On the use of the FNTF algorithm in subband acoustic echo cancellation

Benoît Champagne and François Caron
INRS-Télécommunications, Université du Québec, 16 place du Commerce
Verdun, Québec, Canada, H3E 1H6; champagne@inrs-telecom.uquebec.ca

1. Introduction

Acoustic echo cancellation (AEC) is an effective approach for the control of acoustic echoes generated by hands-free terminals [1]. Even though RLS adaptive filtering achieves faster convergence, the NLMS algorithm is used in most AEC applications because of its low complexity of $2N$ multiplies per iteration (mpi), where N is the number of filter taps. The identification of the long echo paths found in teleconference applications (e.g. $N = 1000$ at 8kHz sampling rate) further requires the use of subband processing to bring down the computational complexity within acceptable limits [2].

Recently, the fast Newton transversal filter (FNTF) algorithms have been proposed in an attempt to bridge the performance gap between the NLMS and the fast RLS (FRLS) algorithms [3]. FNTF models the excitation signal as an AR(M) process, where $0 \leq M \leq N$, and achieves a complexity of $2N + 12M$. In [4], FNTF is shown to be an attractive candidate for AEC applications in the mobile context (short filters, $N = 250$ at 8kHz sampling) since significant improvements in convergence speed over NLMS can be obtained with small values of M. This conclusion does not hold for teleconference applications, where FNTF may lead to a loss of performance.

The investigation in [4] is limited to the use of a single (i.e. full-band) transversal adaptive filter. In the case of long impulse responses, FNTF might benefit from a subband implementation because of the reduced adaptive filter length in subbands, as a result of downsampling. In this work, we investigate the performance of FNTF in a subband AEC structure, with emphasis on the identification of long echo paths typical of teleconference applications.

2. The FNTF algorithm

Application of adaptive system identification to AEC is illustrated in Fig. 1. The unknown system $\mathcal{H}$ consists of the loudspeaker, acoustic medium and microphone. The system input is the far-end signal $u(n)$, where $n \in \{1, 2, ...\}$ is the discrete-time, and the system output is the microphone signal $d(n)$, which contains additive noise and possibly local speech. The unknown system is modeled by an adaptive transversal FIR filter operating on $u(n)$. The time-varying coefficients of the FIR filter are denoted by $h_k(n)$, $k = 0, 1, ..., N-1$, and the filter output is computed as $\hat{d}(n) = \mathbf{h}(n)^T \mathbf{u}(n)$, where $\mathbf{h}(n) = [h_0(n), ..., h_{N-1}(n)]^T$ and $\mathbf{u}(n) = [u(n), u(n-1), ..., u(n-N+1)]^T$. The filter weight vector is recursively adjusted in real-time so as to minimize the power of the error signal, defined as $e(n) = d(n) - \hat{d}(n)$. Practical operation of the adaptive filter requires the use of a double-talk detector (not considered in this study).

The FNTF algorithms belong to a modified class of stochastic Newton (SN) adaptive algorithms:

$$\mathbf{c}_N(n) = -\lambda^{-1} R_N^{-1}(n-1)\mathbf{u}(n), \quad \gamma(n) = 1 - \mathbf{c}_N^T(n)\mathbf{u}(n) \quad (1)$$

$$e(n) = d(n) - \mathbf{h}^T(n)\mathbf{u}(n), \quad \epsilon(n) = e(n)/\gamma(n) \quad (2)$$

$$\mathbf{h}(n+1) = \mathbf{h}(n) - \epsilon(n)\mathbf{c}_N(n) \quad (3)$$

[1] Support for this work was provided by FCAR.

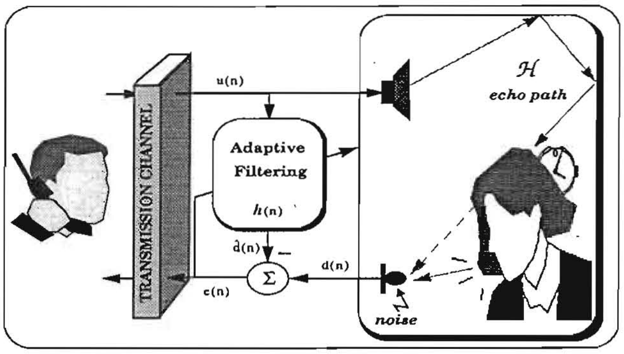

Figure 1: Adaptive identification applied to AEC

where $R_N(n)$ is an estimate of the data covariance matrix, $0 < \lambda \leq 1$ is a forgetting factor, $\mathbf{c}_N(n)$ is a generalized dual Kalman gain, $e(n)$ is the *a priori* estimation error, $\epsilon(n)$ is the *a posteriori* error, and $\gamma(n)$ is a conversion factor. Both NLMS and FRLS can be obtained from (1)-(3) with a proper choice of $R_N(n)$ in (1). In [3], an additional AR(M) assumption on $u(n)$, where $0 \leq M \leq N$, is exploited to derive time-order recursions for the extension of an $(M+1)$th order covariance matrix into the desired N-order one, i.e. $R_N(n)$. Upon substitution of these extension formulae in (1), three distinct FNTF versions are obtained with complexity $2N + O(M)$.

In the context of AEC, practical considerations point to the use of FNTF Version 1 [4]. The latter, used in our work, is summarized below: (a) Using a FRLS forward predictor of order M applied to $u(n)$, update the forward predictor weight vector $\mathbf{a}_M(n-1)$, the residual error $e_M^f(n)$ and the error power $\alpha_M^f(n-1)$ and compute

$$\mathbf{s}_{M+1}(n) = \frac{e_M^f(n)}{\lambda \alpha_M^f(n-1)} \begin{bmatrix} 1 \\ -\mathbf{a}_M(n-1) \end{bmatrix} \quad (4)$$

(b) Using a FRLS backward predictor of order M applied to $u(n_d)$, where $n_d = n - N + M$, update the backward weight vector $\mathbf{b}_M(n_d - 1)$, the residual error $e_M^b(n_d)$ and the error power $\alpha_M^b(n_d - 1)$, and compute

$$\mathbf{t}_{M+1}(n_d) = \frac{e_M^b(n_d)}{\lambda \alpha_M^b(n_d - 1)} \begin{bmatrix} -\mathbf{b}_M(n_d - 1) \\ 1 \end{bmatrix} \quad (5)$$

(c) Update the dual Kalman gain $\mathbf{c}_N(n)$ and $\gamma(n)$:

$$\begin{bmatrix} \mathbf{c}_N(n) \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ \mathbf{c}_N(n-1) \end{bmatrix} - \begin{bmatrix} \mathbf{s}_{M+1}(n) \\ \mathbf{0}_{N-M} \end{bmatrix} + \begin{bmatrix} \mathbf{0}_{N-M} \\ \mathbf{t}_{M+1}(n_d) \end{bmatrix} \quad (6)$$

$$\gamma(n) = \gamma(n-1) + s_{M+1}^1(n) e_M^f(n) - t_{M+1}^{M+1}(n_d) e_M^b(n_d) \quad (7)$$

(c) Filtering part: Same as (2)-(3) above.

To define initial conditions for this algorithm, a soft constraint approach is described in [3]. Assuming that two distinct FAEST algorithms [5] are used in steps (a) and (b), the total complexity of FNTF is $2N + 12M$.

3. Weaver SSB subband structure

Subband adaptive filtering offers several advantages over a conventional full-band approach, including reduction of computational complexity and improved signal conditioning [2].

Design of analysis/synthesis (A/S) filter banks for a subband AEC system is a complex problem involving several tradeoffs/requirements: near-perfect reconstruction, low processing delays, oversampling in the subbands, low complexity. For ease of implementation, it is also desirable that the subband signals be real. Based on these considerations, we have found it convenient to use Weaver SSB A/S banks in our subband filtering structure; the latter is illustrated in Fig. 2.

The loudspeaker signal $u(n)$ and the microphone signal $d(n)$, with sampling rate F_s, are each decomposed into B real subband signals by analysis filter banks based on Weaver SSB modulators. Each bank consists of B band-pass filters followed by decimators by $K \leq B$. The corresponding subband signals are denoted by $u_b(m)$ and $d_b(m)$, where $b = 0, ..., B-1$ and m denotes sampling-time at the lower rate $F_s' = F_s/K$. In each subband, an FNTF adaptive algorithm operating at the reduced rate F_s' is used to identify the corresponding subband component of the echo path. The subband error signals $e_b(m)$ are finally recombined by a dual synthesis bank to produce a full-band error signal $e(n)$, at the original rate F_s.

In our implementation, the digital spectrum $[-\pi, \pi]$ is divided into B real subbands, with bandwidth $\omega_\delta = \pi/B$. The center frequency of the bth subband (positive sideband) is given by $\omega_b = (b + \frac{1}{2})\omega_\delta$, $b = 0, 1, ..., B-1$. Each narrowband filter in the analysis bank is a Weaver modulator with center frequency ω_b; a corresponding Weaver demodulator is used in the synthesis bank (see [6] for details). In theory, subband aliasing may be avoided with critical downsampling, i.e. $K = B$, provided an ideal low-pass filter $h(n)$ with cut-off $\omega_\delta/2$ is used. In practice, because of non-ideal filters, oversampling is necessary, i.e. $K < B$; we chose $K = B/2$. A window technique is used to design $h(n)$. For $B = 16$, the resulting A/S bank has the following properties: processing delay of 16ms, amplitude distorsion within $\pm 0.05dB$, linear phase.

4. Results and discussions

The sampling rate is set to F_s = 8kHz. Different loudspeaker signals $u(n)$ are used in the experiments. Here, we show results for the composite source signal (CSS), a speech-like signal [7]. The echo path $\mathcal{H}$ in Fig. 1 is simulated with a synthetic room impulse response of duration 2048 samples. To produce the microphone signal, this response is convolved with $u(n)$ and independent white noise is added. The convergence performance is evaluated in terms of the short term power of $e(n)$ (32ms window, dB relative to the echo level). The parameter values (N, M, λ) are the same for all the subband FNTFs. Following [4]: λ is set to a large value ($\lambda = 1 - 1/2N$), an acceleration mechanism is used in (3) and the filtering part of FNTF is initially frozen (for 0.5s). Note that for $M = 0$, FNTF corresponds to a modified form of NLMS, while in the case $M = N$, FNTF corresponds to FAEST (FRLS family).

Fig. 3 shows convergence curves of the full-band FNTF for $N = 2048$ (256ms), SNR= $30dB$ and different M. The top and bottom curve represent the microphone signal $d(n)$ and the additive noise. It is seen that a large value of M (≥ 256), is needed to obtain a performance comparable to RLS. Fig. 4 shows convergence curves of the subband FNTF for $N = 256$ (256ms), SNR= $30dB$ and different M. Again, a relatively large value of M (about 128), is required to achieve RLS-like performance, despite the fact that in subbands, the effective prediction order of $u(n)$ is smaller due to decimation.

Our study points to the following general conclusions: (1) the selection of the parameter M in FNTF can not be based strictly on the necessary AR modeling order of the source signal $u(n)$; (2) the use of subband processing is not effective in reducing the ratio M/N necessary for efficient operation of FNTF with long filters; (3) subband FNTF appears to be of limited practical value for AEC in teleconference applications.

References

[1] E. Hansler, *Signal Processing*, vol. 27, pp. 259-271, 1992.
[2] A. Gilloire, *Proc. IEEE Int. Conf. ASSP*, Dallas, Texas, 1987, pp. 2141-2144.
[3] G. V. Moustakides *et al.*, *IEEE Trans. on SP*, vol. 39, pp. 2184-2193, Oct. 1991.
[4] T. Petillon *et al.*, *IEEE Trans. on SP*, vol. 42, pp. 509-518, March 1994.
[5] G. Carayannis *et al.*, *IEEE Trans. ASSP*, vol. 31, pp. 1394-1402, Dec. 1983.
[6] R. E. Crochiere *et al.*, *Multirate Digital Signal Processing*, Prentice-Hall, Englewood Cliffs, NJ, 1983.
[7] H. Gierlich, *Signal Processing*, vol. 27, pp. 281-300, 1992

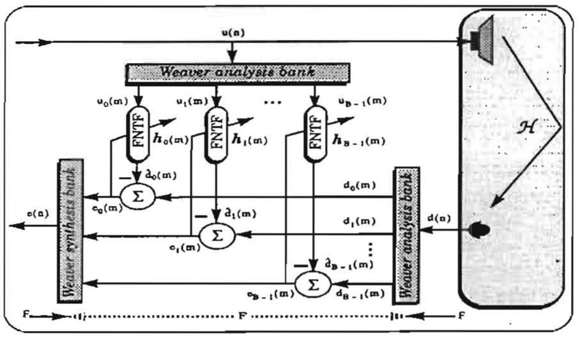

Figure 2: Weaver SSB subband structure for AEC.

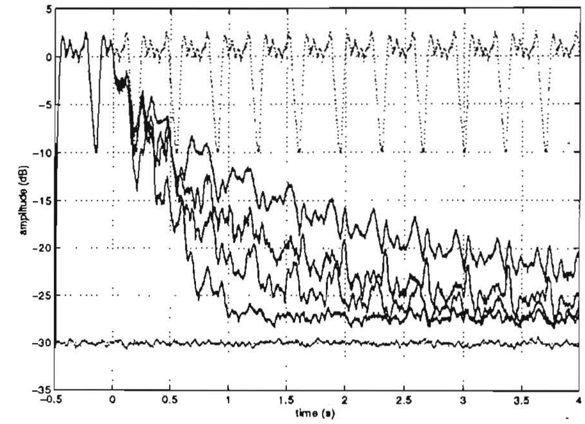

Figure 3: Convergence curves for fullband FNTF (top to bottom: echo; FNTF with M=0,16,256,2048; noise).

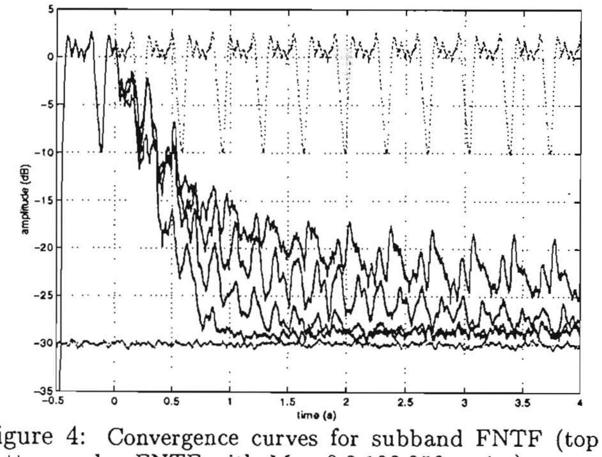

Figure 4: Convergence curves for subband FNTF (top to bottom: echo, FNTF with M = 0,8,128,256; noise).