IMPLEMENTATION OF A LOW-COST ACOUSTIC ECHO CANCELLER

Thomas Schertler *Gerhard Uwe Schmidt*

Fachgebiet Theorie der Signale, Darmstadt University of Technology, Germany

schert@nesi.e-technik.th-darmstadt.de schmidt@nesi.e-technik.th-darmstadt.de

ABSTRACT

Hands-free telephone sets with adaptive echo cancellers need an adaptive loss control in order to guarantee a loss of acoustic echo of about 45 dB (ITU-T Recommendation G.167 [1]). The quality of the speech transmission depends on the coupling between the echo canceller and the loss control. In this contribution we present an implementation of a hands-free telephone algorithm, using a low-cost fixed-point DSP which is well suited for integration in consumer devices.

1. INTRODUCTION

It is a well known fact that in realistic environments one cannot achieve, even with the best available DSP's, an echo loss of more than 20–25 dB with an adaptive echo canceller. This is due to changes in the echo path, disturbing background noise, limited number of filter taps etc.

In order to provide a constant total echo loss of about 45 dB, one can use an adaptive loss control. An adaptive loss control estimates whether the local speaker or the far-end speaker is talking and attenuates the channel that is not in use at the moment. A problem arises when both speakers are active (*double talk*). In this case, the speech quality decreases because the loss control keeps on switching and thus tears the speech to pieces. These artefacts often diminish the acceptance of the hands-free mode of the telephone.

With a combination of an echo canceller and a loss control (see figure 1), the attenuation of the loss control can be lowered by the amount of echo loss achieved by the echo canceller, so that the artefacts are less disturbing. The following algorithms are implemented on a 16 bit fixed-point DSP. Due to external requirements, the computational load of the DSP had to be restricted to 20 MIPS for the echo cancelling part.

2. LOSS CONTROL AND ECHO CANCELLER

The loss control used in our implementation is the ARCOFI–algorithm, which has been implemented [2] in DSP code to enable the combination with other DSP algorithms. The loss control also incorporates an automatic gain control (AGC). It has been used for years in commercially available comfort telephones and has proven its reliability. However, due to the required echo loss of 45 dB, the communication during double talk is somewhat uncomfortable.

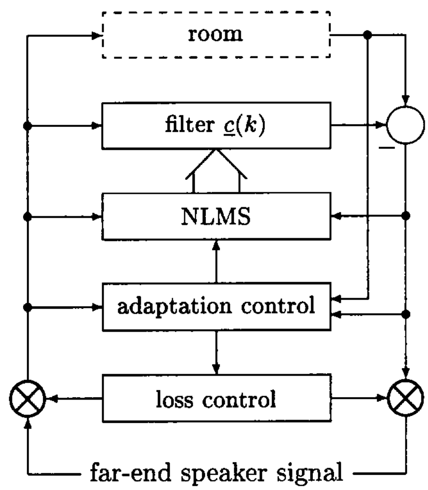

Figure 1: block diagram

As the basis of the acoustic echo canceller, we implemented the well-known NLMS algorithm. The algorithm tries to estimate the acoustic echo by minimizing the mean square adaptation error $e_f(k)$ (see equ. 1). It is well suited for real-time implementations because of its simplicity. The computational load in terms of multiplications is $2N$, where N is the number of filter taps. There are some constraints of the NLMS algorithm, which will be discussed later in this contribution, but it seems to be the algorithm used most often for acoustic echo cancellers.

The NLMS algorithm converges best when the far-end signal is white noise. Therefore, one can try to increase the adaptation performance of the echo canceller by filtering the speech signals with prewhitening filters. The best way to do that would be to estimate the autocorrelation function of the excitation and to compute the Levinson–Durbin–coefficients in periodic intervals. Using these periodically changing coefficients, one could implement an adaptive prewhitening filter which diminishes the correlation between successive values of the excitation [3]. The more coefficients are used, the better is the whitening effect. Due to the time-variance of the prewhitening filter, its inverse cannot be calculated so that the echo cancellation filter has to be carried out twice: one convolution of the estimated room impulse vector with the prewhitened excitation vector and one convolution with the original excitation vector. This leads to an increased need of resources which is $3N$ in terms of multiplications and $3N$ in terms of memory.

In our implementation, we do *not* use adaptive prewhitening decorrelation filters but a time-invariant filter of order one. Thus, we only have to calculate the convolution of the filter vector and the signal vector once instead of twice. Of course, in comparison to the prewhitening filter with adaptive filter coefficients, the performance decreases, but the computational load caused by the fixed prewhitening filter is minimal.

Let $\underline{c}(k)$, $\underline{x}(k)$ and $\underline{x}_f(k)$ be the vectors of the filter coefficients, the far-end speaker signal and the prewhitened signal of the far-end speaker. The index "f" always denotes prewhitened signals. With this, we can annotate the filtering part and the update of the filter coefficient vector $\underline{c}(k)$ by the following equations:

$$\hat{y}_f(k) = \underline{c}^T(k)\,\underline{x}_f(k)\,, \quad e_f(k) = y_f(k) - \hat{y}_f(k)\,, \quad (1)$$

$$\underline{c}(k+1) = \underline{c}(k) + \mu(k)\,\frac{e_f(k)}{\underline{x}_f^T(k)\,\underline{x}_f(k)}\,\underline{x}_f(k)\,, \quad (2)$$

with

$$\underline{c}(k) = (c_0(k), c_1(k), \cdots c_{N-1}(k))^T,$$
$$\underline{x}(k) = (x(k),\, x(k-1),\, \cdots,\, x(k-N+1))^T,$$
$$\underline{x}_f(k) = (x_f(k),\, x_f(k-1),\, \cdots,\, x_f(k-N+1))^T,$$

N: Number of filter taps, $\mu(k)$: stepsize.

In order to achieve convergence with white noise as excitation (which is the best case), $\mu(k)$ has to be chosen between 0 and 2. As shown in [4], best convergence is obtained by choosing the stepsize factor $\mu(k)$ according to the following equation:

$$\mu(k) = \frac{\mathrm{E}\{\varepsilon_f^2(k)\}}{\mathrm{E}\{e_f^2(k)\}}\,, \quad (3)$$

with $\varepsilon_f(k) = e_f(k) - n_f(k)$, $n_f(k)$: local signal.

Unfortunately, the undisturbed error power $\mathrm{E}\{\varepsilon_f^2(k)\}$ is not known. There are several possibilities to estimate this value [3, 5], but with respect to the limited resources in our implementation we decided to use a simpler control mechanism, using only two different stepsizes. One stepsize μ_{high}, which is reasonably large (between 1/8 and one), and another stepsize μ_{low}, which is zero or close to zero. Normally, the large stepsize is used for the adaption. Without excitation or during double talk the filter taps may diverge when this large stepsize is used. Therefore, we need a detection algorithm in order to switch the stepsize $\mu(k)$ to its lower value in these situations.

3. STEPSIZE CONTROL

For the detection of sufficient excitation we calculate the short-term average magnitude $\overline{x}_{f_s}(k)$ of the far-end speaker signal:

$$\overline{x}_{f_s}(k) = (1-\alpha_{sam})\,|x_f(k)| + \alpha_{sam}\,\overline{x}_{f_s}(k-1) \quad (4)$$

$$\text{with } \alpha_{sam} = \begin{cases} \alpha_r & \text{if } |x_f(k)| > \overline{x}_{f_s}(k-1) \\ \alpha_f & \text{if } |x_f(k)| \le \overline{x}_{f_s}(k-1)\,. \end{cases}$$

$$\text{and } 0 \ll \alpha_r < \alpha_f < 1\,.$$

This value $\overline{x}_{f_s}(k)$ is then compared to an excitation threshold $x_0(k)$ which adaptively tracks the far-end background noise level [6]. If $\overline{x}_{f_s}(k)$ is lower than $x_0(k)$, we assume that there is no far-end speaker signal and set $\mu(k)$ to its lower value.

To determine double talk, a coupling factor $l^*(k)$ describing the power coupling between $x_f(k)$ and $e_f(k)$ is calculated, which incorporates the attenuation of the echo canceller and the attenuation of the room. This is done by dividing $\overline{x}_{f_s}(k)$ by the short-term average magnitude $\overline{e}_{f_s}(k)$ of the adaptation error signal $e_f(k)$:

$$l^*(k) = \frac{\overline{x}_{f_s}(k)}{\overline{e}_{f_s}(k)}\,.$$

In order to avoid a poor estimation of the coupling factor, this calculation has to be done only during single talk of the far-end speaker. For this reason, we calculate a correlation factor $\rho(k,l)$ (described in [6, 7]), whose maximum value $\max_l \rho(k,l)$ has to exceed a limit ρ_0 before we assume single talk.

With the mentioned correlation factor, single talk can be detected with a sufficient low error rate, but it is not well suited to control the stepsize in a more direct way, because of its inert behaviour. Additionally, in the case of single talk, it does not reliably detect single talk, which leads to slow convergence. Actually, the coupling factor is smoothed by a nonlinear filter:

$$l(k) = \begin{cases} (1-\alpha_l)\dfrac{\overline{x}_{f_s}(k)}{\overline{e}_{f_s}(k)} + \alpha_l\, l(k-1) \\ \qquad \text{if } \max_l \rho(k,l) > \rho_0 \\ l(k-1) \qquad \text{else}\,. \end{cases} \quad (5)$$

Double talk is assumed when the short-term average magnitude of the echo is larger than the estimated echo, which is calculated as the quotient of $\overline{x}_{f_s}(k)$ and the coupling factor $l(k)$ times an additional factor p_α. In this case, the stepsize $\mu(k)$ is set to its lower value:

$$\mu(k) = \begin{cases} \mu_{high} & \text{if } \overline{e}_{f_s}(k) \le \dfrac{\overline{x}_{f_s}(k)}{p_\alpha \cdot l(k)} \\ \mu_{low} & \text{if not.} \end{cases} \quad (6)$$

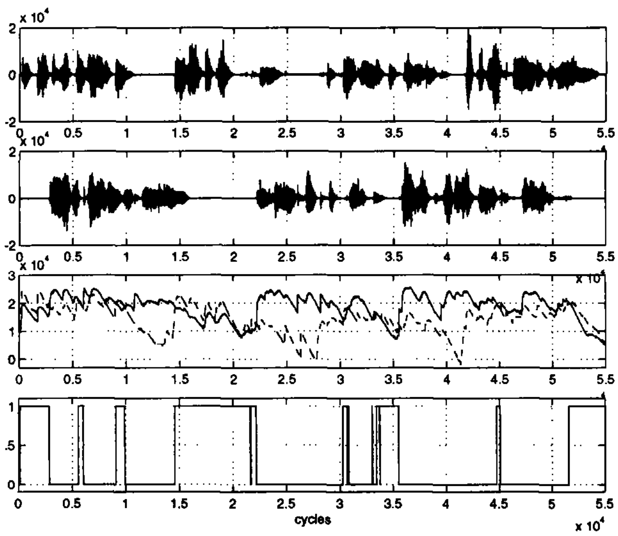

Figure 2: Subfigure 1: far-end speaker signal. Subfigure 2: local speaker signal. Subfigure 3: $\overline{e}_{f_s}(k)$ (solid line) and estimated error (dashed line), logarithmic scaling. Subfigure 4: resulting adaptation parameter $\mu(k)$.

Choosing p_α greater than one diminishes the probability of adaptation during double talk but can also decrease the adaptation speed during single talk.

Figure 2 shows an example of a double talk situation. If there is no excitation (subfig. 1), the stepsize parameter $\mu(k)$ is set to zero. Most of the time during double talk, the solid line ($\overline{e}_{f_s}(k)$) of subfig. 3 exceeds the dashed line (estimated echo). Consequently, the stepsize parameter $\mu(k)$ is set to zero too. There are a few moments during double talk, when $\mu(k)$ is set to one. In these moments, the far-end speaker signal is louder then the local signal.

The reliability of the described double talk detection algorithm increases when the system mismatch between the room impulse response and the filter taps decreases. After changes in the echo path, the detector possibly stops adaptation during single talk, since the coupling factor $l(k)$ does not reflect the changes in the system mismatch. Due to the recalculation of the coupling factor, using the correlation factor, $l(k)$ will be updated within ca. 100 ms [6] and the adaptation will be started again.

4. COUPLING OF ECHO CANCELLER AND LOSS CONTROL

The coupling factor, as described above, should not be used directly to diminish the attenuation of the adaptive loss control. In order to avoid an overestimation of the coupling factor, we proceed by smoothing the coupling factor by means of an additional nonlinear filter. The smoothed factor $l_{lc}(k)$ can be determined as follows:

$$l_{lc}(k) = \begin{cases} (1-\beta_r)\, l(k) + \beta_r\, l_{lc}(k-1) \\ \quad \text{if } l(k) > l_{lc}(k-1) \text{ and } \max_l \rho(k,l) > \rho_0 \\ (1-\beta_f)\, l(k) + \beta_f\, l_{lc}(k-1) \\ \quad \text{if not.} \end{cases} \tag{7}$$

By choosing $0 \ll \beta_f < \beta_r < 1$, we cautiously follow rising values of $l(k)$. If the values of $l(k)$ are falling, we follow these values without delay (see figure 3). By means of this "worst case" assumption, we are able to achieve an echo loss of 45 dB in accordance with the ITU-T Recommendation G.167, (section 5.4.1).

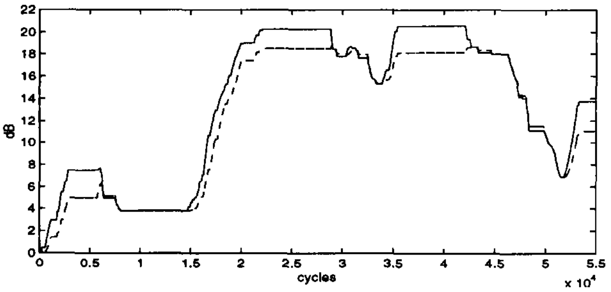

Figure 3: solid line: coupling factor $l(k)$, dashed line: coupling factor $l_{lc}(k)$

In section 5.4.2, the ITU-T Recommendation G.167 describes the weighted terminal coupling loss during double talk. In this situation, the requested attenuation of 45 dB can be lowered to 30 dB. Using this recommendation can lead to a less disturbing behaviour of the loss control.

If we want to make use of this recommendation, there should be no doubt of being in the state of double talk when diminishing the attenuation. Otherwise, the echo loss during single talk does not reach 45 dB as recommended.

On the other hand, in order to preserve the echo loss of the adaptive filter, the double talk detection algorithm must be able to stop the adaptation very quickly. Therefore, the double talk detector mentioned above has to stop adaptation at the first indication of double talk.

To resolve this discrepancy, we need to modify the double talk detector for use with the adaptive loss control, which is done by two criteria and some logic elements.

As the first criterion we use a modification of equ. 6 by adding a second factor p_β, which simply increases the distance between the estimated error and $\overline{e}_{f_s}(k)$:

$$\overline{e}_{f_s}(k) > \frac{p_\beta\, \overline{x}_{f_s}(k)}{p_\alpha\, l(k)} \tag{8}$$

With $p_\beta > 1$, the local speaker signal has to be stronger than in equ. 6, before double talk is assumed.

The second criterion is only checked when $l_{lc}(k)$ exceeds a specified limit l_{lim}, e.g. 6 dB. In this case we examine wether the adaptive filter has reduced the value of $\bar{e}_{f_s}(k)$ by a factor p_{dt} in comparison to the short-term average magnitude $\bar{y}_{f_s}(k)$ of $y_f(k)$. If $y_f(k)$ contains no local speaker signal, this would be possible. Otherwise, the adaptive filter is not able to do such a reduction. This leads to the formulation of the second criterion:

$$l_{lc}(k) > l_{min} \text{ and } \frac{|\bar{y}_{f_s}(k) - \bar{e}_{f_s}(k)|}{\bar{y}_{f_s}(k)} < p_{dt} \qquad (9)$$

In addition before lowering the total attenuation from 45 to 30 dB criterion 1 and criterion 2 have to hold for several consecutive sampling instances. On the other hand, after double talk is ended, the attenuation has to to re-increased quickly in order to satisfy ITU-T Recommendations. This is the reason why we use a counter $c_{dt}(k)$, which is determined according to the following equation:

$$c_{dt}(k) = \begin{cases} c_{dt}(k-1)+1 & \text{if crit. 1 and 2 hold} \\ c_{dt}(k-1)/2 & \text{else.} \end{cases} \qquad (10)$$

The resulting coupling factor $l_{dt}(k)$ which is actually passed to the loss control is calculated as:

$$l_{dt}(k) = l_{lc}(k) + l_{add}(k) \qquad (11)$$

$$\text{with } l_{add}(k) = \begin{cases} \beta_a\, l_{add}(k-1) + (1-\beta_a)\cdot 15\,\text{dB} \\ \qquad \text{if } c_{dt}(k) > c_{lim} \\ \beta_b\, l_{add}(k) \qquad \text{else} \end{cases} \qquad (12)$$

$$\text{and } 0 \ll \beta_a < \beta_b < 1\,.$$

Figure 4 shows the behaviour of $c_{dt}(k)$ (subfig. 3) and the coupling factors $l_{lc}(k)$ (subfig. 4, solid line) and $l_{dt}(k)$ (subfig. 4, dashed line). One can see that during double talk, the attenuation of the loss control will be decreased. Up to cycle 20,000, the criterion 2 is not met because the adaptive filter has to converge and the coupling factor $l_{lc}(k)$ was lower than the limit.

5. CONCLUSIONS

Adaptive loss controlls seem to be indispensable for hands-free telephone sets when the ITU-T Recommendation has to be met, but it is possible to lower their artefacts using adaptive filter algorithms. The presented implementation follows the ITU-T Recommendation and enhances the double talk behaviour of a hands-free telephone set drastically. All described parts of the presented algorithm are implemented on a low-cost fixed-point DSP which is ready to use in commercial products.

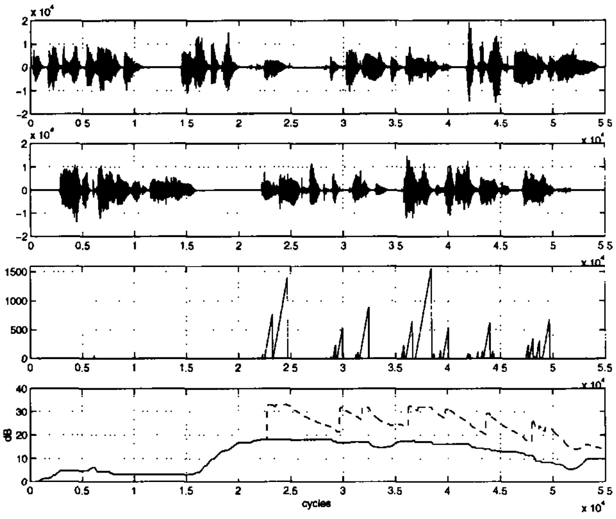

Figure 4: Subfigure 1: far-end speaker signal. Subfigure 2: local speaker signal. Subfigure 3: counter $c_{dt}(k)$. Subfigure 4: $l_{lc}(k)$ (solid line), $l_{dt}(k)$ (dashed line), logarithmic scaling.

REFERENCES

[1] "Rec. G.167: Acoustic echo controllers." ITU-T, WTSC, Helsinki, Mar. 1993. http://www.itu.ch.

[2] A. Hauenstein, "Implementierung eines Freisprechalgorithmus auf dem Siemens DSP–Core SPC," Diplomarbeit D103, FG Theorie der Signale, TH Darmstadt, Nov. 1995.

[3] T. Burger, "Practical applications of adaptation control for NLMS-algorithms used for echo cancellation with speech signals," in *Proc. 4th Int. Workshop on Acoustic Echo and Noise Control*, (Røros, Norway), pp. 87–91, June 1995.

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

[5] P. Heitkämper, "An adaptation control for acoustic echo cancellers," *IEEE Signal Processing Letters*, vol. 4, pp. 170–172, June 1997.

[6] P. Heitkämper, *Freisprechen mit Verstärkungssteuerung und Echokompensation.* Fortschritt-Berichte, Reihe 10, Nr. 380, Germany: VDI, 1995.

[7] P. Heitkämper, "Optimization of an acoustic echo canceller combined with adaptive gain control," in *Proc. ICASSP-95*, (Detroit, Michigan), pp. 3047–3050, 1995.