AN IMPROVED ECHO SHAPING ALGORITHM FOR ACOUSTIC ECHO CONTROL

Rainer Martin and *Stefan Gustafsson*
IND, Aachen University of Technology
52056 Aachen, Germany
Tel: +49 241 806984; fax: +49 241 8888186
e-mail: martin@ind.rwth-aachen.de

ABSTRACT

This paper describes and analyses an improved algorithm for hands-free telephony which uses an acoustic echo canceller combined with an additional FIR-filter (called "echo shaping filter") in the sending path of the hands-free telephone. The algorithm controlling the filter is motivated by an approximation of an optimal least squares filter. Simulation results show that the algorithm allows to reduce the order of the echo canceller significantly, still providing high echo attenuation and low distortion of the near end speech signal during double talk. The modulation of the background noise caused by the echo shaping filter can be reduced by adding artificially generated noise to the output signal ("comfort noise").

1 Introduction

A hands-free telephone needs an echo suppression device which eliminates the feedback of the far end speech signal via the electro-acoustic loop. The acoustic echo canceller is considered to be the ideal solution for the echo reduction task because it allows true full duplex ("double talk") speech transmission. It requires, however, a large amount of computational resources and its echo attenuation is not always sufficient.

This paper deals with an acoustic echo control device which uses an acoustic echo canceller with a reduced number of filter taps and an additional FIR filter in the sending path of the hands-free telephone [5, 6, 7, 8]. This filter is adjusted to suppress those frequencies where the residual echo left by the echo canceller has more power than the near speech signal. As a result the residual echo attains a spectral shape similar to the near end signal and is therefore partially masked by the near end signal. In analogy with the well known noise shaping method used in speech coding [3] we call this approach "echo shaping" and the filter in the sending path the echo shaping filter. The adaptive echo shaping filter can be seen as a device to either increase the echo attenuation in case the echo canceller does not deliver enough attenuation, or, as a means to reduce the overall complexity since the echo canceller might be equipped with less coefficients. Compared to the conventional voice activated attenuator the echo shaping filter provides better speech quality and intelligibility during double talk situations.

2 Principles of the Echo Shaping Approach

Figure 1 shows a block diagram of the acoustic echo canceller C combined with an echo shaping filter H. We assume that the echo canceller C delivers a minimum echo attenuation of 8-10 dB.

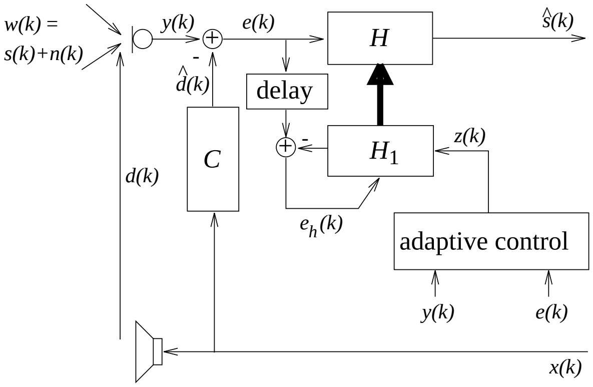

Figure 1: Block diagram of combined echo canceller and echo shaping filter.

The input signal of the adaptive filter H is the compensated signal,

$$\begin{aligned} e(k) &= s(k) + n(k) + d(k) - \widehat{d}(k) \\ &= w(k) + d(k) - \widehat{d}(k) \quad , \end{aligned} \tag{1}$$

where $s(k)$ is the near end speech signal, $n(k)$ is the near end noise signal, $d(k)$ is the echo signal at the microphone input, and $\widehat{d}(k)$ is the echo signal as estimated by the echo canceller C. In what follows we always assume that the signals $s(k)$, $n(k)$, and $d(k)$ are statistically independent and short time stationary.

The echo shaping filter H is adapted by means of a background filter H_1. The coefficients of H_1 are copied into filter H at each time instance. The input signal

$z(k)$ of H_1 is an adaptively mixed linear combination of the microphone signal $y(k)$ and the compensated signal $e(k)$. The reference signal of the adaptive filter is the compensated signal $e(k)$. To adapt the filter H_1 we exploit the fact that the compensated signal $e(k)$ contains less echo than the input signal $z(k)$.

3 An Ideal Control Algorithm

To develop insight into the proper control strategy for the echo shaping filter we first consider the least squares approach (Wiener filter). Figure 2 shows the signal flow diagram of the Wiener filter approach. We assume that the input to the Wiener filter is the compensated signal $e(k)$ and the reference signal is the near end signal $w(k) = s(k) + n(k)$.

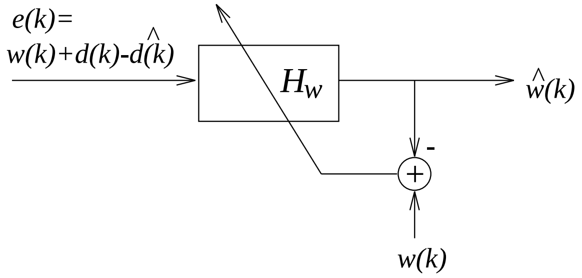

Figure 2: Block diagram of the Wiener filter.

The coefficients of filter H_w are adjusted to minimize the mean square error $E\{(w - \widehat{w})^2\}$. The frequency response of the resulting filter can be written as a function of power spectral densities $R(\Omega)$ [2]

$$H_w(\Omega) = \frac{R_{we}(\Omega)}{R_{ee}(\Omega)} = \frac{R_{ww}(\Omega)}{R_{ww}(\Omega) + R_{(d-\widehat{d})(d-\widehat{d})}(\Omega)} \quad . \tag{2}$$

If we approximate the echo attenuation of the compensator by $\delta \approx 10^{ERLE_C/20dB}$ ($ERLE_C$ denotes the Echo Return Loss Enhancement of the compensator C) we may write

$$H_w(\Omega) \approx \frac{R_{ww}(\Omega)}{R_{ww}(\Omega) + \delta^2 R_{dd}(\Omega)} = \frac{\rho(\Omega)}{\rho(\Omega) + \delta^2} \quad , \tag{3}$$

where $\rho(\Omega)$ denotes the near-signal-to-echo ratio at the microphone input,

$$\rho(\Omega) = \frac{R_{ww}(\Omega)}{R_{dd}(\Omega)} = \frac{R_{ss}(\Omega) + R_{nn}(\Omega)}{R_{dd}(\Omega)} \quad . \tag{4}$$

Figure 3 plots the attenuation characteristic of the Wiener filter $H_w(\Omega)$ as a function of the near-signal-to-echo ratio $\Theta(\Omega) = 10\log_{10}(\rho(\Omega))$ and the compensator attenuation δ. It can be seen that the attenuation of the echo shaping filter is reduced as the compensator attenuation gets larger

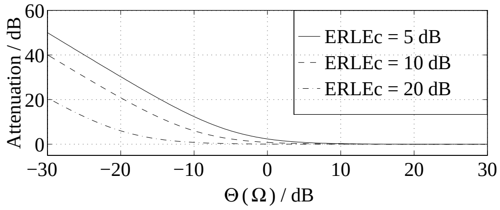

Figure 3: Attenuation characteristics of the Wiener filter ($ELRE_C = -20\log_{10}(\delta)$).

3.1 Development of the Control Algorithm

The input signal $z(k)$ of the NLMS-adapted background filter H_1, as shown in Figure 1, is a combination of the microphone signal $y(k)$ and the compensated signal $e(k)$:

$$z(k) = \alpha(k)y(k) + (1 - \alpha(k))e(k) \quad , \tag{5}$$

where the time varying non-negative control factor $\alpha(k)$ ($\alpha(k) \geq 0$) is controlled such that a required minimum echo attenuation during single talk and an admissible near end speech signal distortion during double talk is achieved.

We assume in a first order approximation that the echo attenuation of the echo canceller can be modelled by a scalar factor

$$d(k) - \widehat{d}(k) = \delta(k)d(k) \quad . \tag{6}$$

Then the combined input signal in equation (5) can be rewritten as

$$\begin{aligned} z(k) &= \alpha(k)y(k) + (1 - \alpha(k))e(k) \\ &= s(k) + n(k) + d(k)\phi(k) \quad , \end{aligned} \tag{7}$$

where the control parameter $\phi(k)$ is given by

$$\phi(k) = \delta(k) + \alpha(k)(1 - \delta(k)) \quad . \tag{8}$$

Since the adaptive filter H_1 approximates a non causal Wiener filter, the frequency response $H(\Omega, k)$ at time instance k can be written as a function of the time varying control factor $\phi(k)$,

$$H(\Omega, k) = \frac{R_{ez}(\Omega)}{R_{zz}(\Omega)} \approx \frac{\rho(\Omega) + \delta\phi(k)}{\rho(\Omega) + \phi^2(k)} \quad , \tag{9}$$

where $\rho(\Omega)$ is defined as in equation (4).

Now we want the control algorithm to approximate the ideal control algorithm as found above. If we postulate $H_w(\Omega) = H(\Omega, k)$, substitute $\rho(\Omega)$ by its estimated mean $\bar{\rho}(k)$, and solve for $\phi(k)$ we obtain

$$\phi(k) = \frac{\delta(k)\bar{\rho}(k) + \delta^3(k)}{2\bar{\rho}(k)} + \sqrt{\frac{(\delta(k)\bar{\rho}(k) + \delta^3)^2}{4\bar{\rho}^2(k)} + \delta^2(k)} \quad , \tag{10}$$

which can be simplified by a truncated Taylor series to

$$\phi(k) = 1.618\delta(k) + 0.9472\frac{\delta^3(k)}{\bar{\rho}(k)} \quad . \tag{11}$$

Hence, in equation (8), $\alpha(k)$ should be choosen to

$$\alpha(k) = \frac{1}{1-\delta(k)} \left(0.618\delta(k) + 0.9472\frac{\delta^3(k)}{\bar{\rho}(k)}\right) \quad . \tag{12}$$

3.2 The Estimation of $\delta(k)$ and $\bar{\rho}(k)$

$\delta(k)$ can be estimated using recursive first order power measurements of the signals $y(k), e(k)$ and $\widehat{d}(k)$. Recall that

$$\begin{aligned} \sigma^2_{yy}(k) &= \sigma^2_{ss}(k) + \sigma^2_{nn}(k) + \sigma^2_{dd}(k) \\ \sigma^2_{ee}(k) &= \sigma^2_{ss}(k) + \sigma^2_{nn}(k) + \sigma^2_{(d-\widehat{d})(d-\widehat{d})}(k) \quad . \end{aligned} \tag{13}$$

Using the assumption in equation (6) we obtain

$$\begin{aligned} \sigma^2_{dd}(k) &= \frac{1}{(1-\delta(k))^2}\sigma^2_{\widehat{dd}}(k) \\ \sigma^2_{(d-\widehat{d})(d-\widehat{d})}(k) &= \frac{\delta(k)^2}{(1-\delta(k))^2}\sigma^2_{\widehat{dd}}(k) \quad , \end{aligned} \tag{14}$$

and by combining the above equations we get an expression for $\delta(k)$,

$$\delta(k) = \frac{\sigma^2_{yy}(k) - \sigma^2_{ee}(k) - \sigma^2_{\widehat{dd}}(k)}{\sigma^2_{yy}(k) - \sigma^2_{ee}(k) + \sigma^2_{\widehat{dd}}(k)} \quad . \tag{15}$$

It turns out, however, that this estimation procedure is very sensitive, as is the control law in equation (12) with respect to estimation errors of $\delta(k)$. Therefore, during single talk of the far end speaker $\delta(k)$ is estimated by the simplified expression

$$\delta(k) = \sqrt{\frac{\sigma^2_{ee,speech}(k)}{\sigma^2_{yy,speech}(k)}} \quad , \tag{16}$$

where $\sigma^2_{ee,speech}(k)$ and $\sigma^2_{yy,speech}(k)$ are the power of the speech components in the signals $e(k)$ and $y(k)$, respectively. These can be estimated using the method described in [4].

With the above estimate for $\delta(k)$ it is now possible to calculate $\bar{\rho}(k)$ as

$$\bar{\rho}(k) = \frac{\sigma^2_{ss}(k) + \sigma^2_{nn}(k)}{\sigma^2_{dd}(k)} = \frac{(1-\delta(k))^2\sigma^2_{yy}(k)}{\sigma^2_{\widehat{dd}}(k)} - 1 \quad . \tag{17}$$

Now $\alpha(k)$ can be estimated according to equation (12). The above calculations are relevant only when the far speaker is active, because otherwise neither $\delta(k)$ nor $\bar{\rho}$ are well defined. Therefore $\alpha(k)$ is set to $\alpha(k) \equiv 0$ whenever the far speaker is not active. This is controlled by a voice activity detector.

3.3 Comfort Noise

The time varying echo shaping filter may give rise to modulations of the background noise at the near side. To make this less irritating to the far end listener, a comfort noise generator can be used. Ideally, the noise in the signal transmitted to the far side should have the same power density spectrum as the noise $n(k)$ recorded by the microphone.

In our system the LPC-coefficients [3] of the noise are estimated during speech pauses. When the far speaker is active, a white noise is filtered with a LPC synthesis filter and amplified to obtain a noise $\widehat{n}(k)$ that approximates the power density spectrum of $n(k)$. To adapt the noise to the time varying echo shaping filter, $\widehat{n}(k)$ should be filtered with a filter $G(k)$, where $|G(\Omega)|^2 = 1 - |H(\Omega)|^2$. However, due to the complicated calculations of the coefficients of the filter G we use the approximation $\widehat{G}(k) = \delta(k - \frac{N_H}{2}) - H(k)$, which yields $|\widehat{G}(\Omega)|^2 = |e^{-j\frac{N_H}{2}} - H(\Omega)|^2$, where $N_H + 1$ is the order of the filter H. The filtered noise is then added to $\widehat{s}(k)$.

Our experiments show that the order of the LPC-estimator and synthesis filter must be about 15-20 to successfully estimate the background noise density spectrum if it contains several strong harmonic components. If it is almost white, fewer coefficents are sufficient.

Experiments also show that the power of the comfort noise may be 3 to 6 dB less than the power of the input noise $n(k)$ without sever impact on the auditive perception.

4 Experimental Results

4.1 Instrumental Assessment

Our evaluation method is based on a separate processing of the acoustic echo and the near end signal [1, 7]. This evaluation scheme requires, however, that the near end speech signal is recorded independently from the far echo signal. Having these signals we can calculate:

- $ERLE_C$: the time average of the echo return loss enhancement of the compensator $\mathbf{C}(k)$,
- $ERLE_{CH}$: the time average of the echo attenuation of the combined system (compensator C + echo shaping filter H), and
- $SEGSNR$: the distortion of the near end signal caused by the echo shaping filter measured by the segmental SNR.

To increase the significance of the SEGSNR as a distortion measure the echo shaping filter is implemented as a linear phase FIR filter. Thus, the $SEGSNR$ criterium measures the spectral amplitude distortions of the near end signal. For more details about the simulation conditions see [8].

4.2 Single Talk Results

During single talk of the far end speaker we are mainly interested in the echo attenuation as a function of the compensator order, i.e. as a function of the overall algorithm complexity, and furthermore in the performance at high near end noise levels. To test our algorithms we measured real signals in a car environment (reverberation time approximately 0.07 s). The simulations were run with compensator orders $N_C = 100, 200, 300$ and near-signal-to-noise ratios $SNR_n^s = 25$ dB and $SNR_n^s = 10$ dB.

The mean values of the Echo Return Loss Enhancement averaged over all 16 simulation runs are presented in Table 1. Each speech sample had a duration of 4 s.

	$SNR_n^d = 25$ dB			$SNR_n^d = 10$ dB		
N_C	100	200	300	100	200	300
$ERLE_C$ dB	7.6	20.6	28.5	7.4	20.1	25.9
$ERLE_{CH}$ dB	34.6	42.3	42.2	30.9	37.2	38.1

Table 1: $ERLE_C$ and $ERLE_{CH}$ during single talk.

It can be seen that the algorithm can adjust to the insufficient attenuation of the compensator (e.g. for $N_C = 100$). For $N_C = 200$ and $N_C = 300$ the $ERLE_{CH}$ is about the same, only depending on the SNR_n^d. When more background noise is present a lower attenuation is tolerable as the echo will be partly masked by the near signal. Since the echo shaping filter is used with a compensator of relative low order the convergence of the compensator in the presence of high noise levels is still very satisfactory. Due to the time varying echo shaping filter in the sending path the output signal will be modulated.

4.3 Double Talk Results

To test the algorithms under double talk conditions eight speech samples of far end speech were processed whereas eight other samples served as near end speech. The duration of each sample was about 4 s. The near-speech-to-echo ratio SNR_d^s was varied to simulate different distances of the near speaker to the microphone, or different coupling strengths between the loudspeaker and the microphone. The compensator was of order $N_C = 200$ for all experiments. The near-speech-to-noise ratios SNR_n^s were 25 dB and 10 dB in these experiments. The results for double talk are summarized in Table 2.

	$SNR_n^s = 25$ dB			$SNR_n^s = 10$ dB		
SNR_d^s dB	-20	0	20	-20	0	20
$ERLE_C$ dB	21.7	18.8	16.2	21.7	17.1	8.5
$ERLE_{CH}$ dB	31.1	22.6	22.1	30.5	21.0	12.3
$SEGSNR$ dB	4.6	13.1	15.7	5.9	18.5	24.6

Table 2: $ERLE_C$, $ERLE_{CH}$ and $SEGSNR$ during double talk.

If the power of the near end signal is significantly higher than the power of the far echo there is only a low additional echo suppression, and consequently a low distortion of the near end signal. When the near end signal has significantly less power than the echo the near signal will be distorted but still highly intelligible. As the near end signal power rises the noise modulations are less disturbing. During double talk the algorithm suffers from estimation errors in $\delta(k)$ [8].

5 Conclusions

Since the FIR acoustic echo canceller alone does not deliver sufficient echo attenuation under all circumstances an additional echo attenuation device is needed. Using the attenuation characteristics of an ideal Wiener filter a control algorithm for the additional filter is developed. Simulation results show that the combined configuration delivers a high echo attenuation even when using a relatively short echo compensator, and that the system complexity in this way may be reduced. The disadvantage of the system is that background noise will be modulated. This can be countered by using comfort noise.

References

[1] S. Gustafsson, R. Martin, and P. Vary, "On the Optimization of Speech Enhancement Systems Using Instrumental Measures." Proc. Workshop on Quality Assessment in Speech, Audio and Image Communication, Darmstadt, Germany, March 11-13, 1996.

[2] S. Haykin, "Adaptive Filter Theory." Prentice Hall, 1986.

[3] N. Jayant and P. Noll, "Digital Coding of Waveforms." Prentice-Hall, 1984.

[4] R. Martin, "An Efficient Algorithm to Estimate the Instantaneous SNR of Speech Signals.", Proc. EUROSPEECH '93, pp. 1093-1906, Berlin.

[5] R. Martin and J. Altenhöner, "Coupled Adaptive Filters for Acoustic Echo Control and Noise Reduction." Proc. Int. Conf. Acoustics, Speech, Signal Processing, Detroit, pp. 3043-3046, May 1995.

[6] R. Martin, "Combined Acoustic Echo Cancellation, Spectral Echo Shaping, and Noise Reduction." Proc. Fourth Int. Workshop on Acoustic Echo and Noise Control, pp. 48-51, Røros, Norway, June 1995.

[7] R. Martin, "Hands-free Telephones based on Multi-Microphone Echo Cancellation and Noise Reduction." Dissertation (in German), Aachen University of Technology, June 1995.

[8] R. Martin and S. Gustafsson, "The Echo Shaping Approach to Acoustic Echo Control." Submitted to Speech Communication.