Katholieke Universiteit Leuven

Departement Elektrotechniek
ESAT-SISTA/TR 1998-38

Real-Time Implementation of an Acoustic Echo Canceller [1]

Koen Eneman, Marc Moonen [2]

May 1998
Published in the Proceedings of the COST#254 Workshop on Intelligent Communications,
L'Aquila, Italy, June 4-6, 1998

[1]This report is available by anonymous ftp from *ftp.esat.kuleuven.ac.be* in the directory *pub/SISTA/eneman/reports/98-38.ps.gz*

[2]ESAT (SISTA) - Katholieke Universiteit Leuven, Kardinaal Mercierlaan 94, 3001 Leuven (Heverlee), Belgium, Tel. 32/16/321809, Fax 32/16/321970, WWW: *http://www.esat.kuleuven.ac.be/sista*. E-mail: *koen.eneman@esat.kuleuven.ac.be*.
Marc Moonen is a Research Associate with the F.W.O. Vlaanderen (Flemish Fund for Science and Research). This research was carried out at the ESAT laboratory of the Katholieke Universiteit Leuven and was partly funded by the Concerted Research Action MIPS (Model-based Information Processing Systems), F.W.O. project nr. G.0295.97 of the Flemish Government, the Interuniversity Attraction Pole
(IUAP-nr.02) initiated by the Belgian State, Prime Minister's Office for Science, Technology and Culture.

Real-Time Implementation of an Acoustic Echo Canceller

Koen Eneman *Marc Moonen*

ESAT - Katholieke Universiteit Leuven
Kardinaal Mercierlaan 94, 3001 Heverlee - Belgium
koen.eneman@esat.kuleuven.ac.be *marc.moonen@esat.kuleuven.ac.be*

Abstract

Acoustic echo cancellation is an essential signal enhancement tool in hands-free communication. Loudspeaker signals are picked up by a microphone and are fed back to the correspondent, resulting in an undesired echo. Nowadays, adaptive filtering techniques are typically employed to suppress this echo.

In acoustic applications long filters need to be adapted for sufficient echo suppression. Classical adaptation schemes such as LMS are quite expensive for accurate echo path modelling in highly reverberating environments. Cheaper algorithms were proposed and are mainly based on subband and frequency-domain techniques. However, due to nonlinearities and the time-dependence of the echo path some residual echo will always remain.

Apart from the adaptive filter also a steering algorithm has to be included to remove the residuals and to ensure proper operation during double-talk.

Different adaptive algorithms have been implemented on DSP. They are compared based on a cost/performance analysis. A steering algorithm is used that can withstand the non-stationarities of the acoustic environment.

1 Introduction

Hands-free teleconferencing systems such as hands-free telephones (in cars), tele-classing and video-conferencing systems provide a comfortable way of communicating. However, signal deterioration occurs when loudspeaker signals are picked up by a microphone and sent back to the correspondent. This results in an undesired echo as shown in figure 1.

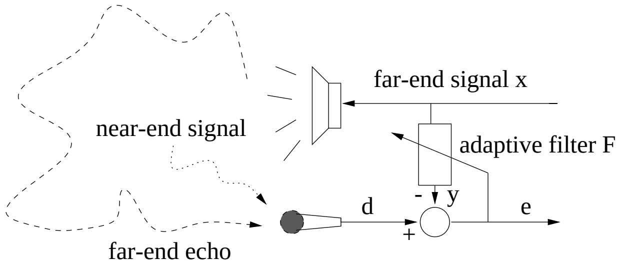

Figure 1 : Echo cancellation setup

Conventional techniques used in classical telephony such as clipping and voice controlled switching [1] only have a limited performance. More advanced techniques using powerful digital signal processing equipment are expected to provide a better signal quality.

2 Adaptive Filtering Techniques

2.1 Least Mean Squares

Nowadays, acoustic echoes are typically suppressed by means of adaptive filtering techniques [2]. An adaptive filter iteratively converges to an estimate of the impulse response of the acoustic path (see figure 1). Of all existing adaptive algorithms the Least Mean Squares algorithm may be best known. An FIR filter F is updated iteratively according to

$$F_{new} = F_{old} + \mu(d - F_{old}^T \mathbf{x})\mathbf{x}.$$

LMS-based algorithms have a complexity that is linear in the filter length, but they suffer from a rather slow convergence for 'coloured' signals such as speech. In order to cope with dynamic signals, stepsize μ is often normalised by taking it inversely proportional to the energy of **x**. This normalised version of LMS (NLMS) is typically used in practice.

As acoustic echo cancellers have to operate in real-time, they should fit on a (single) DSP processor with limited computational capacity and memory. In acoustic applications long filters need to be adapted for sufficient echo suppression. Classical adaptation schemes such as LMS are quite expensive for accurate echo path modelling in highly reverberating environments. More efficient structures have been proposed over the last 15 years. They are mainly based on subband or frequency-domain techniques.

2.2 Subband Adaptive Filtering

A general setup for subband acoustic echo cancellation is shown in figure 2.

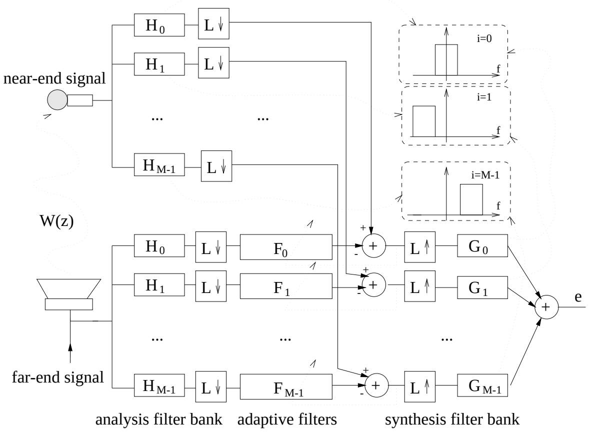

Figure 2 : Subband adaptive echo canceller

The input signals are first processed by identical *M*-band analysis filter banks and then downsampled with a factor *L*. The far-end subband signals are passed through a set of adaptive filters F_i. The subband error signals are then finally recombined in the synthesis filter bank. The ideal frequency amplitude characteristics of the analysis bank filters H_i and synthesis bank filters G_i are shown (ideal bandpass filters). Due to aliasing effects, this setup will only work for $M \geq L$.

2.2.1 Critically Downsampled Subband Schemes

If *L* is chosen equal to *M* a critically downsampled subband adaptive filter is being implemented. It seems attractive because optimal computational savings can be made as *L* is as high as possible. In [3] it is shown that critically downsampled subband systems lead to a residual modelling error which is considerable unless cross filters are included between neighbouring subbands. Cross filters again increase the complexity. Furthermore, cross filters fail to converge quickly. This suggests the use of oversampled subband schemes for which $M>L$.

2.2.2 Oversampled Subband Schemes

Splitting signals into subbands seems very promising, since for coloured input spectra the convergence of fullband LMS is slow. Here, each downsampled subband signal will have a flatter spectrum, leading to improved convergence if an LMS updating algorithm is used to adapt the subband weights. As all computations can be done at the lower sampling rate, this subband approach is supposed to give a better performance at a lower cost.

In practice a considerable residual error remains. It appears that an extra delay has to be inserted in the near-end signal path and that the subband filters need to be larger than expected in order to remove the error [4][5]. The effective computational gain w.r.t. LMS is therefore smaller than expected.

2.3 Frequency-Domain Adaptive Filters

2.3.1 FDAF

By applying block processing techniques, implementation cost can be exchanged for extra delay. BLMS is a block version of LMS. When it is translated in frequency domain it leads to the frequency-domain adaptive filter (FDAF) [6]. The FDAF is only computationally attractive if the block length equals the filter length approximately. In practice this leads to unacceptable input/output delays.

2.3.2 *PBFDAF*

By partitioning the adaptive filter a canceller with acceptable delay and low implementation cost can be obtained. It was called the Partitioned Block Frequency-Domain Adaptive Filter (PBFDAF) [7][8]. The N-taps fullband adaptive filter $w(k)$ is partitioned in N/P equal parts $w_p(k)$: [1]

$$w_p(k) \overset{p=0\to\frac{N}{P}-1}{=} \begin{cases} w(k) & k = pP \to (p+1)P-1 \\ 0 & \text{elsewhere} \end{cases}$$

The equations for the PBFDAF are : [2]

$$\mathbf{X}_{n-p} \overset{\forall p}{=} \operatorname{diag}\left\{ \mathbf{F} \begin{bmatrix} x((n+1)L-pP-M+1) \\ \vdots \\ x((n+1)L-pP) \end{bmatrix} \right\}$$

$$\mathbf{y} = \begin{bmatrix} \mathbf{0}_{P-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_L \end{bmatrix} \mathbf{F}^{-1} \sum_{p=0}^{\frac{N}{P}-1} \mathbf{X}_{n-p} \mathbf{W}_{p_n}$$

$$\mathbf{d} = \begin{bmatrix} \mathbf{0} \\ \mathbf{d}_n \end{bmatrix}, \quad \mathbf{d}_n = \begin{bmatrix} d(nL+1) \\ \vdots \\ d((n+1)L) \end{bmatrix}$$

$$\mathbf{e} = \mathbf{d} - \mathbf{y}$$

$$\mathbf{W}_{p_{n+1}} \overset{\forall p}{=} \mathbf{W}_{p_n} + \mathbf{F} \begin{bmatrix} \mathbf{I}_P & \mathbf{0} \\ \mathbf{0} & \mathbf{0}_{L-1} \end{bmatrix} \mathbf{F}^{-1} \Delta \mathbf{X}_{n-p}^H \mathbf{F} \mathbf{e}$$

The block length is L, the corresponding input/output delay equals $2L$-1. $\mathbf{F}$ is an $M \times M$ DFT matrix, Δ=2diag(μ_n) and $M=P+L-1$.[3] Ideally, the first of the equations above requires only 1 DFT operation, which corresponds to p=0. $\mathbf{X}_{n-p}$ for p>0 can be recovered from previous iterations if P is divisible by L.

It was shown in [9] that the PBFDAF scheme can be put into the oversampled subband framework. The PBFDAF implements a simple DFT modulated perfect reconstruction filter bank with filters having sinc-like frequency characteristics.

There exists two variants, called the constrained and the unconstrained PBFDAF. For the unconstrained version

$$\mathbf{F} \begin{bmatrix} \mathbf{I}_P & \mathbf{0} \\ \mathbf{0} & \mathbf{0}_{L-1} \end{bmatrix} \mathbf{F}^{-1}$$

is left out from the last equation. The unconstrained updating requires 3 FFTs whereas the constrained PBFDAF is more expensive, having an extra $2N/P$ FFTs to compute. The latter on the other hand has better convergence properties.

Stepping several times through the 5 equations that define the PBFDAF with n kept constant leads to an approved weight update. This algorithm, which of course enhances the convergence behaviour, will be called the PBFRAP[4].

Introducing stepsize normalisation is another way of improving convergence. As the PBFDAF takes on the form of an oversampled subband adaptive filter more or less, applying different stepsizes for each subband, dependent on the subband energy, improves the convergence.

In practical design, block length L is constrained by the maximal tolerable delay. For a sampling frequency of 8 kHz and a maximal delay of 16 ms L is constrained to be smaller than or equal to 64. A value for P that minimises the implementation cost is then preferred. In general, one can state that $L=P=M/2$ is a good choice. Only when N becomes large or the maximal tolerable delay is small, P/L>1 can be put forward.

[1] We assume that N/P is integer.

[2] For signal conventions : see figure 1.

[3] $P+L$-1 is in fact a lower bound for M, so also $M>P+L$-1 will work.

[4] RAP stands for Row Action Projection [10].

3 Robust Operation and Control

Until now the design of adaptive filtering schemes was discussed. An echo canceller however, operates in a time-varying environment and has to cancel highly non-stationary signals such as speech. A robust system is then required. Some extra control parameters have to be included, which are basically used to steer the adaptation speed. Intensive testing and tuning should eventually lead to a cheap control system which is as robust as possible. A more elaborated scheme replacing figure 1 is shown in the figure below.[5]

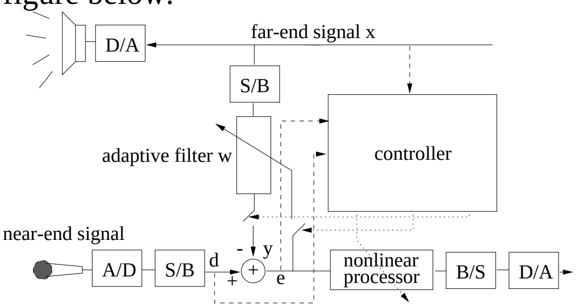

Figure 3 : Acoustic echo canceller

For control the energy of the **x**-, **d**- and **e**-buffer can be tracked.

Once the energies are computed, some control decisions can be made. For instance, if the far-end signal energy is lower than a certain threshold ε, no far-end stimulus is supposed to be present. The adaptation process is frozen and the near-end signal **d** is passed to the output without correction (**e**=**d**). By continuing the adaptation the filter(s) could drift away from the acoustic path replica due to the activity of background noise at the near-end side.

Also when a local near-end speaker is active or in double-talk situations, i.e. when both speakers are active, the adaptation process must be frozen. Otherwise, the adaptive filter is again driven away from its Wiener solution by the local non-stationary source. The adaptive coefficients would be whirling around on the rhythm of the local source resulting in an annoying echo-like disturbance. Near-end speech detection is thus crucial for correct operation. In case of double-talk, this is far from easy as it comes down to a detection of speech in speech. The onsets of speech are often difficult to detect and to discriminate from a non-stationary part of the far-end signal. A double-talk detector which is too sensitive will generate a lot of false alarms. The adaptation is regularly stopped, so the overall convergence speed will be low. On the other hand, when the detector is critically tuned, even a slightly too late detection of the onset of near-end speech could lead to a significant misfit of the adaptive filter.

The echo path is supposed to attenuate the far-end signal level. Therefore, a comparison between far-end and near-end instantaneous energy gives an idea about near-end source activity. If $E_d > \tau E_x$ double talk is detected[6]. Fine tuning threshold τ is crucial however. Another measure could be [8]

$$\rho = \frac{E_x E_e}{E_x^2 + E_y^2}.$$

It is smaller than 1 in absence of double talk. When a local speaker starts to speak, ρ will rise.

The problem so far is that the adaptation is switched either off or on. A sliding stepsize μ may be more appropriate. μ can vary between 0 (near-end activity) and μ_{max} (only far-end activity) based on the probability that the near-end source is active. In [11] a correlation based method was proposed. In the absence of near-end speech the loudspeaker and microphone signal are highly correlated. An estimate

[5] The A/D and D/A units are analog-to-digital and digital-to-analog converters respectively. S/B and B/S stand for serial-to-block and block-to-serial conversion.

[6] E_d is the near-end frame energy. E_x is the energy corresponding to the far-end frame, $\tau<1$.

of the attenuation $\alpha=E_e/E_x$ can be updated now. If the short-time energy at the output of the adaptive filter E_e is significantly larger than expected ($E_e>\alpha E_x$) adaptation must be stopped. By comparing the short-time and long-time energies of both **x** and **e**, the activity at the far-end side as well as the level of near-end background noise can be estimated.

4 Real-Time Implementation on DSP

A real-time echo canceller was programmed on DSP as a demo for adaptive echo cancellation and hands-free communication. The canceller basically consists of an adaptive filtering core and some surrounding control software (fig. 3). Several adaptive filters can be plugged in for evaluation and comparison.

4.1 DSP equipment

Two DSP boards are placed in a VME-rack. They are accessible through our local network via a Sun Sparc station. For this application two DSPs are used. A 25-MIPS TMS320C44[7], clocked at 50 MHz is responsible for the data acquisition. The loudspeaker and microphone channel are first sampled at 16 kHz and then digitally downsampled to 8 kHz to avoid aliasing distortion. The input channels **x** and **d** are sent to a second DSP, a 25-MIPS TMS320C40 @ 50 MHz, which does the echo cancellation. The output samples **e** are transferred back to the first DSP and after digital upsampling, they are sent to a loudspeaker for evaluation.

4.2 Software

The algorithms were first tested in matlab en C and then ported to DSP. The control algorithm is based on [11] and was mainly programmed in C. Some of its features were already described in a previous paragraph.

Different adaptive algorithms were implemented, mainly programmed in assembly. In this way some specific DSP operations such as circular addressing and parallel instructions are optimally used. At this moment NLMS, (un)constrained PBFDAF and PBFRAP are available on DSP. The longest filter that could be adapted in real-time using an unconstrained PBFDAF with $L=P=M/2=64$ was 325 ms. For NLMS this reduces to 100 ms : the on-chip memory of the C4x is very fast, but rather small and limits the filter length.

4.3 Experiments

Some tests were carried out in the ESAT speech laboratory, which has a recording room with variable damping and the necessary equipment to set up an experiment. Referring to figure 3, a loudspeaker (far-end signal **x**) and a microphone (near-end signal **d**) were placed 40 cm apart. The near-end speaker was replaced by another loudspeaker, fed by a CD-player, to avoid unwanted time-variations in the echo path by speaker's motion. The impulse response of the room was determined. The room was found to be moderately damped.

In a first experiment band filtered[8] white noise was put through the far-end loudspeaker, the near-end speaker remained silent. The acoustic path was estimated with an FIR filter of 768 taps (96 ms) using 4 adaptive algorithms ($L=P=M/2=64$, $f_s=8$ kHz) : unconstrained PBFDAF, constrained PBFDAF, NLMS (block length=64) and unconstrained PBFRAP (2 iterations).

After a fast initial convergence a residual error remains. This is mainly because the

[7] The TMS320C4x-family are standard floating-point DSPs from Texas Instruments, suitable for audio processing.

[8] The passband was chosen to be [200,3700] Hz.

infinite path length is modelled with a finite length filter. The PBFRAP algorithm apparently has the best convergence properties. Nevertheless, in practice the echo suppression will not come below approximately 30 dB, because of nonlinear distortion (loud-speaker), the non-stationarity of the acoustic path and wrong control decisions. Identifying long acoustic paths is therefore not advised. It will slow down convergence, lowering the error level just a little bit. By identifying only the dominant part of the acoustic path -100 ms e.g., as was done in this experiment- sufficient echo suppression is obtained and more expensive adaptive algorithms such as NLMS may be reconsidered for acoustic echo cancellation.

The echo canceller was further validated through listening tests in the speech laboratory. In this case both the far-end and near-end source are active. Several tests were done under different conditions: different signals ((coloured) noise, speech, music, ...), sampling frequencies and source volume. Also the effect of a time-varying echo path was verified. The steering algorithm came up to our expectations.

5 Conclusions

A real-time acoustic echo canceller was implemented on DSP. Different adaptive algorithms were compared. Acoustic channels up to 325 ms could be modelled with an unconstrained PBFDAF. In practice there will always be a residual error. Therefore, by modelling only the dominant part of the echo path sufficient echo suppression can be obtained. More expensive adaptive filters such as NLMS may be reconsidered then.

Acknowledgements

Marc Moonen is a Research Associate with the F.W.O. Vlaanderen (Flemish Fund for Science and Research). This research was carried out at the ESAT laboratory of the Katholieke Universiteit Leuven and was partly funded by the Concerted Research Action MIPS (Model-based Information Processing Systems), F.W.O. project nr. G.0295.97 of the Flemish Government, the Interuniversity Attraction Pole (IUAP-nr.02) initiated by the Belgian State, Prime Minister's Office for Science, Technology and Culture. The scientific responsibility is assumed by its authors.

References

[1] E. Hänsler, “The hands-free telephone problem - An annotated bibliography,” Signal Processing, vol. 27, pp. 259-271, June 1992.

[2] S. Haykin, “Adaptive Filter Theory, 3th ed.,” Englewood Cliffs, New Jersey, USA: Prentice Hall, 1996.

[3] A. Gilloire and M.Vetterli, “Adaptive Filtering in Subbands with Critical Sampling: Analysis, Experiments and Application to Acoustic EchoCancellation,” IEEE Trans. Signal Processing, vol.40, pp. 1862-1875, August 1992.

[4] W. Kellermann, “Analysis and Design of Multirate Systems for Cancellation of Acoustical Echoes,” in Proceedings of the 1988 IEEE Int. Conf. on Acoust., Speech and Signal Processing, (New York, NY, USA), pp. 2570-2573, April 1988.

[5] K. Eneman, M. Moonen and I. Proudler, “Frequency-domain adaptive echo cancellation as a special case of subband echo suppression,” in Proceedings of the ProRISC/IEEE Benelux workshop on Circuits, Systems and Signal Processing, Mierlo, The Netherlands, pp.131-136, November 1996.

[6] J. Shynk, “Frequency-Domain and Multirate Adaptive Filtering,” IEEE Signal Processing Magazine, pp.15-37, January 1992.

[7] J.-S. Soo and K. Pang, “Multidelay Block Frequency Domain Adaptive Filter,” IEEE Trans. Acoust., Speech and Signal Processing, vol.38, pp. 373-376, February 1990.

[8] J. Páez Borrallo and M. García Otero, “On the implementation of a partitioned block frequency domain adaptive filter (PBFDAF) for long acoustic echo cancellation,” Signal Processing, vol.27, pp.301-315, June 1992

[9] K. Eneman and M. Moonen, “A Relation Between Subband and Frequency-Domain Adaptive Filtering,” in Proceedings of the 13th Int. Conf. on Digital Signal Processing, (Thira, Santorini, Greece), July 2-4 1997.

[10] S. Gay, “Fast Projection Algorithms with Application to Voice Echo Cancellation”. PhD thesis, Rutgers, The State University of New Jersey, New Brunswick, New Jersey, USA, October 1994.

[11] P. Heitkämper, “An Adaptation Control for Acoustic Echo Cancellers,” IEEE Signal Processing Letters, vol.4, pp. 170-172, June 1997.