A MULTI-FILTER APPROACH TO ACOUSTIC ECHO CANCELLATION FOR TELECONFERENCING

A. John Usher *

McGill University
Multichannel Audio Research Laboratory
Department of sound recording
Montréal, Canada

B. Jeremy R. Cooperstock , C. Wieslaw Woszczyk

McGill University
Centre for Interdisciplinary Research
in Music Media and Technology
Montréal, Canada

ABSTRACT

We have designed and tested an acoustic echo cancellation system for speech teleconferencing. Our algorithm is based on a least-mean-square (LMS) frequency domain adaptive filter (FDAF) and uses a novel filter-update technique using many (at least 3) simultaneously running filters. We find the new multi-filter to converge faster than similar LMS FDAF's for echo cancellation, and find it to be especially robust during double-talk conditions.

1. INTRODUCTION

The increase in data band-width for telecommunications today has created a need for high quality audio teleconferencing. Echo-cancellers are a common feature to teleconferencing systems which use "hand-free" operating systems, whereby the users at each end of the conference can freely interact with each other. As described in figure 1, the purpose of an acoustic echo-canceller for these applications is to reduce the amount of sound which a far-end teleconferencer transmits from returning to them. Traditional approaches to acoustic echo cancellation have used filtering algorithms which try to estimate the impulse response of the acoustic path, h(t), and filter the incoming signal from the far-end, x(n) [1, 2]. The near-end input y(n), e.g., from a microphone, will contain both the far-end sound and the new near-end sound. The far-end sound is convolved with the estimated h(t), and subtracted from y[n] before being sent to the far-end. the estimate is refined by updating the filter according to its output. A common approach for estimating h(t) is the Least Mean Square (LMS) algorithm. This has been discussed in depth [3], and we will now briefly summarize what is relevant to this paper.

By processing the far-end and near-end sound data in blocks, we can exploit the computational advantages of performing the filtering convolutions using the fast Fourier transform

*This work was supported by grants from Valorisation Recherche Québec and Heritage Canada. This support is gratefully acknowledged.

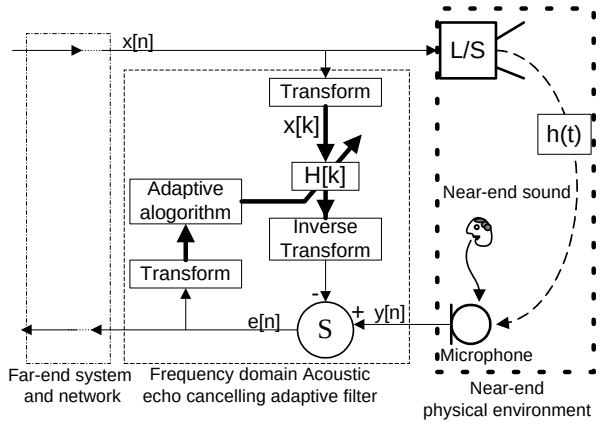

Fig. 1. Functional overview of a basic frequency domain block adaptive filter for acoustic echo-cancelling.

(FFT), though using a time-domain approaches would give similar results [4]. A discussion of the relative merits of time and frequency domain LMS adaptive filtering can be found in [3]. For a block length of N samples, the Frequency Domain Adaptive Filter (FDAF) approach described in figure 1 will give an output delay of N samples.

We define the frequency domain filter weight vector as **H**

$$\mathbf{H}(k)=[H_0(k),...,H_{M-1}(k)]^T \tag{1}$$

and the input signal matrix as

$$\mathbf{X}(k)=\mathrm{diag}\{X_0(k),...,X_{M-1}(k)\} \tag{2}$$

The number of elements, M, is chosen to be of order 2, so that we can use the FFT to transform the input sequence x[n] into the frequency domain. We represent **X**(k) as a matrix so that the output of the filter can be represented simply by:

$$\mathbf{Y}(k)=\mathbf{X}(k)H(k) \tag{3}$$

$$H(k+1)=H(k)+2G\mu(k)X^{H}(k)E(k) \tag{4}$$

The step-size variable μ_m(k) is inversely proportional to the signal power of the m$^t h$ bin of the input signal X(k), P$_m$(k), and a constant μ, where

$$P_m(k)=\lambda P_m(k-1)+\alpha|X_m(k)|^2 \tag{5}$$

Using multiple echo cancelling (EC) filters to estimate the "best" **H**(k) is not a new idea. The so-called two-path model [5, 2] uses a background adaptive filter and a foreground non-adaptive filter. Here, if the background filter is deemed to be better, then its coefficients are copied into the foreground filter (e.g., if double talk is detected). Eneman [6] also suggests using "a sliding step-size μ", and we extend this idea by having n filters, $\hat{H}_n$(k), running simultaneously and independently, which allows the chosen **H**(k) filter to converge at a rate dependant on the chosen variables for that filter algorithm. Each $\hat{H}_n$(k) filter has a unique combination of two constants: the "forgetting factor" α, and the step-size μ. We can therefore change both how much the filter coefficients change from block-to-block as well the weighting of the last filter used. Both α and μ will therefore affect the convergence rate of the filter and will allow the previous filter to remain if the algorithm deems this filter to give the best output. Our criteria for choosing which $\hat{H}_n$(k) to copy to **H**(k) is discussed later.

2. EXPERIMENT DESIGN

2.1. Algorithm details

In our experiments, the frequency-domain multiplication of the transformed far-end input sequence **X**(k) and the time-varying vector **H**(k) was done using the overlap-save sectioning method (as described in [7, 3]. The criteria for choosing which $\hat{H}_n$(k) filter to copy to **H**(k) was decided by the filter which gives the lowest energy output, **Y**(k). This can be calculated in the frequency domain, for example by the summing energy in the output block **H**(k). Alternatively, we can sum the time-domain output block e(k), which is computationally cheaper as we have less than half as many samples to sum when we discard the later half of the 2N product of **X**(k) and **H**(k). We have also experimented with a sub-band energy minimization of **H**(k), whereby only the lowest frequency taps across $\hat{H}$(k)'s are copied to the chosen filter **H**(k), though we found this increased the audible distortion. The block length we used in all experiments was 139 ms, that is, 6144 samples.

2.2. Set-up

All our experiments were conducted in an acoustically isolated room with dimensions of approximately 5*6*2.2 metres. The T60 reverberation time was measured to be approximately 0.34 seconds at 1kHz, which is slightly less than might be expected in a typical conference room. The location of the loudspeaker was 1.5 m from a wall on a 60cm high table.
All measurements on the system we report in this paper were conducted with the microphone held by a stand 1.3m high and 1.5 m away from the speaker, facing the speaker . The sound was sampled and processed at 44.1 kHz at a resolution of 16 bits. The background noise level of the room was approximately 45 dBA, and the sound pressure level at the microphone approximately 65 dB. The recorded sound was processed off-line using MATLAB.

3. RESULTS

Here we will investigate how the EC filter response can be optimized for higher fidelity sound output.
The adaptive filter needs initializing with two data: The time-varying step size used to update the filters $\hat{H}$(k) require an initial estimate of the power levels in the input DFT frequency bins P$_m$[k-1] (equation 5), and the filters $\hat{H}$[0]. We use the metric echo-return-loss-enhancement (ERLE) to show the output of the EC filter. We calculate this as as a power logarithmic ratio of the energy in the output block **Y**(k) to that of the energy in the near-end block **Y**(k) using white noise. This has been used much as a quality metric for EC filter evaluation [8].

3.1. Selection of multi-filter parameters

We created a matrix of 31 different forgetting factors α, and step-sizes μ; each ranging form 0-1 in steps of 0.2.
Figure 2 shows clearly that high step-size and forgetting factors are used as often as very low ones so that the filter can adapt rapidly initially, and then stop adapting once it has settled.

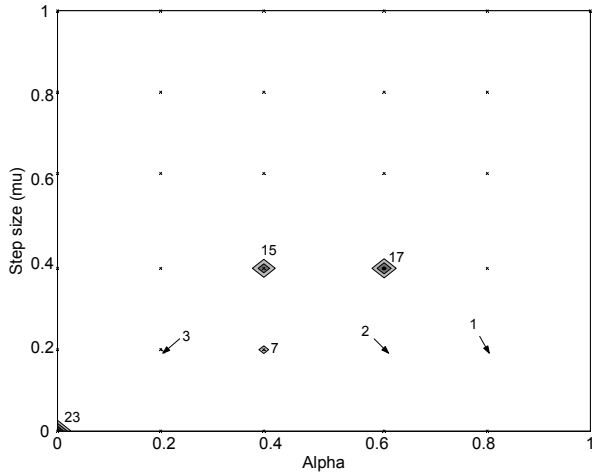

Fig. 2. Plot showing which combinations of α and μ were used most frequently with EC filter for white noise far-end input.

3.2. Initialization of EC filter

If we initialize each of the filters $\hat{\mathbf{H}}(0)$ and $P_m(0)$ with a certain vector, we can increase the rate of initial convergence of the EC filter, as figure 3 shows. The initializations of $\hat{\mathbf{H}}(0)$ are based on known transfer functions from the loudspeaker to the microphone in the near-end environment.

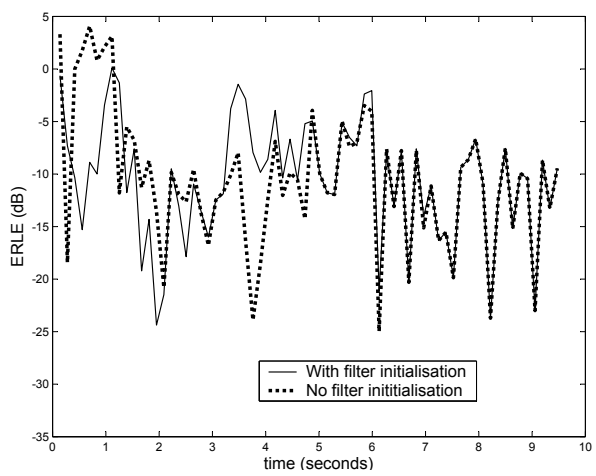

Fig. 3. ERLE for white noise showing the effect of a filter initialized with a selected transfer function

3.3. Updating of filters

Rather than updating each of the $\hat{\mathbf{H}}(k)$ filters with the last chosen filter, i.e. $\mathbf{H}(k)$, we let each filter run independently. This allows a greater range of filter coefficients to be selected.

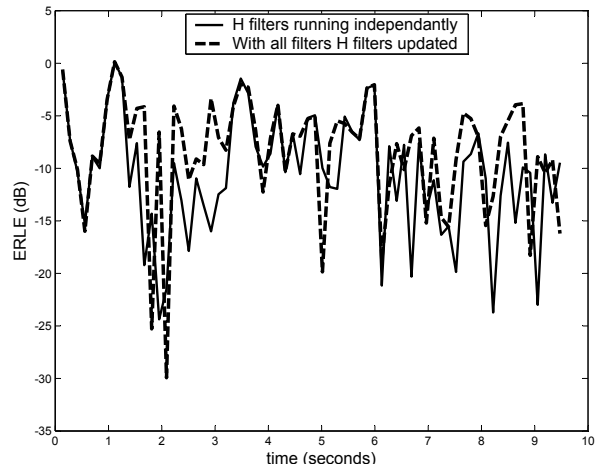

Fig. 4. ERLE for white noise showing the effect of updating all $\hat{H}_n(k)$ filters with the last $\mathbf{H}(k)$.

3.4. Testing of the EC system with speech

Figure 5 shows how the multi-filter system performs with a real speech stimuli at both the far and near ends. We used the 31 filter model, i.e. the same as in section 3.1.

4. DISCUSSION

Figure 5 shows how the new multi-filter approach suppresses the echoes caused by far-end sound better than using a single filter. By comparing the time-domain responses of the multi and single filter (if plots c and d), we can see that during periods of double talk, such as from 11 to 13 seconds, the filter $\mathbf{H}$ stops updating. The ability of the filter to converge rapidly would help in instances when the microphone-loudspeaker transfer function changes rapidly, such as for a moving microphone or when there are large moving objects in the environment which would affect the rooms impulse response.

We can see from plots e and f that α and μ values chosen are highly correlated, as would be expected. We can also see from these plots, as well as from figure 2, that most of the 31 filters are not used. In fact, we tried using just 4 filters with α - μ combinations of 0.02-0.02, 0.1-01, 0.6-0.4, and 0.8-0.5, and found the sound quality to be as good with a normal speech-speech conversation. We were able to run the 4 filter model in real time using a 2 GHz PC and not thoroughly optimized MATLAB code. Dynamic resource allocation is a pertinent design consideration for EC systems [9] such as the one described herein, and a multi-filter system such as ours would benefit computationally by taking advantage of certain temporal patterns in the filter choosing algorithm.

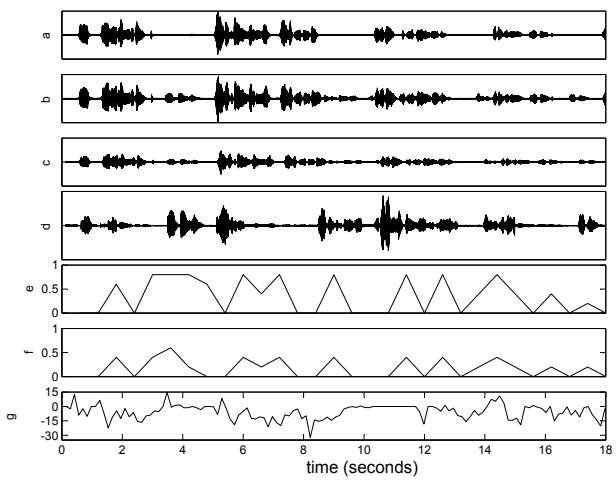

Fig. 5. a: Far-end speech (x[n]); b: Near-end speech (y[n]); c: Output from single filter (alpha=0.1, μ=0.1); d: output from 20-filter model; e: Chosen forgetting factor α ;f: Chosen step value μ; g: ERLE for 20-filter system.

5. CONCLUSION

An acoustic echo cancelling system based on a frequency domain block least mean squares algorithm was designed and tested. The updating of the filter was done by running many filters with different update parameters simultaneously, and using the filter which provides the smallest output. We evaluated the filter with white noise and speech, and found the multi-filter approach to adapt quickly to double-talk conditions. We suggest values for parameters which may be suitable for designing multi-filter echo cancellers.

6. REFERENCES

[1] S. Haykin, *Adaptive Filter Theory*, PrenticeHall, Englewood Cliffs, N. J., 4th edition, 2001.

[2] S. L. Gay and J. Benesty (eds.), *Acoustic Signal Processing for Telecommunication*, Kluwer Academic Publishers, Boston, 2000.

[3] S. Parker G. Clark and S. Mitra, "A unified approach to time- and frequency-domain realization of FIR adaptive digital filters," *IEEE Trans. on Acoustics, Speech and Signal Processing*, vol. ASSP-31, no. 5, pp. 1073–1083, 1983.

[4] B. Widrow and S. Stearns, *Adaptive Signal Processing*, Prentice-Hall, Englewood Cliffs, NJ, 1985.

[5] T. Araseki K.Ochiai and T.Ogihara, "Echo canceller with two echo path models," *IEEE Trans. Communications*, vol. 25, pp. 589–595, 1977.

[6] K. Eneman and M. Moonen, "Real-time implementation of an acoustic echo canceller on dsp," in *Proceedings of the ProRISC/IEEE Workshop on Circuits, Systems and Signal Processing*, Mierlo, the Netherlands, 1997.

[7] J. Shynk, "Frequency-domain and multirate adaptive filtering," *IEEE Signal Processing Magazine*, pp. 15–36, Jan. 1992.

[8] J. Li J. Song and Y.-K. Chen, "Quality-delay-and-computation trade-off analysis of acoustic echo cancellation on general-purpose CPU," in *ASSP*, 2003, pp. 837–840.

[9] DR Morgan MM Sondhi J. Benesty, T. Gänsler and SL Gay, *Advances in Network and Acoustic Echo Cancellation*, Springer, 2001.