MULTIRATE ACOUSTIC ECHO CANCELLATION: WHICH ADAPTIVE FILTERS FOR WHICH SUBBANDS?

Dr. Desmond K. Phillips,
Dept. of Electronic Engineering,
Loughborough University,
Leicestershire LE11 3TU, UK.

Tel: +44 1509 228119; Fax: +44 1509 222854
d.k.phillips@lboro.ac.uk

Prof. Colin F. N. Cowan (RAE / NORTEL chair, Telecomms),
Dept. of Electronic Engineering,
Queen's University of Belfast,
Ashby Building, Stranmillis Road,
Belfast BT9 5AH, UK.
Tel: +44 1232 335475; Fax: +44 1232 667023
c.f.n.cowan@ee.qub.ac.uk

ABSTRACT

Multirate signal processing is one of the best ways to minimise the computational complexity of Acoustic Echo Cancellation (AEC). The subband decomposition of the excitation signal (speech) results in different statistical properties in each subband. As there are a wide range of Adaptive Filters (AFs) with different performances and costs, a challenge is posed in how to match AFs to subbands so as to maximise performance while minimising computational cost. This paper discusses a hypothetical automated benchmarking methodology that addresses some of the issues raised by the attempt to optimise multirate AEC architectures as a function of desired performance. In this context, a performance comparison of the NLMS and RLS algorithms is conducted.

1. INTRODUCTION

1.1 Overview of the AEC problem

The task of an AEC is to neutralise the disconcerting echo that a far-end talker experiences of their own speech when the near-end talker is using a 'hands-free' telephone. The microphone and speaker signals ($x[n]$ and $y[n]$) are coupled acoustically via a reverberant enclosure (e.g. a car cabin) with transfer function $H(z)$. A traversal FIR AF of L taps performs a system identification of $H(z)$ to yield an error residual $e[n]$ whose power the AF seeks to minimise: L is chosen to match the length of the impulse response of $H(z)$. However, any near-end speech activity should be transmitted unaffected so that the AEC process is transparent to both talkers, particularly during double talk. An AEC algorithm should have fast convergence but, perhaps more importantly, should also be able to track a time-varying $H(z)$ (caused by changes in, say, near-end talker position). [3].

1.2 Multirate AEC

The principal obstacle to economical AEC implementation is that, typically, values of L>1000 are required in order to model $H(z)$ accurately resulting in a high computational load relative to other digital audio operations (e.g. GSM) that one would expect of a mobile handset. An elegant solution is to split $x[n]$ and $y[n]$ into N integer-spaced subbands using critically-sampled multirate analysis filterbanks. N independent AFs, each of length L/N, are then deployed to generate N subband error residuals $e_i[m]$:$\{1 \le i \le N\}$ which are recombined by a synthesis filterbank into $e[n]$ as illustrated in Figure 1 for the simple case of N=2. If the filterbanks have negligible overheads, then the computational burden of such a Multirate AEC (MAEC), compared to the fullband case, is only $1/N$.

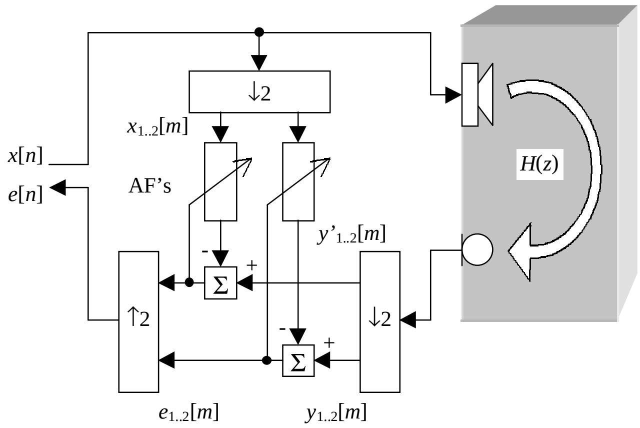


Figure 1. An MAEC with N=2 subbands

A high performance MAEC is predicated on well-designed analysis / synthesis filterbanks. The allpass polyphase IIR structures developed at Imperial College [5] are oriented towards specific application in a MAEC. Based upon QMF principles, they have low computational cost, low delay and imperceptible phase distortion while possessing sharp transition widths. Inter-subband aliasing at the synthesis filterbank output is manifest as narrow peaks coinciding with the analysis QMF transition regions and is eliminated by a notch filter at the input to each QMF analysis stage. If the subband decomposition is relatively coarse (i.e. $N \le 16$), then the resulting spectral nulls are narrow enough to give no objective deterioration in speech intelligibility.

1.3 Choice of AF Algorithm

An additional property of a MAEC is that each subband of the speech signal has unique statistical properties. The general $1/f$ spectral envelope of speech implies that the relative power contribution of each subband to the echo residual $e[n]$ rolls off with frequency. Related to this statistical variety is the diversity of AFs available to the engineer for AEC implementation: a review is provided in [3]. The most popular algorithm is the Normalised Least-Mean Squares (NLMS) algorithm which has the lowest computational cost of order($2L$) and is numerically robust in short wordlength implementations.

At the other end of the spectrum are higher complexity algorithms such as Fast-RLS (Recursive Least-Squares) which has fastest convergence but is sensitive to finite precision effects, requiring computation of order ($10L$) in its stabilised form [4]. Between these two poles lie a variety of AFs which have captured recent attention which seek to provide RLS performance at NLMS cost, and include a design variable that controls the tradeoff between the two: the Fast Newton Traversal Filter (FNTF) is a popular choice [3]. The important question arises, therefore, of which AF to use in which subband (and similarly of how many subbands N to use) leading to the concept of heterogeneous MAEC architectures and of robust methods for deriving them [7].

2. A BENCHMARKING STRATEGY

A conceptual tool that would be of advantage to an AEC designer would be a graph of AEC performance *versus* computational complexity. It would represent (i) the maximum achievable performance for a specified computational cost and by symmetry, (ii) the minimum computation required to achieve a specified performance target. AEC performance is quantified succinctly in the traditional metric of Echo Return Loss Enhancement (ERLE) expressed in eqn. (1). Computational cost for an FIR AF is usually measured in terms of an integer multiple of L, but an additional factor is the precision (wordlength) necessary for stable operation: a benchmark combining both is the ALU bandwidth required in Mbytes / second.

$$ERLE = 10\log_{10}\left(E\left[x[n]^2\right] / E\left[e[n]^2\right]\right) \tag{1}$$

The first stage towards automated benchmarking is to select a subset of C candidate AFs deemed suitable for AEC implementation. Next, a limit must be imposed on N to limit QMF depth before artefacts of the subband decomposition become perceptually apparent (i.e. nulls caused by notch filters): $\max(N)=16$ is considered appropriate. Therefore $N\in\{1,2,4,8,16\}$. To provide a complete search space for the optimisation problem, the performance and computational cost of each architectural permutation of AFs at each value of N must then be benchmarked by extensive simulation. However, the number of permutations at a particular QMF depth is C^N implying an intractable amount of simulation at even moderate values of C and N. A simpler approach is to exploit the mutual orthogonality of subbands and benchmark each independently, in parallel, during the simulation of a homogeneous MAEC composed of N identical AFs. The total number of simulation runs thus required is a tractable $2C(\max(N)\text{-}1)$. The fullband benchmark for any architectural permutation may then be generated theoretically by a simple look-up and accumulation of the appropriate subband benchmarks and computational costs based on the additive power property expressed in eqn. (2).

$$E\left[e[n]^2\right] = \sum_{i=1}^{N} E\left[e_i[m]^2\right] \tag{2}$$

At our disposal is a large database (FREETEL) numbering R=463 synchronous microphone / speaker (SM) pairs of 16-bit 8kHz sample rate recordings taken in a variety of real acoustic environments with ambient noise [2]. Male and female talkers enunciate phonetically balanced utterances in one of three languages (English, French and Spanish). It is a valuable resource for off-line AEC benchmarking since it permits the compilation of large-scale ensemble statistics. With too small a database, benchmarks risk becoming data-specific and inapplicable to the general case. For each simulation of a candidate AF with an SM pair, a set of N cumulative error and microphone powers permit a mean ERLE to be computed for each subband. Mean ERLE is a useable metric since it encapsulates both AF convergence speed and misadjustment error in a single meaningful value. The next problem is how to integrate the results from R SM pairs into a single value. Mean ERLE histograms for each subband yield a consistent pattern manifest as a normal-like distribution. The median value of this set is chosen as the final ensemble-averaged value. This benchmark, denoted B, is expressed in eqn. (3) where the i:{$1\le i\le N$}and j:{$1\le j\le R$} respectively index individual subbands and SM pairs in the FREETEL database.

$$B(i, N, \text{AF Type}) = \operatorname{med}_{j=1}^{R}\left(10\log_{10}\left(E\left[y_i^j[m]^2\right] / E\left[e_i^j[m]^2\right]\right)\right) \tag{3}$$

Assuming a complete set of benchmarks, MAEC architectural optimisation begins with the specification of a desired mean fullband ERLE, denoted D (in dB) which is, explicitly, a wide-sense interpretation of eqn. (1) for single-talk. Next, it is required that the fullband error residual $e[n]$ should be as ‘white’ as possible. The justification is that (i) the echo is ‘smeared’ across frequency making it perceptually less annoying, particularly with a large number of subbands (N) and (ii) it is intuitively attractive since higher-performance AFs are favoured for subbands containing the bulk of the echo power. Prior to AEC, the expected power contribution of each subband $y_i[m]$ to the uncancelled fullband echo $y[n]$, denoted $P(i,N)$ is expressed by eqn. (4) using, for consistency, the same approach as in eqn. (3). After AEC, the desired power ratio of each subband echo residual $e_i[m]$ to $y[n]$, denoted $Q(N)$, is given in eqn. (5) and is a function of D and the equal power division between N subbands for a white signal. Therefore the target ERLE for each subband, denoted $T(i,N)$, is given by eqn. (6). Optimisation thus consists of (on a subband-by-subband basis) identifying the set of AFs which satisfy eqn. (7) and then choosing the one with lowest computational cost.

$$P(i,N) = \text{med}_{j=1}^{R}\left(10\log_{10}\left(E\left[y_i^j[m]^2\right]/E\left[y^j[n]^2\right]\right)\right) \quad (4)$$

$$Q(N) = -D - 10\log_{10}(N) \quad (5)$$

$$T(i,N) = P(i,N) - Q(N) \quad (6)$$

$$B(i,N,\text{AF Type}) > T(i,N) \quad (7)$$

3. EXPERIMENTAL RESULTS

3.1 Method

A single-talk subset (R=128) of FREETEL was chosen (omitting the complexities of double-talk detection) which used the same acoustic front-end and room environment requiring an AF of approximately L=500 taps for accurate modelling. Five individual talkers are featured. For the purposes of benchmarking, an over-determined fullband AF of length L=1024 was used to minimise the risk of subband AFs becoming under-determined due to impulse response length variation between subbands. The NLMS stepsize parameter μ was fixed, but tried at three values of $\mu\in\{0.1,0.3,1.0\}$ to take into account the different properties of subbands: a variable stepsize is ideal, but the chosen range was found to be satisfactory in practice. Eqn. (3) was applied for each value of μ and the maximum benchmark for each subband was recorded.

For the RLS simulations, the 8L Fast-RLS [4] algorithm was implemented in double-precision. Instability was not encountered as the run-length of each SM pair was finite, typically of the order of 10^5 samples. The forgetting factor λ was varied over the range $\lambda\in\{1, 0.99999, 0.99997, \ldots 0.99, 0.97\}$ in each subband in order to identify the optimum time constant. It is a well-known fact that for a non-stationary $H(z)$, NLMS and RLS often have a comparable tracking performance governed by the 'memory' effect of, respectively, μ and λ, often leading to a preference for NLMS [2][3]. An alternative to covariance domain algorithms such as RLS, which involve the numerically ill-conditioned step of 'squaring' and inverting the data matrix are data-domain algorithms using QR factorisation which are less sensitive to roundoff error [4]. However, given the luxury of off-line double-precision arithmetic, it was found that the results of Fast-QR [4] were almost identical to Fast-RLS. This is because that both are different ways of rewriting the same least-squares problem involving λ.

Alongside NLMS and RLS, a modified NLMS-type order(2L) algorithm developed at Loughborough was benchmarked in order to evaluate its performance [6]. It shares the premise made by the FNTF [3] in assuming a low-order autoregressive excitation model but has an implementation involving a simple adaptive preconditioning filter to NLMS: a similar structure has also been proposed in [1] with differing normalisation.

3.2 Analysis of the Results

The benchmarks for the three AF methods are plotted in Figure 2 for each subband i up to a maximum of N=16. At N=1 (i.e. fullband) both Fast-RLS and the modified NLMS exhibit a performance improvement of +2.5dB over standard NLMS. The FREETEL subset has a low ambient S/N ratio of the order of 15-20dB and explains the small improvement margin which might be expected to be larger. However, the modified NLMS result is comparable to Fast-RLS despite their large difference in computational complexity.

As the subband decomposition becomes progressively deeper, a clear difference emerges in that Fast-RLS gives superior performance to NLMS methods in the lower 30%-40% of the spectrum. However, in the upper spectrum, the three methods are roughly comparable. The reason is that speech has the highest signal power in the lower subbands and thus the ambient S/N ratio is much better; a property that can be exploited by fast-converging RLS methods. In the upper subbands, near-end additive noise leads to considerable AF misadjustment.

In order to be able to combine subband results into a fullband version, $P(i,N)$, the power contribution of each subband to the fullband echo must be estimated. $P(i,N)$ is plotted in Figure 3 and shows some interesting ensemble characteristics as $N\rightarrow 16$. At N=16, maximum $P(i,N)$ occurs in the 3rd subband spanning 500-750Hz and, in constrast, there is a minimum $P(i,N)$ in the region of the 8th and 9th subband spanning 1750-2250Hz. The former is characteristic of the speech spectrum, but the latter phenomenon is partly a filterbank artefact [5].

3.3 Interpreting the Results

N	1	2	4	8	16
Best ERLE (dB)	12.7	12.4	13.5	13.4	12.4

Table 1. Estimated Fullband ERLEs for Optimal MAECs

Specifying our desired ERLE at, say, D=15dB, and applying the benchmarking strategy of section 2, we obtain the results of Table 1. The best option (in terms of D) is N=4: there is a qualitative recommendation of Fast-RLS for the lowest frequency subband and of NLMS-type methods for the higher frequency ones. These results are the best achievable for the chosen AF set and FREETEL subset.

4. CONCLUSIONS

The benchmarking strategy is formulaic and should not be applied too rigidly to a complex engineering problem like AEC. However, as we have seen, it brings to light in a quantitative fashion the factors that can bias the choice for one AF over another in individual subbands. For instance, the advantage of RLS over NLMS in low frequency subbands is demonstrated.

ACKNOWLEDGEMENTS

This work is part of the EPSRC research project "Multirate Acoustic Echo Cancellation" (GR/K48617).

REFERENCES

[1] Farhang-Boroujeny B., "Fast LMS-Newton Algorithms Based on Autoregressive Modeling and Their Application to Acoustic Echo Cancellation", *IEEE Trans. on Signal Proc.*, vol. 45, no. 8, pp. 1987-2000, August 1997.

[2] Gilloire A. & Petillon T., "A Comparison Of NLMS And Fast-RLS Algorithms For The Identification Of Time-Varying Systems With Noisy Outputs - Application To Acoustic Echo Cancellation", *EUSIPCO-90*, Barcelona, Spain,18 - 21 Sep 1990, pp. 417-420.

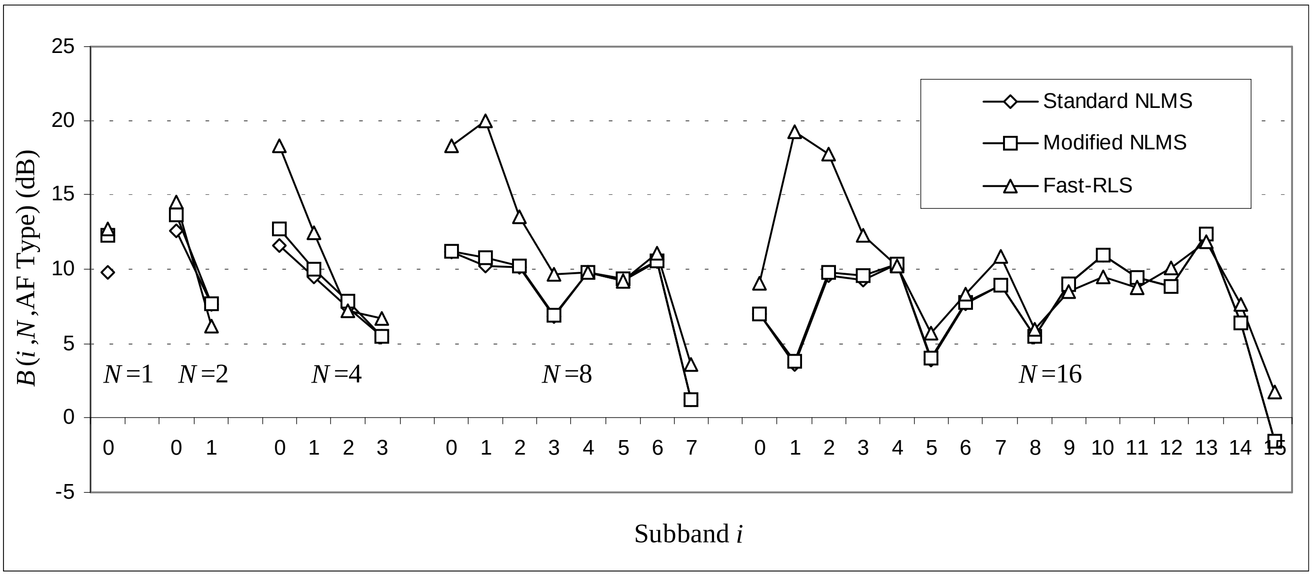
[3] Hänsler E.,"From Algorithms to Systems – It's a Rocky Road", Keynote Speech, *IWAENC'97*, Sept 11-12, 1997, Imperial College, London, pp. K1-K4.

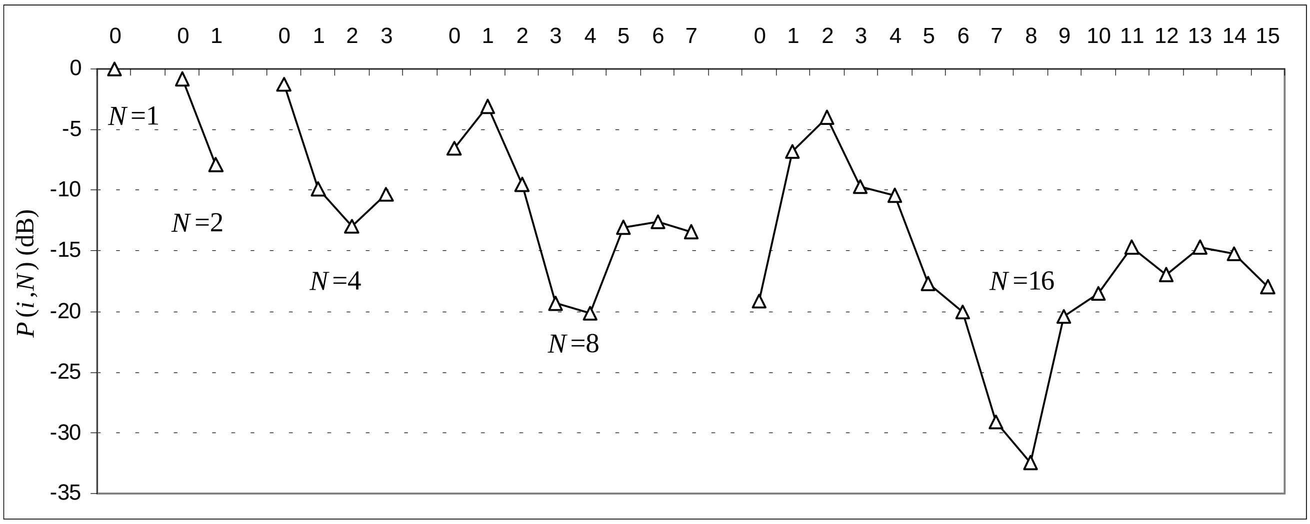
[4] Haykin S. *Adaptive Signal Processing*, Prentice-Hall, Englewood Cliffs, NJ, 1991.

[5] Naylor P., Tanrikulu O. & Constantinides A. G., "Subband Adaptive Filtering for Acoustic Echo Control using Allpass Polyphase IIR Filterbanks", *IEEE Trans. on Speech and Audio Proc.*, vol. 6, no. 2, pp. 143-155, March 1998.

[6] Phillips D. K. & Cowan C. F. N., "Stability Of NLMS with Zero-Phase Signal Conditioning for Acoustic Echo Cancellation", *IEE (E5) Colloq on Adaptive Sig. Proc. for Mobile Comms. Sys.*, 29th Oct, 1997, Savoy Place, London, pp. 2/1-2/4.

[7] Phillips D. K. & Cowan C. F. N., "Heterogeneous Strategies for Multirate Acoustic Echo Cancellation", *IWAENC'97*, Sept 11-12, 1997, Imperial College, London, pp. 164-167.





Figures 2 and 3. Benchmark Results