A COMPARISON OF ADAPTIVE IIR ECHO CANCELLER HYBRIDS

Sharlene Gee* and Markus Rupp

Institut für Netzwerk und Signaltheorie, TH Darmstadt,
Merckstr. 25, D-6100 Darmstadt, Federal Republic of Germany

Abstract

In this paper a comparison of adaptive IIR filters with gradient based adaptation algorithms is presented. The classification structure employed clearly illustrates the relationships of the algorithms to each other; additionally, other new feasible filter methodologies for further investigation become revealed. All algorithms were implemented on a Motorola 56001. Correct normalization of the adaptation stepsize played a critical role in the results, which were obtained by real time measurements. Only the SHARF(Simple Hyperstable Adaptive Recursive Filter) and BRLMS(Bias Remedy LMS) algorithms fulfill the requirements of a low cost hybrid echo canceller.

1 Introduction

The disturbing phenomenon of hybrids in telephone networks is the occurence of unavoidable electric echoes. On the one hand, these echoes may disturb the local speaker; on the other hand, in long distance calls the energy of the echo may largely exceed the energy of the subscriber's signal and thus cause severe quantization problems with the A/D converter which follows the hybrid. Echo cancellation is desired to minimize the impact of imperfect conditions at the hybrid and to ensure a natural speaking environment. Since every new telephone connection changes the transfer function of any given echo path, the echo canceller must be adaptive. The least mean square (LMS) algorithm is commonly used together with a transversal filter because of its well-understood, favorable properties, e.g., guaranteed stability, an unimodal error surface structure, and easy implementation [1,2].

Linear electronic circuits, in this context, hybrids, are known to have impulse responses that are sums of exponentials, which suggests a recursive model [3]. With feedback it is possible to realize a long-duration impulse response with a lower order filter than if a transversal filter were used. This together with the appeal of a better, or even an exact, model has provided the impetus for exploring the possibilities of adaptive IIR filters.

During the last few years several ideas based on IIR filters were proposed as solutions to the echo problem. Of these algorithms, several have recently been implemented on a fixed point DSP and tested on the Motorola 56001.

*The author is now with Electronic Tools, Zum Blauen See 7, D-4030 Ratingen.

This paper compares the real-time behavior and emphasizes those algorithms with promising performances. In particular, the following algorithms were investigated: Series Parallel LMS (SP-LMS) [4], Equation Error Formulation (EEF-LMS) [5], Bias Remedy LMS (BR-LMS) [6], Alternate Filtering Mode (AFM) [7], Simple Hyperstable Adaptive Recursive Filter (SHARF) [8], and Normalized LMS (NLMS), which served as a FIR comparison case.

2 Gradient Based Algorithms

To easily illustrate the commonalities and differences of the various algorithms, the model in Figure 1 will be utilized.

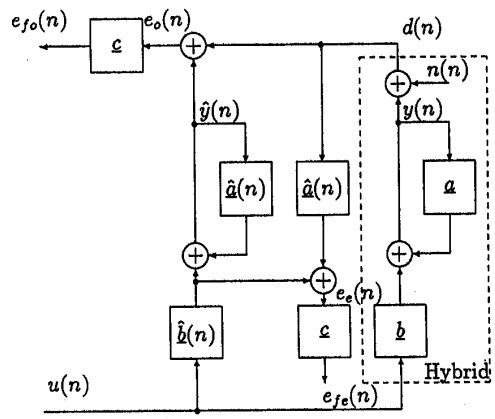

Figure 1: Adaptive filter structure for echo cancelling

For all algorithms two sets of coefficients, one for the transversal and one for the recursive part of the model, are used to estimate the echo path. In the case of NLMS, the set of recursive coefficients is empty. The investigated adaptation algorithms can all be expressed in the same manner. The estimated coefficients $\underline{\hat{w}}(n)$ are adapted in the direction of the negative error gradient that is proportional to a vector $\underline{\psi}(n)$:

CH2977-7/91/0000-1541 $1.00 © 1991 IEEE

$$\hat{\underline{w}}(n+1) = \hat{\underline{w}}(n) + \mu(n)e_a(n)\underline{\psi}(n) \tag{1}$$

where $\mu(n)$ is the stepsize and $e_a(n)$ is the adaptation error at time n, which will be described later in this section. The variables used in this paper are defined as follows:

$$\begin{aligned}
\underline{w}^T &= [a_1, ..., a_{M_a}, b_0, ..., b_{M_b}] && (2)\\
\hat{\underline{w}}^T(n) &= [\hat{a}_1(n), ..., \hat{a}_{M_a}(n), \hat{b}_0(n), ..., \hat{b}_{M_b}(n)] && (3)\\
\underline{x}^T(n) &= [y(n-1), ..., y(n-M_a), u(n), ..., u(n-M_b)] && (4)\\
\underline{x}_e^T(n) &= [d(n-1), ..., d(n-M_a), u(n), ..., u(n-M_b)] && (5)\\
\underline{x}_o^T(n) &= [\hat{y}(n-1), ..., \hat{y}(n-M_a), u(n), ..., u(n-M_b)] && (6)\\
\underline{x}_c^T(n) &= \underline{x}_e^T(n) - \tau(n)\underline{e}_o^T(n) && (7)\\
\underline{x}_{fo}(n) &= \underline{x}_o(n) + \sum_{i=1}^{M_a} \hat{a}_i(n)\underline{x}_{fo}(n-i) && (8)\\
\underline{x}_{fe}(n) &= \underline{x}_e(n) + \sum_{i=1}^{M_a} \hat{a}_i(n)\underline{x}_{fe}(n-i) && (9)\\
e_o(n) &= d(n) - \hat{\underline{w}}^T(n)\underline{x}_o(n) && (10)\\
d(n) &= y(n) + n(n) && (11)\\
\underline{e}_o^T(n) &= (e_o(n-1), ..., e_o(n-M_a), 0, ..., 0) && (12)\\
e_e(n) &= d(n) - \hat{\underline{w}}^T(n)\underline{x}_e(n) && (13)\\
e_{fo}(n) &= e_o(n) + \sum_{i=1}^{M_c} c_i e_o(n-i) && (14)\\
e_{fe}(n) &= e_e(n) + \sum_{i=1}^{M_c} c_i e_e(n-i) && (15)\\
\tau(n) &= \min\left(1, \frac{\sum_{k=1}^{M_a} d(n-k)^2}{\sum_{k=1}^{M_a} e_o(n-k)^2}\right) && (16)
\end{aligned}$$

As can be seen in Figure 1, $u(n)$ and $d(n)$ are the input and output signals, respectively, of the hybrid. The coefficient vectors $\hat{\underline{w}}(n)$ and $\underline{w}(n)$ each consist of $M_a + M_b + 1$ elements (M_a and $M_b + 1$ elements for the recursive and transversal parts, respectively). In [3] it is shown that $M_a = 3$ and $M_b = 8$ is a good choice, and these values were used in the algorithms described here except for the NLMS, for which $M_a = 0$ and $M_b = 31$ were used. The signal $n(n)$ stands for the subscriber's speech but can also be seen as a perturbation like noise, interchannel interference, or signalling impulses.

With this paper's notation, 20 different gradient based algorithm are possible, depending on gradient and adaptation error assumptions. Table 1 illustrates these possibilities and shows the algorithms already reported upon. Many new algorithms reveal themselves from this presentation. Feintuch's [9] and Stearns'-algorithms [10] are not further examined.

	gradient $\underline{\psi}(n)$				
e_a	x_o	x_e	x_{fo}	x_{fe}	x_c
e_o	Feintuch's		Stearns'	AFM	
e_e		SP-,EEF-LMS			BRLMS
e_{fo}	SHARF				
e_{fe}					

Table 1: Possibilities for gradient based algorithms.

The stepsize μ was replaced with a normalized stepsize $\alpha/\text{norm}(n)$ because of the following considerations:

1. The behavior of the algorithm should be independent of the input signal power. The expectation of $\underline{\psi}^T(n)\underline{\psi}(n)$ as well as of $e_a^2(n)$ are proportional to the input power. Therefore, every norm that is proportional to the input power can be used.
2. The stepsize of the various algorithms should be comparable.
3. Within the normalization an improvement of the algorithm performance is possible. For example, the NLMS algorithm is known to be superior in comparison to the unnormalized LMS -algorithm[11].

In general, it is not easy to give the best choice of the norm for every algorithm in Table 1. In some of the literature norms are stated; see, for example, those used for AFM [12] and BRLMS [6]. The norm for SP- and EEF-LMS uses the same justification as that for the NLMS algorithm. However, to be more general a derivation for two cases is given here: For the two algorithms in which the gradient and error product results in a $e_i(n)\underline{x}_i(n)$ term, for $i = \{o, e\}$, the norm is given by $\underline{x}_i^T(n)\underline{x}_i(n)$. The motivation for this choice is as follows. Suppose that the adaptation is done by:

$$\hat{\underline{w}}(n+1) = \hat{\underline{w}}(n) + \mu(n)[d(n) - \hat{\underline{w}}^T(n)\underline{x}_i(n)]\underline{x}_i(n) \tag{17}$$

With the error vector $\underline{\epsilon}(n) = \underline{w} - \hat{\underline{w}}(n)$,

$$\begin{aligned}
\underline{\epsilon}(n+1) &= \underline{\epsilon}(n) - \mu(n)\underline{x}_i(n)\underline{x}_i^T(n)\underline{\epsilon}(n)\\
&+ \mu(n)[\underline{w}^T(\underline{x}(n) - \underline{x}_i(n)) + n(n)]\underline{x}_i(n) && (18)\\
&= [I - \mu(n)\underline{x}_i(n)\underline{x}_i^T(n)]\underline{\epsilon}(n)\\
&+ \mu(n)[\underline{w}^T(\underline{x}(n) - \underline{x}_i(n)) + n(n)]\underline{x}_i(n). && (19)
\end{aligned}$$

If the norm for the stepsize $\mu(n)$ is chosen to be $\underline{x}_i^T(n)\underline{x}_i(n)$, the matrix $\frac{\underline{x}_i(n)\underline{x}_i^T(n)}{\underline{x}_i^T(n)\underline{x}_i(n)}$ is obtained in Eq. 19 and this matrix has the property that its eigenvalues equal either zero or one. For adaptation it is further assumed that reducing the error vector also decreases the difference between the input vector $\underline{x}(n)$ and the vector $\underline{x}_i(n)$. The square of the L_2-norm of the error vector is a Ljapunov function for a certain choice of the stepsize α. Because of the idempotence of the matrix $\frac{\underline{x}_i(n)\underline{x}_i^T(n)}{\underline{x}_i^T(n)\underline{x}_i(n)}$ the following results:

$$\begin{aligned}
\underline{\epsilon}^T(n+1)\underline{\epsilon}(n+1) &= \underline{\epsilon}^T(n)\left(I - \alpha(2-\alpha)\frac{\underline{x}_i(n)\underline{x}_i^T(n)}{\underline{x}_i^T(n)\underline{x}_i(n)}\right)\underline{\epsilon}(n)\\
&+ 2\underline{\epsilon}^T(n)\underline{x}_i(n)\alpha(1-\alpha)\frac{\Delta(n)}{\underline{x}_i^T(n)\underline{x}_i(n)}\\
&+ \alpha^2\frac{\Delta^2(n)}{\underline{x}_i^T(n)\underline{x}_i(n)} && (20)
\end{aligned}$$

with the abbreviation $\Delta(n) = \underline{w}^T(\underline{x}(n) - \underline{x}_i(n)) + n(n)$.

Taking the expectation of both sides of Eq. 20 leads to a simpler equation for which the linear term in $\underline{\epsilon}(n)$ can be

neglected. The expectation $E[\underline{\epsilon}^T(n)\underline{x}_i(n)] = E[e_i(n)]$ is zero, because it is supposed that the input sequence has zero mean. The matrix $E\left[\frac{\underline{x}_i(n)\underline{x}_i^T(n)}{\underline{x}_i^T(n)\underline{x}_i(n)}\right]$ is positive semidefinite with eigenvalues between zero and one. Therefore, for $v(n) = E[\underline{\epsilon}^T(n)\underline{\epsilon}(n)]$ the following inequality is obtained:

$$v(n+1) \leq (1-\alpha(2-\alpha))v(n) + \alpha^2 \frac{\Delta^2(n)}{\underline{x}_i^T(n)\underline{x}_i(n)} \quad (21)$$

which is true for $0 \leq \alpha \leq 2$. Since the term $\Delta^2(n)$ is also dependent on $v(n)$, stability is generally found for a stepsize α much lower than two.

The norm for SHARF is chosen equal to $\underline{x}_o^T(n)\underline{x}_o(n)$. Following the prove of stability in [13] it is possible to show that with this norm the algorithm also fulfils the hyperstability properties.

Not only the gradient and the adaptation error $e_a(n)$ differ amongst the several algorithms, but also the output signal $s_o(n)$, since it is chosen to equal either the output error $e_o(n)$ or the equation error $e_e(n)$. Table 2 expands upon Table 1 by further specifying the signal choices used for the output error and the norm.

3 Measurement Results

All these algorithms were implemented on the Motorola 56001 fixed-point DSP. In Table 3 the results of the measurements and the advantages/disadvantages of the described algorithms are listed. The measurements were made with a white random process as excitation. The differences resulting from using speech signals as excitation are discussed later. All real-time measurements contained a noise disturbance $n(n)$ at the location indicated in Figure 1. Usually the noise power was relatively low and influenced the steady state values. The case of subscriber's speech in $n(n)$ is discussed at the end of the section.

Echo Return Loss Enhancement (ERLE) is defined as $-10\log\frac{E[s_o^2(n)]}{E[d^2(n)]}$ in dB. Because of the difficulties in calculating expectation values in real time measurements, short-time averaged values were taken instead. The values in the ERLE column reflect the variation range of the steady state values when the algorithms were subjected to different connections, which were short ($100m$), long ($9km$), and in-house calls. In-house calls led to the best

Algorithm	$s_o(n)$	$e_a(n)$	$\underline{\psi}(n)$	norm(n)
NLMS	$e_e(n)$	$e_e(n)$	$\underline{x}_e(n)$	$\underline{x}_e^T(n)\underline{x}_e(n)$
SP-LMS	$e_e(n)$	$e_e(n)$	$\underline{x}_e(n)$	$\underline{x}_e^T(n)\underline{x}_e(n)$
EEF-LMS	$e_o(n)$	$e_e(n)$	$\underline{x}_e(n)$	$\underline{x}_e^T(n)\underline{x}_e(n)$
BRLMS	$e_o(n)$	$e_o(n)$	$\underline{x}_e(n) - \tau(n)\underline{e}_o(n)$	$(1+2\tau(n))\,\underline{x}_e^T(n)\underline{x}_e(n)$
AFM	$e_o(n)$	$e_o(n)$	$\underline{x}_{fe}(n)$	$\underline{x}_{fo}^T(n)\underline{x}_{fo}(n)$
SHARF	$e_o(n)$	$e_{fo}(n)$	$\underline{x}_o(n)$	$\underline{x}_o^T(n)\underline{x}_o(n)$

Table 2: The algorithms as defined by their particular variable choices.

Algorithm	Time	α	bias	convergence	stability	ERLE	remarks
NLMS	$< 100ms$	1	no	yes	sure	$12-26dB$	
SP-LMS	$< 100ms$	1	yes	yes	reached	$14-29dB$	
EEF-LMS	$< 100ms$	1	yes	yes	reached	$7-22dB$	
BRLMS	$< 100ms$	0.125	no	yes	reached	$9-21dB$	
AFM	$< 1s$	0.01	no	no	reached only with low noise	$3-23dB$	often instable
SHARF	$< 100ms$	1	no	yes	reached	$8-23dB$	difficult to specify c_i

Table 3: Algorithm properties and implementation results.

results because of correct impedances on the subscriber side.

Adaptation time is specified as the time from the beginning of the adaptation until the output error signal $s_o(n)$ has reached its "steady state" value. The adaptation time depends strongly upon the choice of the constant α. In the cases of BRLMS and AFM a low stepsize is necessary to avoid instability. The adaptation time of the BRLMS algorithm with speech as excitation is a few hundred milliseconds; that of AFM is even slower because an even smaller α must be used to avoid instability.

LMS based algorithms for IIR filters always have a bias in the estimated coefficients [5], which can lead to instability; consequently, only bias-free methods are further compared.

In the above table convergence and stability refer to different conditions. Convergence implies that the algorithm ensures that the coefficients move to a fixed point in the parameter space, whereas stability specifies that this point is a stable filter point. In the SHARF algorithm only one coefficent c_i for the output error filtering was used. The size of this coefficient was found heuristically and cannot be proven to be appropriate for all hybrid wire combinations. The BRLMS and SHARF algorithms show only infrequently instable behavior, which can yet be decreased by two supplements to the algorithms.

Because of parameter drift [14], it is useful to have a leakage compensation in the algorithms. Furthermore, a double talk detector is implemented to set the stepsize α to zero when the subscriber is speaking. All algorithms were tested, when the subscriber was speaking. With an double talk detector the adaptation time can substantially increase, because adaptation is performed only during breaks in the subscriber's speech. However, the steady state ERLE values, once reached, typically decrease less than one dB or two. With these two additions the behavior of the algorithms is much improved, such that additional monitoring to guarantee stability is no longer necessary.

In the case of AFM, which in spite of a low stepsize, often tends to become instable, using leakage compensation and a double talk detector slightly improves the behavior but still does not eliminate the instable tendency. Monitoring, however, is not an attractive solution because the different corrections often corrupt the speech signal to an audibly disturbing degree.

4 Conclusion

In conclusion, together with coefficient leakage compensation and a double talk detector, the BRLMS and SHARF algorithms are the two most stable of the implemented IIR algorithms. With the additional advantages of bias-free coefficients and ERLE in the vicinity of $22dB$, these two algorithms appear to be the most appropriate for compensation of hybrid echoes.

References

[1] Man Mohan Sondhi, David A. Berkley, "Silencing Echoes on the Telephone Network," Proc. IEEE, vol. 68, no. 8, pp. 948-963, Aug. 1980.

[2] Simon Haykin, "Adaptive Filter Theory," Prentice Hall, Englewood Cliffs, N.J., 1986.

[3] Markus Rupp, "Adaptive IIR Echo Cancellers for Hybrids using the Motorola 56001," in Proc. Eusipco Signal Processing V, Barcelona, pp. 1487-1490, 1990.

[4] Maurice G. Bellanger, "Adaptive Digital Filters and Signal Analysis," Marcel Dekker, Inc./ New York Basel, 1986.

[5] John J. Shynk, "Adaptive IIR Filtering," IEEE ASSP Magazine, vol. 6, no. 2, Apr. 1989.

[6] Ji-Nan Lin, Rolf Unbehauen, "Modification of the Least Mean Square Equation Error Algorithm for IIR System Identification and Adaptive Filtering," in Proc. Int. Symp. Circuits and Systems 1990, New Orleans, pp. 3150-3153, 1990.

[7] Hong Fan, "A New Adaptive IIR Filter," IEEE Trans. Circuits Systems, vol. CAS-33, no. 10, pp. 939-947, Oct. 1986.

[8] Michael G. Larimore , John R. Treichler, C. Richard Johnson, "SHARF: An Algorithm for Adapting IIR Digital Filters," IEEE Trans. Acoust., Speech, Sig. Proc., vol. ASSP-28, no. 4, pp. 428-440, Aug. 1980.

[9] P. L. Feintuch, "An Adaptive Recursive LMS Filter," Proc. IEEE, vol. 64, no. 11, pp. 1622-1624, Nov. 1976.

[10] Kai. X. Miao, Hong Fan, "Transforming Sharf etc. into Lattice," in Proc. Int. Symp. Circuits and Systems 1990, New Orleans, pp. 796-800, Mai 1990.

[11] Moshe Tarrab, Arie Feuer, "Convergence and Performance Analysis of the Normalized LMS Algorithm with Uncorrelated Gaussian Data," IEEE Trans. Inform. Theory, vol. IT-34, no. 4, pp. 680-691, July 1988.

[12] Hong Fan, W. Kenneth Jenkins, "An Investigation of an Adaptive IIR Echo Canceller: Advantages and Problems," IEEE Trans. Acoust., Speech, Sig. Proc., vol. ASSP-36, no. 12, pp. 1819-1833, Dec. 1988.

[13] C. Richard Johnson, Jr., "A Convergence Proof for a Hyperstable Adaptive Recursive Filter," IEEE Trans. Inform. Theory, vol. IT-25, pp. 745-749, Nov. 1979.

[14] William A. Sethares, Dale A. Lawrence, C. Richard Johnson, Robert A. Bitmead, "Parameter Drift in LMS Adaptive Filters," IEEE Trans. Acoust., Speech, Sig. Proc., vol. ASSP-34, no. 4, pp. 868-879, Aug. 1986.