier in such a way results in a distortion term $n[k]$, which is uncorrelated with the output $d[k]$ of the linear part. We can thus describe the performance by SNR and ERLE as above and conclude in the same way as for (2), that the ERLE approaches the SNR for small noise powers. We confirmed this by experiments with different SNRs and adaption stepsizes, where $x[k]$ was white and coloured gaussian noise. The SNR can be obtained from the measured power spectral densities of $d[k]$ and $n[k]$.

From this result we can conlude that it is also possible in the nonlinear case to find an arbitrary exact model $\hat{h}$ for the room impulse response h.

3 Solution Method

3.1 General

In this paper we demonstrate that nonlinear echo signals can be reduced by a simple nonlinear compensator, which costs only a few additional multiplications per sample.

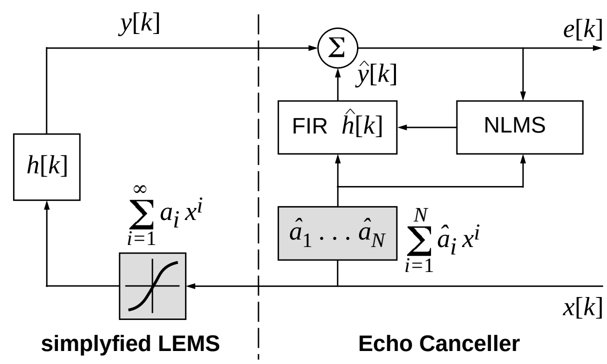

Figure 2: Compensation of memoryless nonlinearities

The left side of Fig. 2 shows a model for the LEMS, depicting the assumptions about nonlinearities made above. The right side proposes a nonlinear preprocessor for the signal $x[k]$ which is fed into the adaptive filter, so that the filter has to model mainly the linear part $h[k]$ of the LEMS. The preprocessor models a memoryless nonlinear function by a polynomial in x with a finite number N of coefficients.

3.2 Determination of Polynomial Coefficients

Assuming that the coefficients are time-invariant, we can determine them offline, and then run the system with a fixed set of coefficients. The measurement of polynomial coefficients is performed in two steps, during which the room impulse response $h_0[k]$ must be time invariant. First we need a linear room impulse response estimate

$\hat{h}_0[k]$. Even in the nonlinear case we can obtain it with arbitrary accuracy using the system in Fig. 1, as discussed in section 2.

In the second step the coefficients $\hat{a}_i$ of the nonlinear preprocessor are calculated. They are determined as estimates of the first N cofficients of the LEMS nonlinearity. This estimate is based on the impulse response $\hat{h}_0[k]$ measured in step 1. To this end, we send an appropriately distributed training sequence $x[k]$ with L samples through the LEMS and record the output y. Care must be taken that no local speech or noise is present during the recording, so that $n[k]$ is produced only by nonlinearities. We then calculate the estimated echo $\hat{y}[k]$, see Fig. 2, and interchange the sum and the convolution with the FIR filter coefficients $\hat{h}_0[k]$:

$$\hat{y}[k] = \sum_{i=1}^{N} \hat{a}_i \left(x^i[k] * \hat{h}_0[k]\right) = \sum_{i=1}^{N} \hat{a}_i \, \hat{y}_i[k] \, . \qquad (3)$$

Here $\hat{y}_i[k]$ is the partial response to the i-th power of the input signal $x[k]$. To express (3) in matrix notation, we arrange the values $\hat{y}_i[k]$, $k = 0, \ldots, L-1$ of each partial response in a vector $\hat{\mathbf{y}}_i$

$$\hat{\mathbf{y}}_\mathbf{i} = \left[y_i[0], y_i[1] \ldots y_i[L-1]\right]^T, \; i = 1, \ldots, N \qquad (4)$$

and combine these vectors to a $L \times N$ matrix

$$\hat{\mathbf{Y}} = \left[\hat{\mathbf{y}}_\mathbf{1}, \hat{\mathbf{y}}_\mathbf{2} \ldots \hat{\mathbf{y}}_\mathbf{N}\right] \, . \qquad (5)$$

With

$$\hat{\mathbf{a}} = \left[\hat{a}_1, \hat{a}_2 \ldots \hat{a}_N\right]^T \qquad (6)$$

we can express the estimated echo as

$$\hat{\mathbf{y}} = \hat{\mathbf{Y}} \, \hat{\mathbf{a}} \, . \qquad (7)$$

To minimize the power of the error signal $\left\|e[k]\right\|_2^2$, we have to solve

$$\mathbf{y} - \hat{\mathbf{Y}} \, \hat{\mathbf{a}} = \mathbf{0} \qquad (8)$$

in the least squares sense. This can be done with a pseudoinverse [8, Appendix 1] of $\hat{\mathbf{Y}}$:

$$\hat{\mathbf{a}} = (\hat{\mathbf{Y}}^\mathbf{T} \hat{\mathbf{Y}})^{-\mathbf{1}} \, \hat{\mathbf{Y}}^\mathbf{T} \, \mathbf{y} \, , \qquad (9)$$

where $\mathbf{y}$ is a vector containing the samples of the microphone signal $y[k]$, arranged as in (4). The resulting polynomial $\sum \hat{a}_i \, x^i[k]$ approximates the nonlinear behaviour of the echo path only if $x[k]$ stays within the amplitude range which has been used for the training sequence.

3.3 Operation of the Echo Canceller

Once we have measured the polynomial coefficients $\hat{a}_i$ using a fixed room impulse response estimate $\hat{h}_0[k]$, we can run the echo canceller shown in Fig. 2. The possibly time-variant room impulse response $h[k]$ is modelled with an adaptive filter $\hat{h}[k]$, while the coefficients $\hat{a}_i$ are fixed. They are unique for a given hardware and given signal distribution. The latter can be held constant in a practical system by a (digital) gain control which exploits the amplitude range of the D/A converter.

4 Results

For an experimental verification of the proposed compensation, the low-cost analog equipment for the far end speaker path according to Fig. 1 consisted of a cardboard-mounted speaker (0.1 Watts) and a standard one-chip amplifier (TBA 820). The equipment was placed in an anechoic chamber, thus avoiding local noise and undermodelling by too few adaptive filter taps. In our experiments, we used a 250 taps adaptive filter which could reach 45 dB ERLE with high quality audio equipment under such conditions.

The coefficients were determined with a white Gaussian training sequence being $L = 40000$ samples long. It was normalized so that its amplitude did not exceed ± 1. Fig. 3 shows the resulting polynomial of order $N = 13$

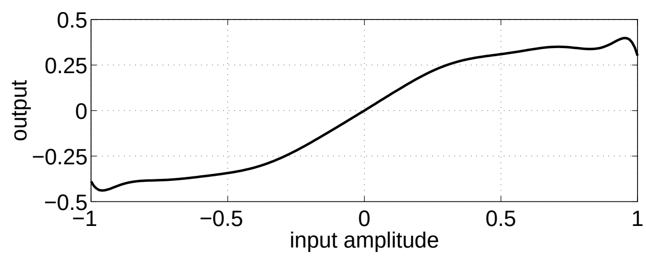

Figure 3: Estimated polynomial

which models a nonlinearity indicating an overdriven amplifier. All following results are obtained with this polynomial. Note that for input amplitudes near ± 1 the curve seems not to be very reliable. This is due to fact that the probability for the occurence of such values is low in our normalized Gaussian training sequence. This effect can be tolerated as experiments have shown that the best ERLE gain is achieved, if the amplitude distributions (i.e. probability density functions) of the training sequence and the test sequence are similar.

The ERLE curves in Fig. 4 compare the adaptation process in the presence of nonlinear distortions with and without compensation. They were obtained with an NLMS algorithm with stepsize 0.05 and 250 taps, and with a white Gaussian test signal $x[k]$ different from the training sequence. In addition, the microphone was placed in a different position than it was placed during the training to obtain a different room impulse response.

Two SNR levels are shown, where the SNR values are

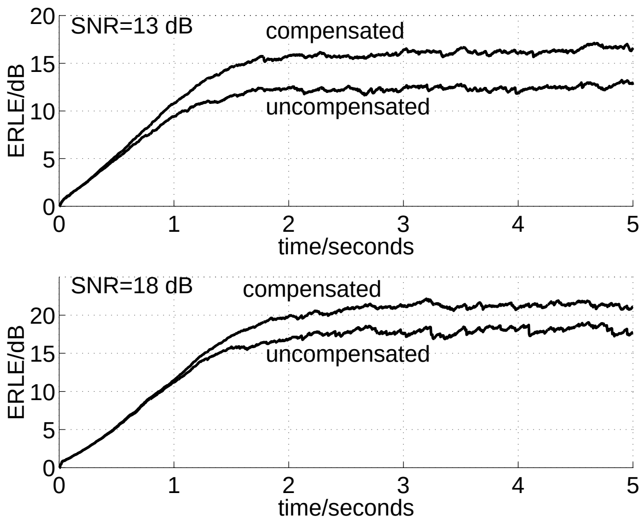

Figure 4: Improvement of ERLE due to compensation

measured using [7]. The value of 13 dB was obtained with full range amplitude, and 18 dB were obtained with amplitude ± 0.7. The ERLE curves of the uncompensated system reach the level of the corresponding SNR. In the compensated case, however, the effects modelled by a Nth order polynomial do not contribute to the distortion signal $n[k]$ any more. This results in a higher SNR value according to (1). Consequently, the ERLE curves with compensation reach a higher level. In addition, this gain is achieved at the same rate of convergence.

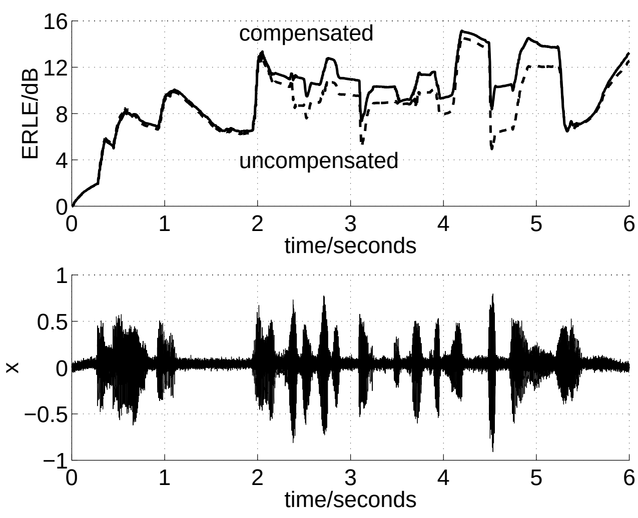

Figure 5: ERLE gain by compensation, speech signal

Fig. 5 shows that a white Gaussian training sequence is still appropriate, if the compensator is run with speech signals. At high amplitudes of the speaker signal x, the benefits of the compensator become visible.

We can summarize, that a nonlinearly distorted echo, which would be fully audible using an uncompensated adaptive filter, is reduced by approximately 4 dB with the new technique.

5 Conclusion

The nonlinear effects of low-cost audio components in speech communication products impair the effectiveness of echo cancellers. The reason for this is the additional distortion signal introduced by the nonlinearity. This results in a lower SNR also at times without local speaker activity and thus in a lowered upper bound on the achievable ERLE. This detrimental effect can be partially offset by a simple nonlinear compensator at the extra cost of a few operations per sample. Although the principle of the compensator and the calculation of its coefficients are based on a number of simplifying assumptions, an ERLE gain of 4 dB over the uncompensated case can be achieved at the same rate of convergence.

References

[1] A. N. Birkett, R. A. Goubran, "Limitations of handsfree acoustic echo cancellers due to nonlinear loudspeaker distortion and enclosure vibration effects", *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, New Paltz, New York*, Oct. 1995, pp. 13-16

[2] A. N. Birkett, R. A .Goubran, "Acoustic Echo Cancellation Using NLMS-Neural Network Structures", *Proc. ICASSP 1995, Detroit, MI*, pp. 2035-3038

[3] C. Breining, "Control of a hands-free telephone set" *Signal Processing*, vol. 61, 1997, pp.131-143

[4] U. Schultheiß, "Über die Adaption eines Kompensators für akustische Echos", *VDI-Fortschritt-Berichte*, Reihe 10, Nr. 90, Düsseldorf, Germany, 1988

[5] Franklin X.Y. Gao, W. Martin Sbelgrove, "Adaptive Linearization of a Loudspeaker," *Proc. ICASSP 91*, pp. 3589-3592.

[6] F. Heinle, R. Rabenstein, A. Stenger, "Measuring the Linear and Nonlinear Properties of Electro-Acoustic Transmission Systems," *Int. Workshop on Acoustic Echo and Noise Control (IWAENC) 1997*, pp. 33-36, London, September 1997.

[7] F. Heinle, R. Rabenstein, A. Stenger, "A measurement method for the linear and the nonlinear properties of electro-acoustic transmission systems", *Signal Processing*, vol. 64, No. 1, pp. 49-60

[8] Robert J. Schalkoff, "Pattern Recognition: Statistical, Structural and Neural Approaches", *Wiley*, 1992