OPTIMIZATION OF AN ACOUSTIC ECHO CANCELLER COMBINED WITH ADAPTIVE GAIN CONTROL

Peter Heitkämper

Institut für Netzwerk- und Signaltheorie, Technische Hochschule Darmstadt,
Merckstr. 25, D-64283 Darmstadt, Germany
pheit@nesi.e-technik.th-darmstadt.de

ABSTRACT

This contribution deals with the application of hands-free algorithms in wide-band telephone systems with limited computing resources. The basis is the combination of an acoustic echo canceller and an adaptive gain control method. The paper describes how the effect of the echo canceller can be optimized without increasing the computational expense. This is achieved by extending the length of the adaptive filter at the cost of a reduced affected frequency range. The study shows to which extent one can concentrate on the low frequency portion of the acoustic echoes in a wide-band application, if a suitable additional gain control method is available. The optimization is accomplished using a simple room model. Real-time measurements were carried out using an implementation of the complete hands-free system on a single DSP.

1. INTRODUCTION

Using hands-free telephone systems, the speech signals to be transmitted are usually disturbed by background noise and reverberation. Besides that, acoustic echoes of the far-end speaker's signal may be transmitted as well. They can disturb the conversation and even cause feedback howling.

There are many suggestions to solve this problem while there is still a need for solutions which operate in wide-band applications with modest computational demands [3]. In this context, a gain control method has been introduced in [4], which can be fruitfully combined with an adaptive echo canceller [5]. With the gain control method acoustic echoes of wide bandwidth can be attenuated consuming little processing power at the cost of a reduced double-talk capability. On the other hand, echo cancellers are known to compensate acoustic echoes without influencing the local signal. However, they consume a considerable amount of computing power, which increases with growing bandwidth and reverberation times.

Modern telephone systems for consumer applications can not be expected to be equipped with more than one single digital signal processor. Therefore, optimized combinations of signal processing algorithms are necessary in order to provide high quality devices even with limited resources. The key idea is to concentrate each signal processing method on that part of the task where it is most effective.

Section 2 shows a profitable partitioning of the frequency domain. In section 3 the minimum compensation error for a given partition is derived. The partitioning is then optimized in section 4 using a simple room model. Experimental results of a real-time implementation in section 5 exemplify the practical relevance of this contribution.

2. SUBBAND ECHO CANCELLER AND GAIN CONTROL

The impulse response of an office room can have a length of several hundred milliseconds. The tap coefficients of the adaptive echo cancellation filter are adjusted to model this impulse response. Usually an FIR-filter is involved, as this is appropriate for modelling a room impulse response [2]. For high quality conversations sampling rates up to f_s=24kHz are used. Under these circumstances a precise modelling of the room impulse response would require an echo canceller with several thousand coefficients.

In telephony, primarily speech signals are transmitted. In these signals most of the power is embedded in the lower frequencies, whereas the high frequency portions are only important for the speech quality [7]. An illustrative example for the time average speech spectrum is given in figure 1.

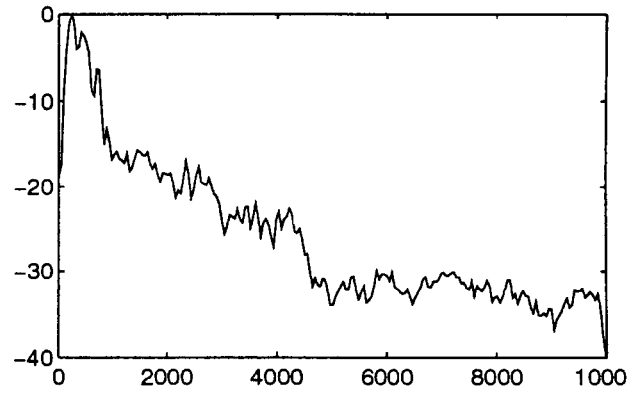

Figure 1: average power spectrum of speech in dB versus f/Hz

It was computed using speech signals of 15s length sampled with 24kHz averaging the short-term fourier transforms of 512 point frames. The speech samples originated from five speakers, female and male.

0-7803-2431-5/95 $4.00

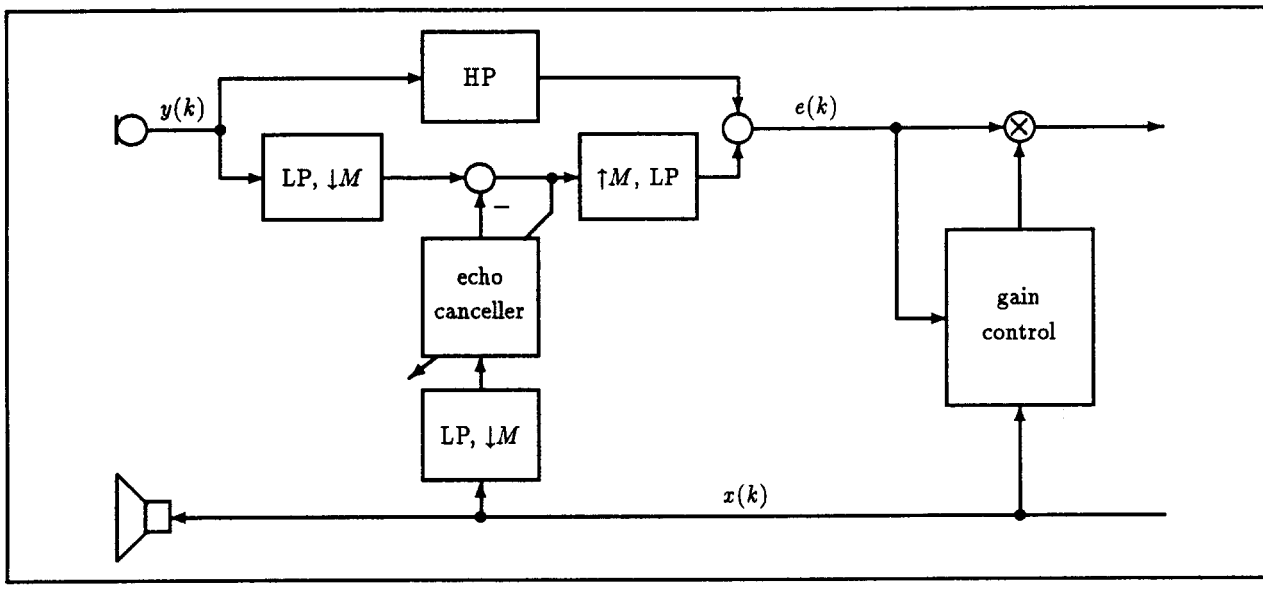

Figure 2: concept of the algorithm

The crucial quantity that causes the reduced double-talk capability in the gain control method is the power of the acoustic echoes [4]. The idea is therefore to run an echo canceller with increased length only in a low frequency subband in order to compensate the main part of the acoustic echoes. The remaining ,weaker high frequency portion can be suppressed easily by the gain control method, whereby the wide transmission-bandwidth is not affected. A similar consideration with different echo control methods was used in an 8kHz-algorithm [1], where no optimization of the downsampling factor is described.

Figure 2 illustrates the concept of the algorithm. LP and HP denote low-pass and high-pass filters, which are assumed to be ideal. $\downarrow M$ and $\uparrow M$ indicate downsampling and upsampling by a factor M, which is possible due to the reduced bandwidth of the echo canceller signals. With the reduced sampling rate, N coefficients of the adaptive FIR-filter can cover NM values of the wide-band system. As the adaptive filter operates with a reduced sampling rate, the calculations for one filter update can be spread over M samples. Consequently, the number N of coefficients which can be realized with a given processing power is almost proportional to M.

The gain control method is described in detail in [4]. This second method is necessary, as for the high frequency portion of the acoustic echoes no attenuation is achieved by the subband echo canceller. The gain control method realizes a relation between the level of its input $e(k)$ and the level of the output to be transmitted as shown in figure 3. Above a threshold S_0, the input is assumed to originate from the local speaker and is transmitted with almost a constant level. Below the threshold the input is distinctly attenuated as desired for background noise and residual echoes. The threshold S_0 is adapted to the background noise level. In the presence of acoustic echoes it is raised automatically to guarantee the attenuation. S_0 has to be raised less the weaker the echoes are.

3. MINIMUM ERROR

In order to achieve the highest possible echo attenuation, a lower bound for the compensation error is derived. Let $x(k)$ denote the far-end speaker's signal that causes the acoustic echoes and drives the adaptation of the echo canceller. The signal can be split in two portions

$$x(k) = x_L(k) + x_H(k), \tag{1}$$

where $x_L(k)$ contains the lower frequencies up to $f_s/2M$ and $x_H(k)$ the range from $f_s/2M$ to $f_s/2$. The system of loudspeaker, room, and microphone is described by its impulse response $g(k)$ leading to the microphone signal

$$y(k) = g(k) * x(k) + n(k), \tag{2}$$

where $*$ indicates the convolution of the signal $x(k)$ and the impulse response $g(k)$. The local signal is described by $n(k)$, but will be neglected during the analysis of the

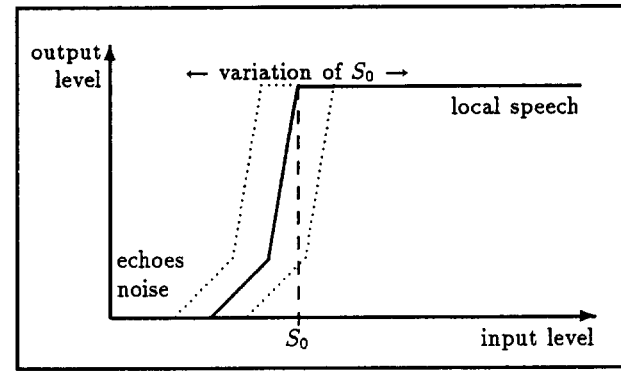

Figure 3: gain control: relation between output and input level

minimum error. The error signal after the compensation is

$$e(k) = y(k) - \hat{y}(k), \tag{3}$$

where $\hat{y}(k)$ is the interpolated highly sampled canceller signal. Assume the impulse response $g(k)$ is partitioned into

$$g_1(k) = \begin{cases} g(k) & 0 \le k < NM \\ 0 & \text{otherwise} \end{cases},$$

$$g_2(k) = \begin{cases} g(k) & k \ge NM \\ 0 & \text{otherwise} \end{cases}, \tag{4}$$

and the part $g_1(k)$ is split in two portions

$$g_1(k) = g_{1L}(k) + g_{1H}(k) \tag{5}$$

which correspond to the frequency contents of $x_L(k)$ and $x_H(k)$. With $\hat{y}(k) = \hat{g}(k) * x_L(k)$, the error 3 can be written

$$\begin{aligned} e(k) &= g(k) * x(k) - \hat{g}(k) * x_L(k) \\ &= g(k) * x_H(k) + g_2(k) * x_L(k) + g_{1L}(k) * x_L(k) \\ &\quad + g_{1H}(k) * x_L(k) - \hat{g}(k) * x_L(k). \end{aligned} \tag{6}$$

In the case of perfect adaptation the interpolated echo canceller impulse response equals the low frequency portion of the subsystem $g_1(k)$:

$$g_{1L}(k) = \hat{g}(k). \tag{7}$$

As $g_{1H}(k)$ and $x_L(k)$ are orthogonal, their convolution vanishes, and we obtain the error in the case of perfect adaptation of the subband canceller of limited length N, neglecting local signals:

$$\begin{aligned} e_{\min}(k) &= g(k) * x_H(k) + g_2(k) * x_L(k) \\ &= \sum_{i=0}^{\infty} g(i)\, x_H(k-i) + \sum_{i=NM}^{\infty} g(i)\, x_L(k-i). \end{aligned} \tag{8}$$

Due to the ideal assumptions, this error can be seen as a lower bound of the compensation error. The first term is the high frequency portion of $e(k)$. It originates from the portion of $x(k)$ which is not affected by the echo canceller. It can be assumed that the power of this term is very small for small values of M, as the signal $x_H(k)$ contains little power. With increasing values of M it becomes stronger, since more of the dominating portions of $x(k)$ contribute to $x_H(k)$.

The second term of 8 stems from the incomplete replica of the system $g(k)$ by the subband canceller of length N. As the number N of coefficients that can be realized with a given processing power is almost proportional to M, the lower summation limit of the second sum increases with M^2. The largest values of $g(k)$ can be expected at the beginning of the impulse response, as the echoes decay exponentially with the propagation time. So the power of this second term is expected to decrease with increasing M.

There is a value of the downsampling factor M for which the power of the lower bound $e_{\min}(k)$ becomes minimum. This value depends on the room impulse response and the spectrum of the excitation and will be analyzed in the following.

4. OPTIMIZATION USING ROOM MODEL

Using a stochastic model $g(k)$ for the impulse response and $x(k)$ for the excitation, the mean power of the minimum error is described by the expectation

$$\begin{aligned} \mathrm{E}\{e_{\min}^2(k)\} = &\sum_{i=0}^{\infty}\sum_{j=0}^{\infty} s_{gg}(i-j) s_{x_H x_H}(i-j) + \\ &\sum_{i=NM}^{\infty}\sum_{j=NM}^{\infty} s_{gg}(i-j) s_{x_L x_L}(i-j) \end{aligned} \tag{9}$$

with the autocorrelation functions $s_{gg}(k)$, $s_{x_H x_H}(k)$ and $s_{x_L x_L}(k)$. A simple model for the impulse response is

$$g(k) = s(k)\, h(k), \tag{10}$$

where the zero-mean sign process $s(k)$ takes the values -1 and 1 and $s(k_1)$, $s(k_2)$ are uncorrelated for $k_1 \neq k_2$. The deterministic function $h(k)$ equals zero for negative k and is exponentially decaying for positive k. With this model the mean power becomes

$$\mathrm{E}\{e_{\min}^2(k)\} = \sum_{i=0}^{\infty} h^2(i) s_{x_H x_H}(0) + \sum_{i=NM}^{\infty} h^2(i) s_{x_L x_L}(0). \tag{11}$$

We are interested in the relation between this expectation and the mean power of the microphone signal

$$\mathrm{E}\{y^2(k)\} = \sum_{i=0}^{\infty} h^2(i) s_{xx}(0), \tag{12}$$

as this is the echo attenuation achieved by the echo canceller.

As an example a single exponential function is used for $h(k)$:

$$h(k) = \begin{cases} A e^{-\gamma k} & i \ge 0 \\ 0 & \text{otherwise} \end{cases}, \tag{13}$$

where the factor γ and the sampling frequency f_s determine the reverberation time $T_r = \ln 10^3 / \gamma f_s$. Expressing the autocorrelation functions by integrals of the power spectrum $S_{xx}(\Omega)$, we obtain the echo attenuation for this example:

$$\frac{\mathrm{E}\{e_{\min}^2(k)\}}{\mathrm{E}\{y^2(k)\}} = \frac{\int_{\pi/M}^{\pi} S_{xx}(\Omega)\, d\Omega}{\int_{0}^{\pi} S_{xx}(\Omega)\, d\Omega} + e^{-2\gamma NM} \frac{\int_{0}^{\pi/M} S_{xx}(\Omega)\, d\Omega}{\int_{0}^{\pi} S_{xx}(\Omega)\, d\Omega}. \tag{14}$$

In an implementation, which is described in section 5, the number of coefficients that can be realized is approximately $N \approx -141 + 127M$, $M \ge 2$. With this relation and the time average of figure 1 for the power spectrum, expression 14 can be plotted as a function of the downsampling factor M for different reverberation times, as given in figure 4. It shows that for a wide range of reverberation times a higher attenuation is achieved by using a subband canceller and increasing the downsampling factor. With further reduced bandwidth less attenuation is achieved as the first term of 8 becomes predominant. In office rooms, T_r usually lies between 100ms and 400ms. In this example the optimum downsampling factor would be between 4 and 6.

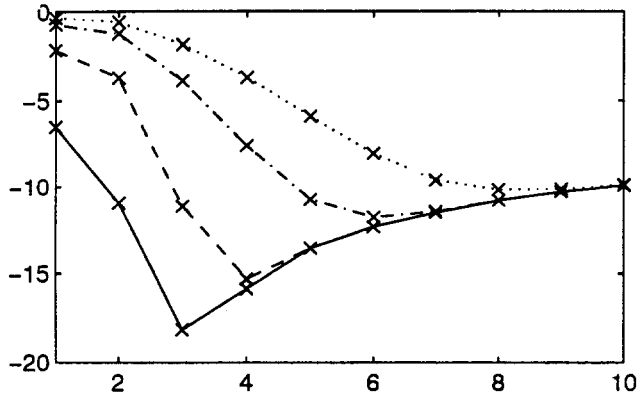

Figure 4: $10\log\left(\mathrm{E}\{e^2_{\min}(k)\}/\mathrm{E}\{y^2(k)\}\right)$ versus downsampling factor M, reverberation time 50ms (—), 150ms (– –), 450ms (– ·), 950ms (···)

5. EXPERIMENTAL RESULTS

The optimum downsampling factor M for which the highest attenuation is achieved depends on the impulse response $g(k)$ and on the spectrum of the excitation. Due to this dependence, no general optimum can be found. For a variety of systems and speakers, however, real-time measurements were carried out, which indicate that the optimal value is approximately the same for many cases. The best measured result is given in figure 5. It shows the echo attenuation as

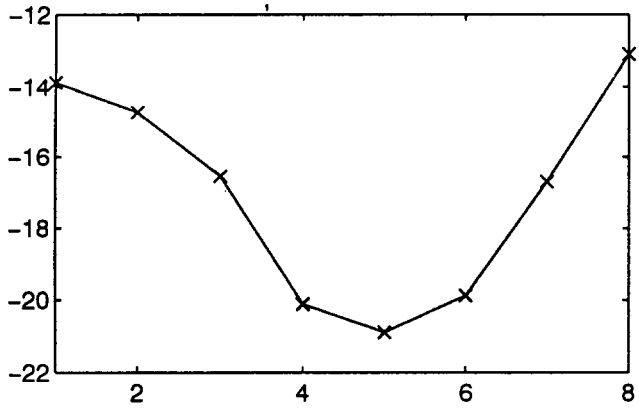

Figure 5: $20\log\left(\overline{|e(k)|}/\overline{|y(k)|}\right)$ versus downsampling factor M

the relation between the average magnitudes of the error signal and the microphone signal $20\log\left(\overline{|e(k)|}/\overline{|y(k)|}\right)$. It was measured in an office room using speech signals $x(k)$ in the presence of background noise. The NLMS algorithm for the adaptation of the coefficients was implemented on a Motorola DSP56156 consuming about half the processing power of this device. The sample frequency was 24kHz and the number of coefficients as given in section 4. An optimal value, which is also valid for many other constellations, can be found for $M = 5$, which corresponds to a filter length of about 100ms and a bandwidth of 2400Hz.

Impulse responses of practical systems usually differ from the simple example used in section 4 in that they have large values at the beginning, which correspond to the coupling of the direct sound wave [4], and that their envelope is better approximated by a sum of exponentials [6]. In the measuring arrangement the distance between loudspeaker and microphone was approximately 35cm, so that the direct sound wave was predominant and higher echo attenuation was possible than in the model of section 4. In this investigation, however, it is the dependence between the attenuation and the downsampling factor that is in the foreground of discussions and not the absolute values.

For the far-end listener, the improved attenuation between the optimized subband canceller and the conventional full-bandwidth version is clearly audible.

6. CONCLUSIONS

In all cases that were investigated a higher echo attenuation could be achieved by limiting the affected frequency range of the echo canceller and increasing the number of filter coefficients. The optimum downsampling factor depends on the given environment as well as on the computing resources that are available. However, in both the theoretical example and the measured result, the curves in the vicinity of the minima are flat. Thus, even if the optimum downsampling factor is not met precisely, an improved attenuation close to the optimum case can be achieved.

Due to the reduced power of the acoustic echoes, the influence of the gain control method on the transmitted signal can be reduced. This is done automatically so that the double-talk capability is maintained.

With the proposed optimized combination of a subband echo canceller and gain control, undisturbed hands-free conversations are possible with wide bandwidth using a single low-cost DSP.

7. REFERENCES

[1] W. Armbrüster, "Wideband Acoustic Echo Canceller with Two Filter Structure," Proceedings EUSIPCO 92, Bruxelles, Belgium, 1992 Vol. 3, pp. 1611-1617.

[2] S. Gudvangen and S.J. Flockton, "Comparison of Pole-Zero and All-Zero Modelling of Acoustic Transfer Functions." Electronics Letters Nr. 28, 1992, pp. 1976-1978.

[3] E. Hänsler, "The hands-free telephone problem — An annotated bibliography update," Annales des Télécommunications T.49, Nr. 7-8, 1994, pp. 360-367.

[4] P. Heitkämper and M. Walker, "Adaptive Gain Control for Speech Quality Improvement and Echo Suppression," Proceedings IEEE ISCAS, Chicago, Illinois, 1993 Vol. 1, pp. 455-458.

[5] P. Heitkämper and M. Walker, "Adaptive Gain Control and Echo Cancellation for Hands-free Telephone Systems," Proceedings EUROSPEECH 93, Berlin, Germany, Sept. 21-23, 1993.

[6] J. Marx, "Kompensation akustischer Echos in Räumen," Fortschritte der Akustik – DAGA 94, Dresden, Germany 1994, pp. 521-524.

[7] L.R. Rabiner and R.W. Schafer, "Digital Processing of Speech Signals," Prentice-Hall, Englewood Cliffs, N.J. 1978.