HETEROGENEOUS STRATEGIES FOR MULTIRATE ACOUSTIC ECHO CANCELLATION

Dr. Desmond K. Phillips,

Dept. of Electronic Engineering,
Loughborough University,
Loughborough,
LEICS, LE11 3TU, UK.

d.k.phillips@lboro.ac.uk

Professor Colin F. N. Cowan,

Dept. of Electronic Engineering,
Queen's University of Belfast,
Ashby Building, Stranmills Road,
BELFAST, BT9 5AH, UK.

c.f.n.cowan@ee.qub.ac.uk

ABSTRACT

Each subband in a multirate acoustic echo canceller has different statistical properties. As there is a wide range of adaptive filters to choose with different cost / performance trade-offs, it is important to target computational resources intelligently depending on the perceptual contribution of each subband in the overall error residual. This results in the scenario of a heterogeneous AEC. For this optimisation problem, both a pertinent methodology and reliable source data through comprehensive benchmarking are required. This paper discusses some of the methodological issues raised by AF benchmarking in subbands.

1. INTRODUCTION

Multirate DSP provides a successful strategy for decomposing the computationally intensive problem of an Acoustic Echo Canceller (AEC) for hands-free telephony. The speaker and microphone signals $\{x[m], y[m]\}$ are decomposed using an analysis filterbank into N integer-spaced subbands allied to N subband Adaptive Filters (AF's). Their output is recombined in a synthesis filterbank to give the error residual $e[m]$. Principal benefits are (i) the computational cost is $1/N$ relative to the fullband counterpart and (ii) subband spectra are relatively 'white' compared to the fullband giving improved AF convergence. Requirements for the analysis and synthesis filterbanks are that they should have low cost, delay and phase distortion (to maintain signal fidelity). Additionally, they should minimise subband transition widths and overlap (to eliminate inter-band aliasing without spectral nulls): these requirements are satisfied by the novel polyphase-allpass QMF filterbank scheme incorporating transition region notches developed at Imperial College [1].

The next logical challenge in this project (EPSRC GR/K48693) is to optimise the subband AF's. Beyond the basic NLMS algorithm, there is a diversity of higher complexity / performance algorithms that are potential candidates; contrasting philosophies are exemplified by [2] [3] [4]. Recent work at Loughborough has also produced some new low-cost pre-whitened NLMS-variants that display promising results for implementing AEC's [5] [6] [7]. To this end, a design tool is under development that assists in determining which AF is best for which subband. Two factors of a multirate AEC hint at heterogeneity. Firstly, the statistics of each subband is unique implying a subband-specific performance bias for certain AF's over others: the lowest frequency subband is characterised by a harmonic signal content in contrast to fricative noise bursts in the highest frequency subband. Secondly, given the $1/f$ spectrum of speech, the perceptual weighting of each subband in $e[m]$ has a roll-off with increasing frequency.

2. METHODOLOGY

2.1. On Asking the Right Question

$$\mathrm{ERLE} = 10\log_{10}(E[y[m]^2]/E[e[m]^2]) \qquad (1)$$

One of the most important design variables in an AEC implementation is the desired performance criterion, usually related to the Error Return Loss Enhancement (ERLE) of eqn. (1). System distance is not an option because the impulse response of the echo path is difficult to determine *a priori* with accuracy. A popular measure is the convergence time to a particular ERLE value. Given such a quantity, say d, which AEC architecture is the most efficient? However, this question raises the higher-level issue of whether the basis for such a pre-emptive election of d is robust. This is because d equates to the subjective level of disturbance caused to the far-end speaker, rather than an objective metric of speech intelligibility that must be satisfied.

Arbitrariness in the selection of d is undesirable. Conceivably, a slight sacrifice in d may yield a much more efficient AEC, or d may tend towards an

asymptotic limit (e.g. the additive noise floor). Though an AEC is optimal for a certain value of d, it may be possible to make computational savings without degrading significantly the perceived AEC quality. By transposing the optimisation problem to a higher level, optimal AEC architecture (and its associated computational cost) becomes a function of d allowing greater rationality in the AEC design process. Therefore, this objective is dominant in our benchmarking methodology.

2.2. Optimising to Mean ERLE

One of the problems in determining ERLE convergence time is that speech is non-stationary and the consequent (block segmented) ERLE envelope from an AEC has a wide dynamic range. For instance, additive noise during the short silences in an utterance can yield an instantaneous negative ERLE. However, if a stationary excitation signal such as USASI noise is used, which has a speech-like spectral envelope, the ERLE envelope displays better convergence properties more amenable to analysis. A problem with signals of this type is that they ignore the non-stationary transient behaviour inherent to natural human speech by which AEC benchmarking ought to proceed.

Our work follows such an approach and is facilitated by the FREETEL database which comprises a phonetically balanced set of $\{x[m], y[m]\}$ speaker microphone (SM) pairs taken across a comprehensive range of environments and talkers. In the light of the problems with other benchmarks, mean ERLE across each SM pair provides an attractive alternative for the following reasons:

- A single intuitive figure represents the total echo power reaching the far end speaker.
- It encapsulates AF convergence, misadjustment and tracking properties (for a non-stationary echo path).
- Dimensionality of the optimisation problem is minimised.

An anticipated refinement to mean ERLE is to integrate some form of auditory model in order to arrive at a perceptual benchmark of echo volume.

2.3. Factorising the Benchmarking Process

For reliable results it is desirable to benchmark the performance of a particular AEC architecture by analysing the ensemble statistics of a large number of SM pairs to smooth out individual talker / environment peculiarities. However, a problem for a heterogeneous AEC is that for K AF prototypes and N subbands, there are K^N potential architectures, leading to an intractable number of simulations even if K and N are low (e.g. $K>10$, $N=4$). However, as each subband AF is independent and subbands have negligible cross-correlation, only K simulation runs are required, each generating N couplets of accumulated $y[m]$ and $e[m]$ powers.

To compute mean ERLE for any of the K^N potential architectures, the pertinent subset of N couplets from the resulting set (size KN couplets) are summed and then applied to eqn. (1). This process has insignificant overheads in comparison to a complete simulation of a particular architecture through FREETEL. Different QMF depths are logged in the range $N \in \{1,2,4,8,16\}$ to maximise the diversity of AEC architectures available for solving the optimisation problem: N, too, is an important AEC design variable. In consequence, AEC architectural optimisation is decoupled from the requirement for time-consuming and repetitive simulations.

3. A BENCHMARKING EXPERIMENT

3.1. Method

An initial step towards fully automated AEC benchmarking along these principles was to investigate the performance of one of the new pre-whitened NLMS variants developed at Loughborough [5] with conventional NLMS. As the extra overheads in [5] are negligible, computational costs are assumed to be identical. The primary purpose was to validate the new algorithm with actual speech in support of this research. Secondly, it raises the question of how to integrate a set of mean ERLEs into a single benchmark (say B). As the human ear has a logarithmic perception of audio power, an intuitive 'first-cut' approach is to compute B (dB) as the mean of mean ERLEs, as expressed in dB's.

A subset of the FREETEL database was chosen comprising 128 single-talk (i.e. no near-end speaker activity) SM pairs in an acoustic enclosure with an impulse response equivalent to 512 taps at f_s=8kHz. NLMS was simulated at three different stepsizes of $\mu_1=\{0.1, 0.3, 1.0\}$ over 3 QMF filterbank depths with $N \in \{1, 2, 4\}$. The new technique [5] was tested over the same range with the additional parameter of a predictor stepsize $\mu_2=\{1e^{-4}, 1e^{-5}...1e^{-12}\}$ which has the default order of 2: higher orders do not yield a significant performance improvement [5]. Hence the number of candidate AF's was K=30.

3.2. Interpreting the Results

Figs. 1 to 7 plot B as a function of three variables μ_1, μ_2 and N. The first noticeable feature is the absence of results for $\mu_2=\{1e^{-4}...1e^{-7}\}$. This is because predictor adaptation noise [5] causes instability in one or more of

the SM pair simulations. In the fullband case of Fig. 1, μ_1=0.3, μ_2=1e^{-10} yields slightly superior results to NLMS with μ_1=1.0. Additionally, a stepsize of μ_1=0.1 is shown to be inferior to the other values of μ_1={1.0, 0.3} which share similar optimality. With μ_1=1.0, the single value of μ_2=1e^{-12} alone yields stability. The best stepsize appears to be μ_1=1.0 with a slender case for using [5] instead of NLMS.

In the *N*=2 case of Figs. 2 and 3, a large improvement of about +2.5dB in *B* over Fig. 1 is evident, probably due to the whitening effect of a subband decomposition. Again, the optimal NLMS stepsize is μ_1=1.0 with both μ_1={0.3, 0.1} significantly inferior. With μ_1={0.3, 0.1}, it can be seen that *B* has a perceptible maximum in subband #1 at μ_2=1e^{-10} indicating this to be an optimal predictor stepsize value. The *N*=4 case of Figs. 4 to 7 yields some interesting results. The optimal value is μ_1=1.0 for subband #1 and (in the region of) μ_1=0.3 for the higher subbands. In subband #2, there is stronger evidence for a performance improvement of [5] at μ_1=0.3, μ_2=1e^{-8} over NLMS than in any other graphs, though only of the order of +0.5dB.

3.3. Conclusion

The variation of optimal μ_1 with subband# indicates the utility, if NLMS is to be deployed, of an adaptive stepsize. Other possible augmentations are exponential coefficient weighting [8] or adaptive tap-assignment [9]. Unfortunately, the case for [5] over NLMS is unproven and further analysis and simulation is underway presently to give a greater insight. However, the validity of the 'mean-of-means' benchmark *B* is supported by the evidence of listening tests when sample SM pairs are played back through the contrasting AEC architectures.

4. FUTURE WORK

A satisfactory benchmarking methodology is established which decouples simulation from optimisation with a consequent reduction in computational overhead. Concurrently, the methodology is framed in the correct context by seeking to solve optimal multirate AEC architecture as a function of fullband *B*. The next step is to include as wide a range of candidate AF's in the simulation stage as feasible and to develop a software tool for solving the multirate AEC architectural optimisation problem.

5. REFERENCES

[1] Tanrikulu O., Buyurman B., Constantinides A. G. & Chambers J. A., "Residual Echo Signal in Critically Sampled Subband Echo Cancellers based on IIR and FIR Filterbanks", submitted to *IEEE Trans. on Sig. Proc.*

[2] Moustakides G. and Theodoridis S., "Fast Newton Traversal Filters - A New Class of Adaptive Estimation Algorithms", *IEEE Trans on Sig. Proc.*, vol. 39, pp. 2184-2193, Oct. 1991.

[3] Tanaka M., Kaneda Y., Makino S. and Kojima J., "A Fast Projection Algorithm for Adaptive Filtering", *IEICE Trans. Fundamentals*, vol. E78-A, no. 10, pp. 1355-1361, Oct. 1995.

[4] Acker C. and Vary P., "Combined Implementation of Predictive Speech Coding and Acoustic Echo Cancellation", *SIGNAL PROCESSING VI*, Theories and Applications, 1992, Elsevier Science Publishers, pp. 1641-1644.

[5] Ndungu E. N. and Cowan C. F. N., "A New Pre-whitened Adaptive Algorithm for Acoustic Echo Cancellation", under review for publication in the *IEEE Trans. on Sig. Proc.*

[6] Phillips D. K., and Cowan, C. F. N. (1996), "A Comparison of Prewhitening Techniques for Subband Acoustic Echo Cancellation", *16th IEE Saraga Colloquium on Digital and Analogue Filters*, Savoy Place, London, 9/12/96, pp. 10/1-10/4.

[7] Phillips D. K. and Cowan, C. F. N. (1997), "Zero-phase Signal Conditioning for Improved NLMS Convergence", invited for the 13th International Conference on DSP, Santorini.

[8] Makino S., Kaneda Y. and Koizumi N., "Exponentially Weighted Stepsize NLMS Adaptive Filter Based on the Statistics of a Room Impulse Response", *IEEE Trans on Speech and Audio Proc.*, vol. 1, pp. 101-108, Jan. 1993.

[9] Sugiyama A. and Hirano A., "A Subband Adaptive Filtering Algorithm with Adaptive Intersubband Tap-Assignment", *IEICE Trans. Fundamentals*, vol. E77-A, no. 9, pp. 1432-1438, Sept. 1994.

6. RESULTS

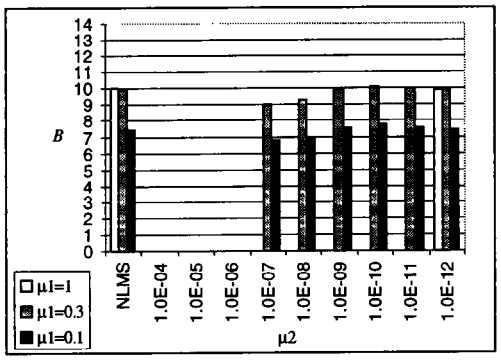

Figure 1. Fullband

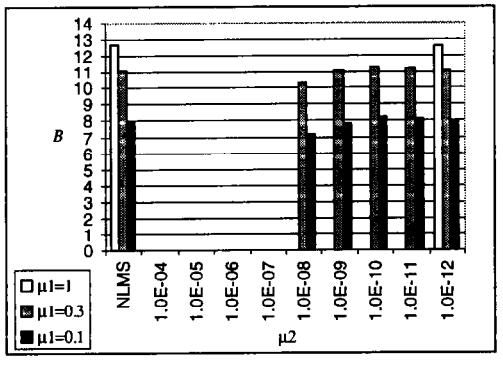

Figure 2. Subband #1 of N=2

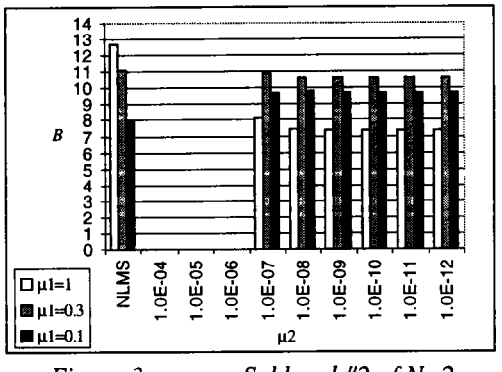

Figure 3 Subband #2 of N=2

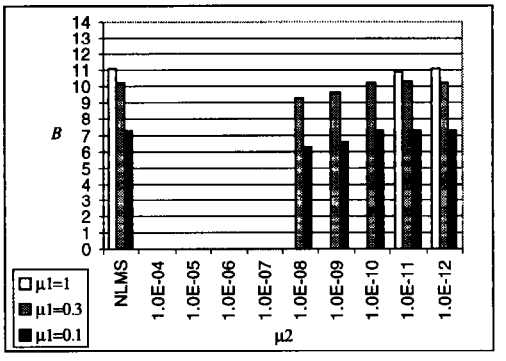

Figure 4. Subband #1 of N=4

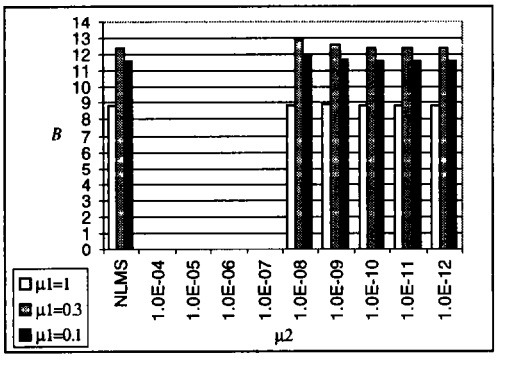

Figure 5. Subband #2 of N=4

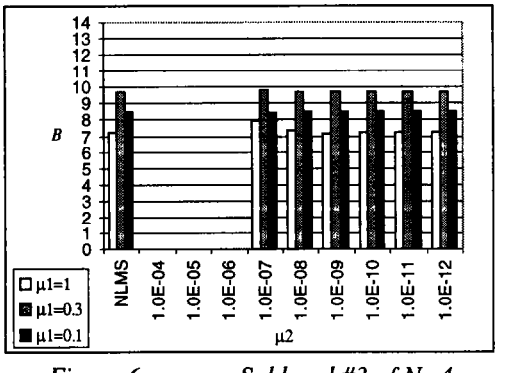

Figure 6. Subband #3 of N=4

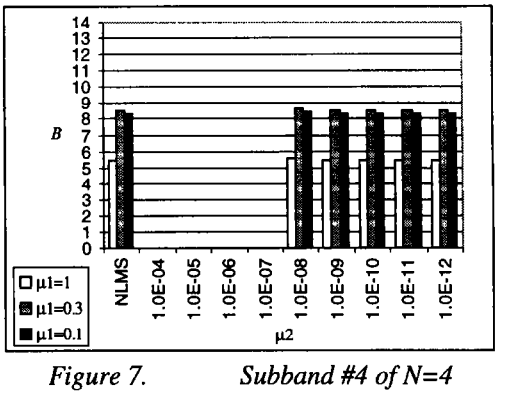

Figure 7. Subband #4 of N=4