A Fast Echo Canceller Initialization Method for the CCITT V.32 Modem

JOHN M. CIOFFI, MEMBER, IEEE

Abstract—**The single most critical subsystem in the CCITT V.32 9600 b/s full-duplex modem is the echo canceller, which must achieve in excess of 60 dB near-end-echo rejection and 25 dB far-end-echo rejection. An important issue in the design of such an echo canceller is the convergence time required to achieve these stringent levels of cancellation. While stochastic-gradient or LMS algorithms should be used for tracking echo variation in the steady-state (after convergence) mode of operation, their use during convergence can result in an unacceptably long training, and especially retraining, period (maximum of 18 s, nominally about 15 s).**

This paper investigates the use of a fast-converging frequency-domain adjustment method for the echo canceller that conforms to the CCITT V.32 standard. The method requires no additional processor cycles (or cost increase) over that already required during steady-state operation of the V.32 modem, and reduces the training period to a maximum of 7.5 s, and nominally about 2.5 s. The presence of intermediate echos can be detected naturally with the presented methods. Furthermore, simple modifications of the new method are also introduced to accommodate the special situations where intermediate echos and/or frequency offset are present.

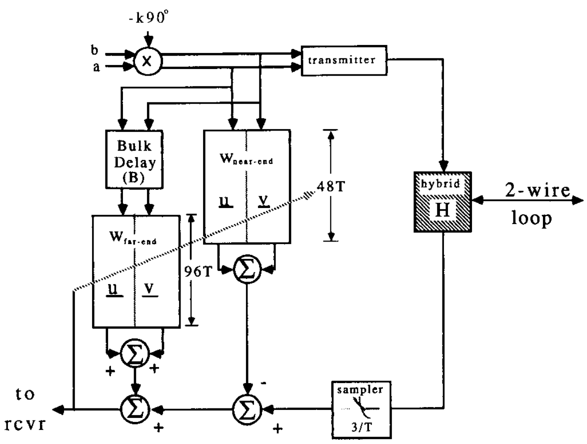


Fig. 1. A V.32 echo data echo canceller.

I. Introduction

FULL-DUPLEX voiceband modems for data rates at or above 4800 b/s require internal echo cancellers for operation over the two-wire General Switched Telephone Network. These echo cancellers must achieve a high level of cancellation. In particular, echo cancellers in modems conforming to the 9600 b/s CCITT V.32 (or the 14400 b/s V.32 bis) standard [1] must achieve in excess of 60 dB of near-end-echo cancellation under worst case conditions. An important issue in the design of the corresponding echo cancellers is the convergence time required to achieve this stringent level of cancellation.

While stochastic-gradient or LMS algorithms should be used for tracking echo variation in the steady-state (after convergence) mode of operation, their use during convergence can result in unacceptably long training, and especially retraining, periods (about a maximum of 18 s, nominally 15 s). Faster converging techniques have been introduced by Salz [2], Honig [3], and Cioffi [4] for the generic full-duplex echo canceller. Nevertheless, in applying these techniques specifically to the V.32 (or the more recent V.32 bis) modem, the designer discovers some serious drawbacks in the form of unacceptably high computation and/or high memory requirements when, as is usual, there is a far-end echo. This paper investigates computationally efficient, fast-converging echo cancellers that are specifically designed for the protocol of the CCITT V.32 modem (with or without trellis coded modulation), and which can readily accommodate the usual situation when a far-end echo is present. In particular, the introduced methods require no increase in computational cost with respect to what must have already been provided for steady-state operation of the V.32 modem. Also, the echo canceller converges[1] in a maximum of 7.5 s and a nominal time of about 2.5 s). The maximum occurs with a satellite channel and 600 ms of bulk delay in the far end. For a nominal 60 ms delay in terrestrial connections, we get the second lower startup time.

The structure chosen for the implementation of the echo canceller is the so-called "Nyquist" in-band canceller, described by Werner [5] and illustrated in Fig. 1. This structure has been demonstrated to have reduced computational and data-memory requirements [5] and better numerical properties [6] than the "transmitter-driven" echo cancellers that are fed with the transmitter output rather than the actual data symbols (rotated transmitter input). In the configuration of Fig. 1, we choose the near-end cancellers to span 20 ms or 48 symbol periods at 2400 baud, and the far-end cancellers to span 40 ms or 96 symbol periods. The bulk delay will typically span up to a maximum of 600 ms or 1440 symbol periods. The canceller under study will be realized as a $T/3$ structure with a sampling rate of 7200 Hz, requiring three 48-tap in-phase near-end subcancellers, three 48-tap quadrature near-end subcancellers, three 96-tap in-phase far-end subcancellers, and three 96-tap quadrature far-end subcancellers. A frequency-domain method, which uses a variable-length-period training sequence, is presented. Basically, the silence of the double talker during V.32 echo canceller training is exploited through the use of frequency-domain adaptive filters, which converge rapidly under this condition. Additionally, the presented methods are numerically stable. We also present minor additions for the cancellation of intermediate echos and for fast acquisition of any echo frequency offset (up to the 14 Hz that would occur if the V.32 maximum of 7 Hz accumulated along the echo return path).

Section II briefly reviews the necessary mathematics for description of the echo canceller configuration in Fig. 1, and then discusses frequency-domain training of the echo canceller with periodic training sequences. Section III is a more detailed investigation of the

Paper approved by the Editor for Channel Equalization of the IEEE Communications Society. Manuscript received February 1, 1988; revised November 28, 1988 and January 23, 1989. This work was supported in part by an IBM Faculty Development Award and in part by the National Science Foundation under Contract MIP 86-57266. This paper was presented at IEEE GLOBECOM'87, Tokyo, Japan, November 1987.

The author is with the Information Systems Laboratory, Department of Electrical Engineering, Stanford University, Stanford, CA 94305.

IEEE Log Number 9035431.

[1] Here, we mean the total elapsed time before useful data can be reliably transmitted, and not just the convergence time for a single adaptive echo canceller, which would be about 20–40 ms in the presented methods.

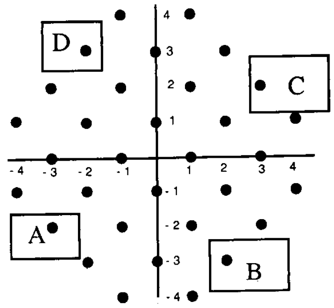


Fig. 2. V.32 EC nominal inputs. Boxed points $ABCD$ only used for training.

frequency-domain training for the V.32 echo canceller. For simplicity, discussion of intermediate echos and frequency-offset tracking is intentionally averated in these first three sections. Nevertheless, Section IV then builds upon the existing structure, with relatively simple modifications, that generalize the presented methods to cancel intermediate echos, as well as to track frequency offset in far or intermediate echos. Section V is a brief summary and conclusion.

II. Echo Canceller Architecture and Algorithms

This section briefly reviews the $T/3$ Nyquist in-band echo canceller for steady-state cancellation of the echo, and then proceeds to introduce further simplifications that occur only during the special training and retraining periods provided in the CCITT V.32 protocol.

Section II-A reviews the structure and introduces definitions. Section II-B investigates the use of frequency-domain computation of the initial echo canceller settings with periodic training sequences. Section II-C examines the effects of rotation (of the signal constellation) on the settings for the canceller responses, dubbed here "canceller rotation."

A. T/3 Canceller Architecture

The hybrid output signal in Fig. 1 can be expressed as

$$s(t) = \text{Re}\left[\sum_m A_m h_{bb}(t - mT)e^{j\omega_c t}\right] + n(t) \qquad (1)$$

where A_m is the complex data signal

$$A_m = a_m + jb_m \qquad (2)$$

formed by taking a_m, the real part, as the horizontal component of a signal chosen from the CCITT signal set shown in Fig. 2, and b_m, the imaginary part, as the vertical component. $h_{bb}(t)$ is the baseband-equivalent pulse response (at $1/T = 2400$) [7] of the combined echo-pulse shaping path or, equivalently, the pulse response of the combined pulse-shaping-echo path is

$$h(t) = \text{Re}\,[h_{bb}(t)e^{j\omega_c t}]. \qquad (3)$$

Also in (1), $n(t)$ is a noise signal that can consist of channel noise, double talk, and other nonideal channel impairments, $\omega_c = 2\pi 1800$ is the radian carrier frequency, and t denotes the continuous-time variable. The signals in Fig. 2 need not be used in *training* the echo canceller, as it will actually be more convenient to sometimes zero either the real or imaginary components, as is the case in the particular V.32 method described in Section III.

To see the equivalence of (1) and Fig. 1, we write

$$s(t) = \text{Re}\left\{\sum_m A_m h_{bb}(t - mT)e^{j\omega_c(t - mT + mT)}\right\} + n(t) \qquad (4)$$

or

$$s(t) = \text{Re}\left\{\sum_m (A_m e^{j\omega_c mT})h_{bb}(t - mT)e^{j\omega_c(t - mT)}\right\} + n(t) \qquad (5)$$

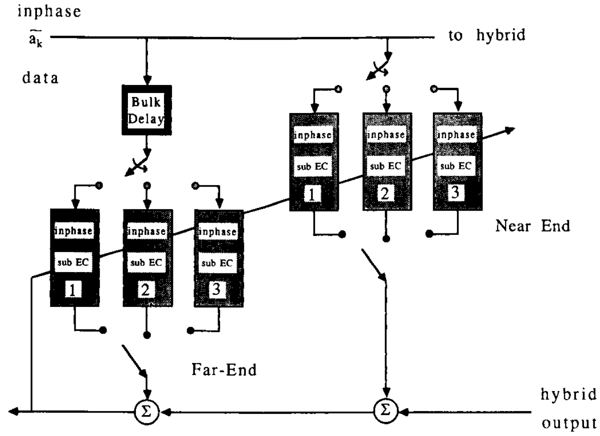


Fig. 3. In-phase echo canceller block diagram (polyphase "subcanceller" implementation").

which can be reduced to

$$s(t) = \text{Re}\left\{\sum_k \tilde{A}_m H(t - mT)\right\} + n(t) \qquad (6)$$

where

$$\tilde{A}_m = A_m e^{j\omega_c mT} \qquad (7)$$

and

$$H(t) = h_{bb}(t)e^{j\omega_c t}. \qquad (8)$$

Note that $H(t)$ can also be written in terms of $h(t)$ in (3) and $\hat{h}(t)$ [the Hilbert transform of $h(t)$] as

$$H(t) = h(t) + j\hat{h}(t). \qquad (9)$$

Thus, $s(t)$ becomes

$$s(t) = \sum_m [\tilde{a}_m h(t - mT) - \tilde{b}_m \hat{h}(t - mT)] + n(t) \qquad (10)$$

which can be estimated using only two filters at rate $3/T = 7200$ Hz (a rate higher than twice the highest passband frequency) in parallel on the rotated in-phase and quadrature inputs $\tilde{a}_m$ and $\tilde{b}_m$, respectively. The two outputs are summed to form an estimate of the echo, which is then subtracted from the received signal. There is no need for a Hilbert transform or phase splitter in the $T/3$ canceller nor is there need for a modulator or demodulator, other than the trivial rotations in (7), trivial as $\omega_c T = 3\pi/2$. A discrete form of the Hilbert transform, however, will be used in the presented fast start, as $\hat{h}$ will be computed from the estimate of h to save computation and training time. In the steady-state (stochastic-gradient algorithm in use) mode of echo canceller operation, it is actually more computationally efficient to estimate $\hat{h}(t)$ as a separate adaptive filter than it is to compute it (via the Hilbert transform) from an estimate of $h(t)$.

Further reductions in computation are possible using a performance- and mathematically equivalent subcanceller configuration [8] (actually, a special case of the polyphase systems in [9]), as illustrated in Fig. 3. If $s(t)$ is sampled at rate $1/T' = 3/T$, we can write

$$s(kT') = \sum_m [\tilde{a}_m h(kT' - mT) - \tilde{b}_m \hat{h}(kT' - mT)] + n(kT') \qquad (11)$$

or by letting $kT' = lT + pT'$ where $p = 0$, 1, or 2,

$$s(lT + pT') = \sum_m [\tilde{a}_m h(lT - mT + pT') - \tilde{b}_m \hat{h}(lT - mT + pT')] + n(lT + pT'). \qquad (12)$$

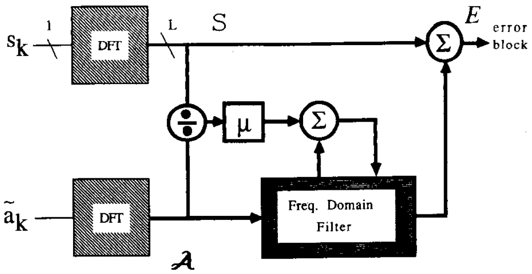


Fig. 4. Fast starting frequency-domain echo canceller (for periodic training sequence, with no double talk).

In (12), there are three phases (one for each value of p) of $s(kT')$ for each symbol interval T, each of which can be construed as a convolution of the same input with a subecho channel indexed by the corresponding value of p. The echo estimate can be similarly configured, leaving only one in-phase and one quadrature subcanceller to be implemented and updated at each sampling instant ($T/3$), resulting in a factor of 3 reduction in the computation used to implement the echo canceller, and a factor of 25% less computation than the fashionable $T/2$ cross-coupled echo canceller.

After the training of the echo canceller to be later described, the echo canceller coefficients, call them

$$W_{p,k} = u_{p,k} + jv_{p,k} \tag{13}$$

for the pth subcanceller, updated (on phase p of T) according to the stochastic-gradient or LMS algorithm:

$$\epsilon_{p,k} = s_{p,k} - u'_{p,k}\tilde{a}_k - v'_{p,k}\tilde{b}_k \tag{14}$$

$$W_{p,k+1} = W_{p,k} + \mu\epsilon_{p,k}A_k^* \tag{15}$$

where $\tilde{a}_k$ and $\tilde{b}_k$ are the vectors of inputs corresponding to the echo canceller coefficients, * denotes complex conjugate, and $'$ denotes transpose. Thus, the echo canceller continues to adapt throughout the entire call, permitting slow variation of the echo channel over the course of the call. Nevertheless, the stochastic-gradient method in (14) and (15) is notorious for its slow convergence. Its use is not desirable during echo canceller training, motivating the work of this paper.

B. Frequency-Domain Training Algorithm

Fig. 4 illustrates the basic principle of the frequency-domain method. The discrete Fourier transform (DFT) of the hybrid output is divided by the DFT of the known periodic training sequence to get the subcanceller tap settings in the frequency domain. An inverse DFT (IDFT) can then be performed to get the time-domain settings of the in-phase canceller. This procedure can be repeated for each of the three in-phase subcancellers. The settings for the quadrature cancellers are found by first interleaving the in-phase subcanceller settings in the appropriate order (actually, summing their DFT's with appropriate rotations) and performing a digital Hilbert transform. The rotated quadrature input $\tilde{b}_m$ is set to zero (equivalently, A_m is chosen such that $\tilde{A}_m$ has no imaginary part).

More precisely, we define the DFT's

$$\mathcal{A}_k = \frac{1}{\sqrt{L}}\sum_{n=0}^{L-1}\tilde{a}_n e^{-(j2\pi nk/L)}, \qquad k = 0, \cdots, L-1 \tag{16}$$

and

$$\mathcal{S}_{p,k} = \frac{1}{\sqrt{L}}\sum_{n=0}^{L-1}s(nT + pT')e^{-(j2\pi nk/L)},$$
$$p = 0, 1, 2, \quad k = 0, \cdots, L-1. \tag{17}$$

There are L possible values of the DFT's in (16) and (17) corresponding to the L possible values of k per block of L data points. In order for overlap effects to be avoided, both $\tilde{a}_n$ and $s_{p,n} = s(nT + pT')$ should be periodic, with period L. The periodicity of $\tilde{a}_n$ is ensured by design of the training sequence. The periodicity of $s_{p,n}$ is very closely approximated because of the very small noise added to the echo when the double talker is silenced during echo canceller training, along with the periodicity of the echo channel input $\tilde{a}_n$.

The vector of echo canceller coefficients ($\mathcal{W}_{p,k}$ in the frequency domain (and $W_{p,k}$ in the time domain) can be estimated for a single block of data as

$$\mathcal{W}_{p,k} = \frac{\mathcal{S}_{p,k}}{\mathcal{A}_k}, \tag{18}$$

while a recursive estimate is given by

$$\mathcal{W}_{p,k+(i+1)L} = (1-\mu_k)\mathcal{W}_{p,k+iL} + \mu_k\frac{\mathcal{S}_{p,k+(i+1)L}}{\mathcal{A}_k} \tag{19}$$

where $\mathcal{W}_{p,k+iL}$ is the estimate of the echo response in the kth frequency "bin" f the ith data block for the pth in-phase subcanceller. μ_k is a step size and $\mathcal{E}_{p,k} = \mathcal{S}_{p,k} - \mathcal{W}_{p,k}\mathcal{A}_k$. Note in (18) and (19) that the factors $1/\sqrt{L}$ effectively cancel and never need be computed. Tacit in (19) is that $\mathcal{A}_k$ is nonzero for all k. If one such value is zero, $\mathcal{W}_{p,k}$ should also be zeroed. Judicious choice of the training sequence, as described later, avoids this choice.

To perform the Hilbert transform necessary for the quadrature echo canceller coefficients, first form the subcanceller DFT at the higher sampling rate $3/T(=7200$ Hz). This higher rate DFT has three times as many points ($3L$), but they are the original length-L DFT points (for rate $1/T = 2400$) repeated successively three times. We call this interpolated DFT $\mathcal{W}_k$ and it has $3L$ points given by

$$\mathcal{W}_k = \mathcal{W}_{0,k} + \mathcal{W}_{1,k}e^{-j(2\pi k/3L)} + \mathcal{W}_{2,k}e^{-j(4\pi k/3L)} \tag{20}$$

where the DFT's $\mathcal{W}_{p,k}$ have been triplicated for the sum in (20). The echo canceller settings are then obtained by inverse DFT on $\mathcal{W}_k$ to get the in-phase coefficients and by inverse transformation on $-j\,\mathrm{sgn}(k - 3L/2)\mathcal{W}_k$ to get the quadrature coefficients. Actually, both IDFT's need not be computed, as separating the frequency-domain coefficients into "positive" and "negative" components of the IDFT sum will allow the computation to proceed with one IDFT and $3L$ additional additions/subtractions, as discussed in Section III-E. The subcanceller settings are obtained by taking the three sets of every third taps.

The reader may note that the estimate in (18) is equivalent to determining $\mathcal{W}_{p,k}$ such that it minimizes the frequency-domain sum of squared errors

$$\sum_{k=0}^{L-1}|\mathcal{S}_{p,k} - \mathcal{W}_{p,k}\mathcal{A}_k|^2, \tag{21}$$

while that in (19) is equivalent to a sum of criteria similar to (21). Because the DFT (IDFT) is an orthogonal transformation on the block of a sequence (DFT of sequence), the frequency-domain sum in (21) is equivalent to a time-domain sum of squared filter outputs (only strictly true when both sequences are periodic, which is true to a high degree of approximation in this situation). Thus, the frequency-domain method, for the special case of a periodic input and silenced double talker, is performance equivalent to the very fast converging recursive least squares (RLS) methods of [4], [2], and [3]. Because the number of parameters is equal to L, the block length, there is no noise averaging over one block, so at least two blocks must be processed according to (19) to get a good estimate. In fact, applying formulas from any of the three RLS references above, one will determine that the mean-square error is guaranteed to be within

$$\left(1 + \frac{L}{2L}\right)\text{MMSE} = \frac{3}{2}\text{MMSE} \tag{22}$$

or 1.8 dB of the best achievable cancellation after two length L estimates are averaged ($\mu = 1/2$ in (19) and in Fig. 4).

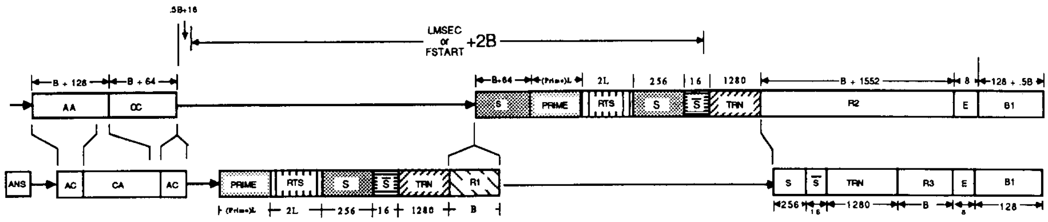


Fig. 5. Training protocol. All delays are in units of symbol periods, $T = 1/2400$.

At this point, stochastic-gradient methods are as effective as frequency-domain or RLS methods in providing the remaining 1.8 dB in the echo cancellation dynamic range. The huge immediate gain is due to the very low noise in estimating the echo when the double talker is silenced. If the double talker were present, the MMSE in (22) would be this double talk power, and the echo canceller would be about (25 + 1.8) dB away from best cancellation, and would require much longer time to further average the echo canceller coefficient estimates to a reasonable level in either the RLS, frequency-domain, or stochastic-gradient methods.

While in principle simple and in performance excellent, this frequency-domain method needs some precaution in execution because of the many possible positions of far-end echo with respect to near-end echo, and also the possibility of intermediate echos. Sections III and IV address these important details.

C. Canceller Rotation

In the signal constellation of Fig. 2, suppose we rotate the real and imaginary axes by some fixed angle θ. Then

$$\bar{A}_k = A_k e^{j\theta} \tag{23}$$

and, thus,

$$\tilde{\bar{A}}_k = \tilde{A}_k e^{j\theta}. \tag{24}$$

Then, in order to maintain the same echo $s(t)$,

$$\bar{H}(t) = H(t)e^{-j\theta}, \tag{25}$$

that is, the echo canceller settings are rotated by the same angle in the opposite direction.

During the initial acquisition of frequency offsets in intermediate and far-end echos (as discussed in Section IV), we will choose θ to be such that the signals A and C fall on the real axis, which corresponds to (see Fig. 2)

$$\theta = -\arctan \frac{1}{4} = -14.04°. \tag{26}$$

The echo canceller can be returned (although this is not necessary) to the use of $\theta = 0$ settings and can be adjusted by "derotating" the $+14°$ in the echo canceller settings. This is achieved by

$$h(t) = \hat{\bar{h}}(t) \cos \theta - \hat{\bar{h}}(t) \sin \theta \tag{27}$$

and

$$\hat{h}(t) = \bar{h}(t) \sin \theta + \bar{h}(t) \cos \theta. \tag{28}$$

Also, if ± 1 are the levels in the rotated input $\bar{A}_k$ and $\tilde{\bar{A}}_k$, then $h(t)$ and $\hat{h}(t)$ must be scaled down by $\sqrt{17}$ after derotation of the canceller settings if such derotation were used. In our presented method, this derotation is never necessary.

III. Frequency-Domain Training for the V.32 Echo Canceller

In this section, frequency-domain training of the V.32 echo canceller is investigated in detail. Section III-A discusses the bulk delay and its effect upon the general frequency-domain training that was described in Section II-B. Section III-B details the structure of the periodic training sequence whose period is a function of the bulk-delay estimate. Section III-C lists the exact recommended protocol and sequence of events in training the echo cancellers. Section III-D details the computational requirements. Intermediate echos and frequency offset in the echos are assumed absent throughout this section; however, they are incorporated through the modifications discussed in Section IV.

A. Bulk Delay

The bulk delay B is the elapsed time in symbol periods between transmission of data and receipt of the corresponding far-end echo.

The CCITT V.32 protocol provides for a crude estimate of the bulk-delay associated with the far-end echo only. Basically, the call modem transmits at 1800 Hz and the answer modem at 600 Hz,[2] with the answer modem starting a counter while reversing the phase of the transmitted 600 Hz square wave near the beginning of the protocol. The call modem detects the phase reversal, then reverses its own phase, and starts its own counter. The answer modem detects the phase reversal in the 1800 Hz signal, stops its counter with the present count equal to the bulk delay estimate B, and again reverses phase on the 600 Hz signal. The call modem detects the final 600 Hz phase reversal, and then freezes its counter with the same bulk-delay estimate ($\pm 4T$). For the far-end bulk-delay estimate to be accurate in this scheme, the far-end hybrid should be colocated with the far-end data set. Intermediate echos are not detected using such a technique.

Here, we assume that the CCITT bulk-delay estimate is valid, and we describe the effect of this estimate on the frequency-domain echo canceller initialization.

B. V.32-Compatible Echo Canceller Training Signal

The relevant portion of the CCITT V.32 initialization and retraining protocols is illustrated in Fig. 5. The optional echo canceller training sequence is inserted just before the "dotting" S and $\bar{S}$ patterns.

The in-phase and quadrature echo canceller settings are related through a Hilbert transform. Initially (see Section II-B), we need only estimate one and then compute the other. Thus, we excite only the in-phase echo channel for the initialization ($\tilde{b}_k = 0$), which keeps the length of the added training sequence to a minimum. Because of the rotation of successive input symbols to the data echo canceller, as given in (7), the in-phase subcancellers will alone be active if we ensure that after rotation, only the in-phase subcanceller inputs are nonzero. This is easily implemented, as an echo canceller implemented by the methods in Section II-A is usually accompanied by a similar "subtransmitter" implementation which also uses the rotated data signal as an input. For the special training sequence, we bypass the rotators employed in both and insert data into only the in-phase subcancellers and in-phase subtransmitters (the far-end modem is dormant at this time). Furthermore, only two bipolar values are used in the trainlng sequence. After $h(kT')$ has been estimated with W_k, $\hat{W}_k$ is computed as the Hilbert transform of W_k (although, as we see later, this transform is performed in the frequency domain directly).

Basically, the data sequence that is inserted should have a reasonably flat spectral density, although the degree of flatness is not critical to performance as long as obvious poor choices like dc or sinusoids are avoided. Furthermore, it is best to choose a sequence such that no matter what the value of B, there are no nulls of $\mathfrak{a}_l$

[2] 3000 Hz is also permitted in the CCITT protocol with the 600 Hz tone.

at any value of k. Also, the designer should recall that if there are a few bad values of B for a particular training sequence, B can be incremented or decremented by a few data symbol periods without any loss in performance of the echo canceller as it is sufficiently long that the bulk delay need only be approximately determined. The choice of training sequence and its relation to bulk delay B has recently been investigated in greater detail by Long and Ling of Codex [11]. For an improvement on the technique here, which can lead to an appearance of overlapping echos, the reader should also see [11] and [12].

Since the far-end bulk delay is known, as B symbol periods, the inserted training sequence must be applied to the channel for at least time B to ensure that remnant far-end echos have decayed prior to the first echo canceller updating. The period of the training sequence, we call it L here, must satisfy the relation

$$L = 3N + (B - N)_{\mathrm{mod}\,(3N)} = N_n + N_f + (B - N)_{\mathrm{mod}\,(N_n + N_f)} \tag{29}$$

where (in the general case) N_n is the length of the near-end canceller and N_f is the length of the far-end canceller. When N is used in our calculations, we assume that $N_n = N$ and $N_f = 2N$. The latter term in (29) is the residual of $B - N$ (or $B - N_n$) modulo $3N$ (or $N_n + N_f$). Accordingly, the maximum length sequence is $6N - 1$ (or $2N_n + 4N_f - 1$), while the minimum is $3N$ (or $N_n + N_f$) [we set the second term to zero if $B < N$ (or $B < N_n$)]. The reason for the dependency on B is that it is desirable to simultaneously estimate the near- and far-end settings, and the choice of L in (29) ensures that the concatenated inputs (with the $B - N$ (or $B - N_n$) intervening quiesent data symbols removed) appear as a cyclic sequence. This periodicity facilitates the application of DFT adaptive methods simultaneously to near-end and far-end echos since the efficient implementation of the DFT depends heavily on the periodicity of the input sequence. For $N = 48$, choosing L successive symbols from the beginning of a length-511 pseudorandom sequence, and then recycling (the sequence is truncated into a new length-L period sequence) those L into the channel satisfies the above objective of reasonable spectral flatness. According to G. Long of Northeastern University, there are a few values of the bulk delay that will produce some zero coefficients in the transform of the input data sequence. These points can be circumvented by studying the training sequence in advance, and knowing the corresponding values of B. Then B is incremented or decremented to the nearest reasonable value without performance loss in the echo canceller, as it is presumed to have a sufficient number of parameters that minor estimation errors in the bulk delay are insignificant. More generally, the maximum length is $2(N_n + N_f) - 1$.

The frequency-domain solution can be equivalently written in the time domain as the solution of the set of linear equations (for one near-plus-far in-phase subcanceller)

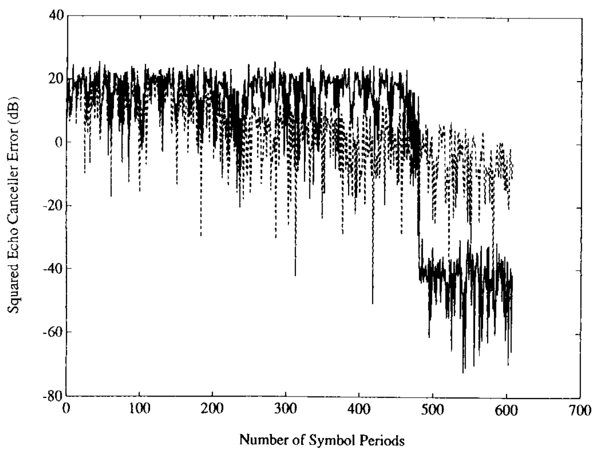


Fig. 6. Convergence comparison for V.32 EC.

period must be dynamically determined as a function of the bulk-delay estimate B, as given in (29).

Note that because of the subcanceller configuration discussed in the Introduction of this paper, the DFT of the length-L input sequence need only be computed once for the entire training period as it is common to all subcancellers and both of the successive blocks used to obtain the near-end, in-phase echo estimate. Also, the IDFT for the canceller time-domain settings need only be computed once upon completion of the fast start after two length-L blocks have been processed. Furthermore, enough length-L cycles of the training sequence must be inserted prior to the blocks used in the computations to clear any residual far-end echos from previous call-setup data. More precisely, we prime the channel with a smallest integer number of cycles of the training sequence that exceeds B/L, which we denote as $\lceil B/L \rceil$.

C. Recommended Protocol for Fast EC Training

Fig. 5 lists the V.32 initialization or retrain protocols over the time period relevent to echo canceller training. We emphasize that the optional echo canceller (EC) training sequence is added just prior to the "dotting S" and $\bar{S}$ sequences, and is shown as PRIME and RTS in Fig. 5. The answer modem precedes EC with a short (too long will enable the poor-quality (20 dB) voice echo cancellers present in the toll network, a highly undesirable event) silence, although previous call-setup data can and usually will generate far-end echos that will overlay the EC sequence. Also in Fig. 5, the EC portion has been expanded as detailed in this subsection.

To ensure that the above remnant far-end echos have dissipated

$$\begin{bmatrix} s_k \\ s_{k-1} \\ \vdots \\ s_{k-L+1} \end{bmatrix} = \begin{bmatrix} a_k & \cdots & a_{k-NN} & \cdots & a_{k-L+1+2N} & a_{k-B} & \cdots & a_{k-B-2N+1} \\ a_{k-1} & \cdots & a_{k-N-1} & \cdots & a_{k-L+2N} & a_{k-B-1} & \cdots & a_{k-B-2N} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{k-L+1} & \cdots & a_{k-N-L+1} & \cdots & a_{k-2L+2N+1} & a_{k-L-B+1} & \cdots & a_{k-B-2N-L+1} \end{bmatrix} \begin{bmatrix} w_{0,k} \\ w_{1,k} \\ \vdots \\ w_{L-1,k} \end{bmatrix} \tag{30}$$

where $w_{i,k}$ is the ith (starting at 0) coefficient in the time-domain settings for the particular subcanceller in question. Upon careful inspection of the matrix in (30) and recalling (29), it is determined that this data matrix is circulant, which is the condition necessary for the DFT division of the last section to be both computationally efficient and to produce a high-quality estimate. The extra $L - 3N$ (or $L - N_n - N_f$) coefficients in positions (N_n) to $(L - N_f + 1)$ are ignored (not produced by inverse transformation) as they should be zero if intermediate echos are absent. The circulant structure to be maintained in (30) is, again, the reason that the training sequence prior to training, the first

$$\mathrm{PRIME} = \left\lceil \frac{B + 2N}{L} \right\rceil \quad \text{or} \quad \left\lceil \frac{B + N_f}{L} \right\rceil \tag{31}$$

cycles of the length-L training sequence prime the channel to ensure that far-end echos have been generated by a "spectrally rich" input (rather than by a sinusoid). Following PRIME, two length-L cycles are used for echo canceller training.

The DFT of $\tilde{a}_k$ (denoted $\mathfrak{A}_k$) can be computed during PRIME.

Since the inverse of $\mathcal{A}_k$ is necessary in the frequency-domain computations, this inverse is also computed during PRIME. These quantities can, in fact, be computed at any time after the bulk delay B (and, thus, L) have been estimated. $\mathcal{S}_{p,k}$ is computed recursively as detailed in the next subsection during the first and second cycles of the training period. $\mathcal{W}_k$ is computed during S.

Finally, the stochastic-gradient method of (14) and (15) is then used on the TRN and ensuing data and continues to operate throughout the duration of the call (except during a retrain). The far-end equalizer is also trained on the TRN sequence. The CCITT minimum of $1280T$ should be sufficient for obtaining the remaining 1.8 dB of echo canceller performance, and should also be sufficient for the far-end equalizer to progress into its decision-directed mode of operation.

D. Training Time

In this subsection, we compute the actual training time for both the fast-starting echo canceller and a nominal echo canceller that uses only stochastic-gradient algorithms.

Regardless of the specific echo canceller adjustment or training method, the V.32 protocol always requires

$$BASE = 1896 + 5.5B \tag{32}$$

symbol periods of overhead. This expression is obtained by adding the delays (in symbol periods, T) along the top of Fig. 5, with the interval *"LMSEC or FSTART"* excluded presently as it alone depends on the algorithm employed to converge the EC. The worst case value of this expression occurs when $B = 1440$, its maximum value.[3] Nominally, $B \cong 144$ (60 ms). Then $BASE = 4.1$ s (maximum) and 0.9 s nominal. If stochastic-gradient adjustment techniques are used, there are two training times that can be cited. The first, preferred by the author, recognizes the slow convergence of the gradient algorithm in this application, and allows for the CCITT maximum training time to ensure convergence to the best performance level before training other components, such as the adaptive equalizer. In this case, an additional

$$LMSEC = 4(8192) + 2(272) + 16 = 33\,328 \tag{33}$$

symbol periods are required (independent of B). Note the B periods of S prior to EC training can be omitted (without violating the protocol), as the $8192T$ symbol periods of extra training maximum is more than sufficient to span the maximum bulk delay. Then, worst case $(BASE+LMSEC-B)T = 17.3$ s, and nominally (with $B = 60$ ms), the training time is 14.9 s. Both figures are indicative of the stochastic-gradient algorithm's well-known slow convergence time in this application [2]-[4]. Some manufacturers use various "gear-shifting" (variable step size) methods, empirically derived, to reduce the extra 8192 learning time. The new method presented requires less computation and is not empirically derived, two advantages with respect to the ad hoc gear shifting. However, presuming that such methods can be used to remove the additional 8192 symbol periods of extra training at both ends, then (33) becomes (and the B symbol periods of S prior to the optional training now cannot be omitted)

$$LMSEC = 2(8192) + 2(272) + 16 = 16\,944. \tag{34}$$

This leads to a worst case LMS training time of $(BASE + LMSEC)T = 11.2$ s, and a nominal startup time of 8.2 s. For the frequency-domain fast start of this paper, (32) still holds, but the equivalent of (33) or (34) is

$$FSTART = 2(PRIME + 2L + 1280T + 272T) + 16T. \tag{35}$$

Worst case, when $B = 1440$, this expression evaluates to $7(240) + 2(240) + 1280 + 272T + 8$ or 3720 symbol periods for one direction. This leaves $7432T$ or 3.1 s for both modems. The total is then $(BASE+FSTART)T = 7.2$ s. Nominally [with $B = 144$ (or 60 ms)], we obtain $FSTART = 2(1(240)+2(240)+1280+272)+16 = 4560$, leaving a nominal total start time of $1.9 + .9 = 2.8$ s. The improvements over the stochastic-gradient methods are substantial, with the latter figure of 2.8 s in the usual (nonsatellite) case particularly attractive, especially for retrains. Of course, this presumes that both modems are fast starting. If one is LMS and the other is fast, then we get approximate figures of 8.8 s (nominal) and 12.3 s (maximum) for the case corresponding to (33), and 7.2 s (nominal) and 9.2 s (maximum) for the case corresponding to (33). We plot an instance of the case corresponding to the nominal case of $B = 144$ in Fig. 6.

[3] While the CCITT specifications require that bulk delays of up to 600 ms be accommodated, some manufacturers design for 1200 ms (1.2 s) to be extra cautious (and allow for two satellite hops).

E. Computational Requirement

The proposed fast initialization method is extremely efficient, even if fast DFT methods (or FFT's) are not used. In fact, because of the variable length training sequence, it is advisable not to use FFT methods as the specific program used to implement them efficiently is too strongly a function of L for them to be of practical use in this V.32 application.

We thus compute the DFT's of the real sequences $\tilde{a}_k$ and $s_{p,k}$ in a straightforward manner according to (16) and (17). This requires $L^2/2$ real operations (multiply-accumulates). For $\mathcal{S}_{p,k}$, it is most convenient to compute each of the L (actually, $L/2$ complex) coefficients recursively in time as

$$\mathcal{S}_{p,n}(k) = \mathcal{S}_{p,n}(k-1) + s_{p,k}e^{-(j2\pi kn/L)}. \tag{36}$$

After L iterations of (36), the DFT will have been computed. The evaluation of (36) requires the evaluation of the complex exponential. The values for $\cos(\theta)$ can be stored in a 64-location memory, with entries $k\pi/64$ for $k = 0, \cdots, 63$. The quantity $\sin(\theta)$ is found through the relation

$$\sin(\theta) = \cos(\theta - \pi/2). \tag{37}$$

The maximum fractional quantization error in using only 64 locations can be easily determined using trigonometric identities as

$$1 - \cos(\pi/128) = 0.000301 \tag{38}$$

which is about 70 dB down from nominal signal levels, more than sufficient. The contents of the memory can be prestored or computed at any time after the modem is powered. Actually, (36) is used over the $2L$ cycle times corresponding to the two cycles of the periodic training sequence if k in the complex exponential is replaced by $k_{\bmod L}$.

In the protocol Section III-C, $\mathcal{A}_n$ is computed during the PRIME time period of the echo canceller starting sequence. This DFT is common to all three subcanceller combinations (near-far). One complex DFT coefficient can be computed per two symbol intervals or per $T = 1/2400$, requiring L multiply-accumulates. One complex divide is necessary every $2T$ also over the PRIME segment to compute $1/\mathcal{A}_n$. These complex divides require six real multiplications and one real divide, which is easily determined by noting that a complex divide can be written as the solution of the following 2×2 set of linear equations:

$$(\mathcal{W}^r + j\mathcal{W}^i)(\mathcal{A}^r + j\mathcal{A}^i) = (\mathcal{S}^r + j\mathcal{S}^i) \tag{39}$$

is equivalent to

$$\begin{bmatrix} \mathcal{A}^r & -\mathcal{A}^i \\ \mathcal{A}^i & \mathcal{A}^r \end{bmatrix} \begin{bmatrix} \mathcal{W}^r \\ \mathcal{W}^i \end{bmatrix} = \begin{bmatrix} \mathcal{S}^r \\ \mathcal{S}^i \end{bmatrix} \tag{40}$$

where unnecessary time subscripts have been dropped to simplify this complex divide description, and r and i are used as superscripts to denote real and imaginary parts. Equation (40) becomes

$$\begin{bmatrix} \mathcal{W}^r \\ \mathcal{W}^i \end{bmatrix} = \frac{1}{(A^r)^2 + (A^i)^2} \begin{bmatrix} \mathcal{A}^r & \mathcal{A}^i \\ -\mathcal{A}^i & \mathcal{A}^r \end{bmatrix} \begin{bmatrix} \mathcal{S}^r \\ \mathcal{S}^i \end{bmatrix}. \tag{41}$$

As the matrix in (41) is used $p = 3$ times, it is precomputed, requiring two real multiplies and one real divide. The divide can be computed using a maximum of ten real multiplies. These extra 12 multiplies per $2T$ over the PRIME time period are inconsequential in comparison to the DFT coefficient calculation itself.

On the very first few iterations of S in the protocol, $\mathcal{W}_{p,n}$ is computed using four real multiplies for each of the L values of n. To compute $\mathcal{W}_n$, we use (20), requiring four real multiplications for each of the $3L$ DFT coefficients. Finally, the appropriate $3N$ time-domain coefficients and their Hilbert transform can be computed from the IDFT formula:

$$w_k = \mathrm{Re}\left\{\frac{1}{3L}(I_k + Q_k)\right\} \tag{42}$$

and

$$\hat{w}_k = \mathrm{Re}\left\{\frac{1}{3L}(-jI_k + jQ_k)\right\} \tag{43}$$

where

$$I_k = \sum_{n=0}^{3L/2-1} \mathcal{W}_n e^{j(2\pi kn/3L)} \tag{44}$$

and

$$Q_k = \sum_{n=3L/2}^{3L-1} \mathcal{W}_n e^{j(2\pi kn/3L)}. \tag{45}$$

Equations (44) and (45) are trivially modified for $3L$ an odd number.

The steady-state echo cancellation algorithm (stochastic-gradient) requires approximately $12N$ multiply–accumulate operations per T [2($3N$ in-phase + $3N$ quadrature)]. Worst case, the above computations require $2L = 12N$ per T; thus, no extra computational power is necessary using this fast start. Of course, the appropriate microcode software to implement the fast initialization is necessary, but this is the only additional cost.

All of the discussion to this point has been for the usual case of no frequency offset in the far-end echo and no intermediate echos. If either or both of these unusual impairments are to be tolerated by the high-performance V.32 modem, the modifications of the next section are necessary.

IV. Frequency Offset and Intermediate Echos

In this section, a method for fast acquisition of frequency offset in the far-end echo, without increasing the training time of the modem, is studied. To this objective, it is fortuitous that the V.32 protocol uses sinusoidal tones in the initial handshaking procedures for the bulk-delay estimation. The method recommended here will use these tones to get an initial estimate of any frequency offset in the far-end echo.

Section IV-A specifically investigates the special structure of the echo when a sinusoid excites the echo channel. Section IV-B discusses the distinction of the far-end echo from the near-end echo when the echo channel is sinusoidally excited. Section IV-C discusses the initial acquisition of the frequency offset with a special phase-lock loop.

A. Sinusoidal Echos

The CCITT V.32 protocol mandates that several sinusoidal signals be transmitted after the V.25 handshaking answer tone has been received by the calling modem. The caller does not transmit for more than 1 s during this V.25 interval, thus eliminating the possibility of any residual far-end echos when caller nonzero transmission begins. Simultaneously, the answer modem transmits the V.25 tone at 2100 Hz for more than 1 s. Any residual 2100 Hz answer-modem far echos will be out of band with the sinusoidal bulk-delay-estimating signals described in Section III-A.

Following V.25, the calling modem transmits the sequence $AAAAA\cdots$, until a phase reversal is detected from $ACACAC\cdots$ to $CACACA\cdots$ in the incoming answer-modem-transmitted sequence. Then the caller reverses its phase to $CCCCCC\cdots$. The answer modem detects this call-modem phase reversal, and then again reverses its answer-modem phase back to $ACACAC\cdots$. The call modem, upon detecting the second incoming phase reversal, stops transmitting. Finally, the answer modem detects the absence of caller transmission, and also stops transmitting for $16T$ prior to the beginning of the echo canceller training that was discussed earlier in Section III.

The caller thus transmits a tone at 1800 Hz during the $AAAAA$ and $CCCCC$ signals initially. The answer modem, in effect, transmits two equal energy tones at 600 and 3000 Hz for the $ACACACAC$ and $CACACACA$ sequences. These three tones are separable, as is the residual V.25 far-echo tone at 2100 Hz (if present). Appropriate bandpass filters need to be initially used during these transmissions to ensure that only the 1800 Hz echo is processed by the calling modem's echo canceller (to be described in this section) and only the 600 Hz echo tone is processed by the answer modem's echo canceller. These bandpass filters ensure the absence of a double talker in the simple sinusoidal cancellers to be described.

Recalling (7), the caller's $T/3$ data-driven echo canceller sees the sequence $ADCBADCB\cdots$ at the inputs to its subcancellers when $AAAAA$ are the actual data transmitted. Actually, a baseband canceller followed by a modulator could be used to avoid the input rotation, but we continue with the $T/3$ structure here to avoid complicating the development by introducing yet another structure at this point. Note that $ADCBADCB$ alternately excites the rotated (see Section II-C) in-phase and quadrature subcancellers as shown in Fig. 7. We have picked the angle of our receiver phase such that A and C fall along a rotated in-phase axis, while B and D fall along the rotated quadrature axis, as discussed in Section II-C. We also assume that the input is ± 1 along either of these axes, rather than the nominal 5 as the echo canceller settings to be generated are only of interest during frequency-offset acquisition, and this scaling simplifies computation without affecting the accuracy of this acquisition. On the first nonzero tonal excitation of the channel, the near-end in-phase echo is $-\bar{h}_{r,0}$ (assuming that we start the transmitter with symbol A), while the quadrature echo is 0. This is the same for all subcancellers, with only the quantity $\bar{h}_{r,0}$ changing with the subcanceller index. On the next symbol period, the in-phase echo becomes $-\bar{h}_{r,1}$ and the quadrature becomes $-\bar{h}_{q,0}$. The excitation continues, producing the following sequence of in-phase and quadrature echos:

$$\begin{array}{l} -\bar{h}_{r,0} \\ 0 \\ -\bar{h}_{r,1} \\ +\bar{h}_{q,0} \\ -\bar{h}_{r,2} + \bar{h}_{r,0} \\ +\bar{h}_{q,1} \\ -\bar{h}_{r,3} + \bar{h}_{r,1} \\ -\bar{h}_{q,2} - \bar{h}_{q,0} \\ -\bar{h}_{r,4} + \bar{h}_{r,2} - \bar{h}_{r,0} \\ +\bar{h}_{q,3} - \bar{h}_{q,1} \end{array} \tag{46}$$

and so on. After $N_n T = 48T$, one then determines that the in-phase subcanceller can be separated into two one-tap minicancellers, one for even kT and the other for odd kT. The setting for the even(odd) minicancellers is an alternating-sign sum of the even(odd) echo channel coefficients:

$$(-1)^{(k/2)} \sum_{m=0}^{23} \bar{h}_{r,2m}(-1)^m, \qquad k \geq 48,\ k \text{ even} \tag{47}$$

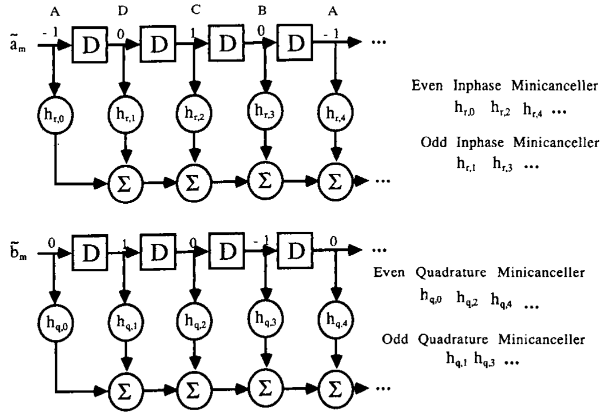


Fig. 7. Minicanceller simplification for sinusoidal echo cancellation.

and

$$(-1)^{(k-1/2)}\sum_{m=0}^{23}\bar{h}_{r,2m+1}(-1)^m, \qquad k > 48,\ k \text{ odd}, \tag{48}$$

respectively. Similar statements hold for the quadrature subcanceller:

$$(-1)^{(k+1/2)}\sum_{m=0}^{23}\bar{h}_{q,2m}(-1)^m, \qquad k > 48,\ k \text{ odd} \tag{49}$$

and

$$(-1)^{(k/2)}\sum_{m=0}^{23}\bar{h}_{q,2m+1}(-1)^m, \qquad k \geq 50,\ k \text{ even}. \tag{50}$$

The sum coefficients in (47) and (48) can be estimated recursively by determining each of the individual coefficients as they are excited:

$$\bar{h}_{r,k} = \frac{-1}{2}((s_k + (\tilde{s})_k) - (y_k + \tilde{y}_k))$$
$$\bar{h}_{q,k} = \frac{-1}{2}((s_k - \tilde{s}_k) - (y_k - \tilde{y}_k)) \tag{51}$$

where the tildes denote Hilbert transforms, and y is the echo canceller output using previously estimated coefficients (the Hilbert is trivially obtaining by exciting the quadrature canceller with the real input and summing with the output of the in-phase canceller when excited with the quadrature input). That is (for k even),

$$\tilde{y}_k = (-1)^{k/2-1}\sum_{m=0}^{k/2+1}\bar{h}_{r,2i}(-1)^m + (-1)^{k/2}\sum_{m=0}^{k/2}\bar{h}_{q,2m}(-1)^m \tag{52}$$

and (for k odd)

$$\tilde{y}_k = (-1)^{k+1/2}\sum_{i=1}^{k-1/2}\bar{h}_{r,2m}(-1)^m + (-1)^{k-1/2}\sum_{m=0}^{k-1/2}\bar{h}_{q,2m+1}(-1)^m. \tag{53}$$

The coefficients h are stored after N iterations, even though the sums in (47) and (48) are sufficient for the sinusoidal cancellation. Then, they are used after the phase reversal to adjust the sums, one term at a time, for the N iterations following such a phase reversal. Equivalently, convolution can be used as no far-end echo computation is occurring, and there is thus available computation cycles, in excess of that required, on the DSP. Because of the bandpass filters mentioned earlier, the process in (51) is basically noiseless; thus, the sinusoidal cancellation will be accurate after N steps as it is virtually deterministic and can be computed without averaging. Two or three steps of a stochastic-gradient algorithm can be continued after time N to introduce a small amount of averaging into the coefficient settings; then the updating should be discontinued and the near-end settings frozen. For the answer modem, the input is constant, rather than alternating as in the call modem, so all the -1's in (52) and (53) become $+1$'s and drop out. Numerical problems are not significant when $N \leq 48$, as here.

B. Distinguishing the Far-End Echo

The near-end echo for the sinusoidal input is thus very quickly estimated and cancelled as described in Section IV-A. The absence of the 1800 Hz (or 600 Hz) sinusoidal near-end echo permits the search for another sinusoid of the same frequency, denoting the onset of the far-end echo. This far-end echo is typically of magnitude much smaller than the near-end echo, which is the reason for cancelling the near-end sinusoidal echo so that this smaller echo can be detected without the larger interfering near-end echo present.

The energy level of the echo canceller error signal is monitored until a significant increase in energy is detected for several successive symbol periods, signifying the arrival of the far echo (or possibly an intermediate echo, as described in Section IV-D). The bulk delay of an intermediate echo would thus be estimated. Of course, the bulk delay of the far-end echo is determined later in the V.32 protocol, but the fast acquisition of any frequency offset requires that we know this bulk delay earlier in the protocol than it would otherwise be computed.

Mathematically, the detection of the far-end echo can be described in detail in terms of a hypothesis test:

$$H0\text{: echo present} \tag{54}$$

versus

$$HA\text{: echo absent.} \tag{55}$$

When the near-end echo has been cancelled, the worst case far-end-echo-detection situation occurs when the far-end hybrid is performing well and the far-end echo power level is comparable to the channel noise level. For simplicity, we will assume that we are distinguishing a Gaussian zero-mean random variable with power σ^2 (echo absent) from a Gaussian zero-mean random variable with power $2\sigma^2$ (or higher). The test statistic under these conditions is well known to the sample-mean-square, which has a chi-square distribution.

Specifically, we form

$$\xi_k = \frac{1}{M}\sum_{m=0}^{M-1}\epsilon_{k-m}^2 \tag{56}$$

where the error signals are taken at the higher sampling rate $3/T$. We note that $(m-1)$ previous near-end canceller errors must be stored and used initially in (56). For the case that the far-end echo immediately follows the near-end echo. This method will work if near and far echos overlap with frequency offset in the far echo. The "optimum" (Neyman–Pierson) test is to compare ξ_k to a threshold, below which we decide that there is no far-end echo, and above which we decide that an echo is present. We define α as the error probability of deciding that no echo is present erroneously, and preset its value to 0.1%, that is,

$$\alpha = P[HA/H0] = 0.001. \tag{57}$$

In other words, 999 times out of 1000 worst case situations, we correctly detect the location of the far-end echo. In this case, by setting M in (56) to 200, one determines that the probability of false detection is below 0.01%. These numbers are probably excessive performance specifications, but the processor is not fully utilized when the bulk-delay estimation takes place, so the extra 200 multiplications are not a computational burden. As the far-end echo may more typically be 10–100 times larger in power than the channel noise level, the designer could probably reduce M to about 10 or 20 with only the reservations of a perfectionist. Of course, the far-end echo is not Gaussian, as it is a shaped data signal, so the analysis

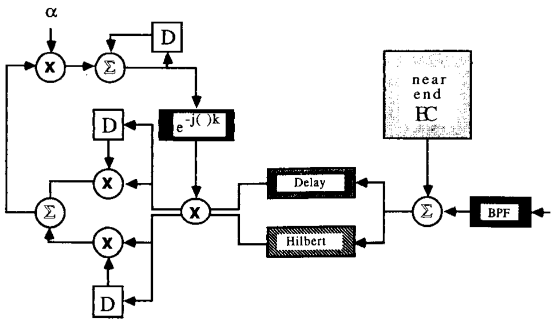


Fig. 8. Fast frequency offset acquisition.

presented in this subsection will only yield a rough figure of merit. The best value of M is probably best determined empirically.

The threshold value for ξ_k is about $1.5\sigma^2$, requiring that the channel noise power be estimated during the initial silent period of the call. This can be done to a reasonable degree of accuracy by sensing the silent line signal for several sampling periods and sample averaging its squared value:

$$\sigma^2 \cong 2^{-7}\sum_{i=1}^{128} n_i^2. \tag{58}$$

However, the setting of the gain control circuit would have to be known (and fixed during the bulk delay estimation) with respect to its value (also fixed, but probably different when (58) is computed so that the proper scaling adjustments could be made).

C. Fast Frequency-Offset Acquisition

Since the near-end canceller taps are frozen at the onset of the far-end echo, the far-end echo tone plus offset frequency 1800 (or 600) + Δ is present in the error signal. Δ must now be determined with a phase-lock loop.

Fig. 8 illustrates a simple PLL that suffices for this application. After waiting $2NT$ for the far-end echo to stabilize in magnitude, the error signal at the subcanceller output sampling instants should be (18/72)360 or 90° separated in the call modem and (6/24)(360), also 90° separated in the answer modem (at the symbol sampling instants, rather than at $T/3$ as the answer modem).

To understand Fig. 8, we focus on the call modem that operates in the frequency acquisition phase at a sampling rate of 7200 Hz on the 1800 Hz carrier. The phase splitter generates a two-dimensional output, for which successive two-vectors should be rotated by 90° with respect to each other if there is no frequency offset (or equivalently, our frequency-offset estimate is accurate). Thus, the inner product between these two successive 2-D vectors should be zero. Note that the circuitry following the derotation by the current frequency-offset estimate computes this inner product using two multiplies an an addition. A positive value for this inner product means that our estimate Δ should be decreased. Correspondingly, a negative value for the inner product means that Δ should be increased. The remaining circuitry (computation) directly accumulates the frequency offset estimates. Note that this is effectively a second-order loop that recovers frequency and some constant phase. Phase is not important because our Nyquist echo canceller (not yet trained) can compensate for any phase as long as we supply an accurate frequency-offset estimate. The sines and cosines necessary for the rotation can be computed with the lookup-table cosine discussed in connection with the DFT's in Section III. The answer modem duplicates the computation associated with Fig. 8 at 1/3 the rate of the call modem.

If the far-end echo level is sufficiently high to be of concern initially, then the SNR is very high for this phase-lock loop, and it thus coverges rapidly. The consultant has simulated this loop, and the rapid acquisition is illustrated in Fig. 9 under worst case conditions.

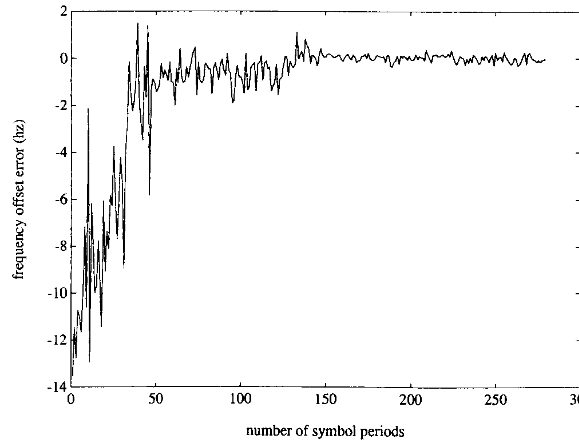


Fig. 9. Fast acquisition of frequency offset using CCITT protocol tones.

A very accurate estimate of the offset can be obtained by averaging the last 50 estimates of the PLL.

D. Intermediate and Overlapping Echos

There can be more than one far-end echo in voiceband data channels, particularly as equal-access network constraints have made possible the existence of two-to-four wire interfaces at intermediate positions within the channel. Also, the use of the V.32 modem in some private networks (not conforming to general switched network standards) can also give rise to intermediate echos. The corresponding echos are called "intermediate" echos here so as to distinguish them from the echo generated by the far-end four-to-two wire interface in the receiving (far-end) data set. Each of the intermediate echos can have a different bulk delay and may require a separate echo canceller (consisting of three in-phase and three quadrature subcancellers).

The bulk delay and frequency offset of the first intermediate echo are determined as discussed previously in this section. Further echo frequency offsets are determined by demodulating the echo by the first frequency $w_c + \Delta_1$ and then low-pass filtering (also rejecting dc) before estimating the offset of the second with respect to the first in an additional phase-lock loop $\Delta_2 - \Delta_1$.

Intermediate echos are difficult as more memory and processor time must be available for the extra echos. The modem manufacturer must make a decision as to whether the extra cost of the processing is merited to handle this situation.

V. Conclusion

The fast start in the V.32 modem is a desirable feature as the user should not have to wait 15–18 s at the call setup time or during a retrain, nor should empirically derived gear-shifting methods be necessary. However, the author is not aware of any marketing surveys that support or refute this claim. Nevertheless, the proposed methods add no computational cost to the implementation, thus enabling a rare opportunity for substantial performance improvement at no cost increase.

Furthermore, the fast start can be very attractive from another limited-precision viewpoint. The frequency-domain start here gives good numerical performance (accuracy) as it is a block-type algorithm for the echo·canceller (see [10]). It requires a substantially smaller coefficient and processor wordlength than the limited-precision slow (LMS or gradient) starting V.32 modem. Thus, if block algorithms, which trade numerical accuracy for convergence speed in this application, are to be used to reduce memory requirements in the V.32 modem, the fast start becomes increasingly attractive as it can also be achieved with low precision and the convergence speed of the ensuing block LMS is not an issue after fast training. In fact, in this practical sense, a fast-starting V.32 modem, such as the one described, can result in a lower cost of implementation for the V.32 modem digital-signal-processing functions.

ACKNOWLEDGMENT

The author would like to thank G. Long and F. Ling of Codex for helpful comments, and also S. Unagami of Fujitsu Laboratories for several interesting conversations related to the content of this paper.

REFERENCES

[1] CCITT, "V.32—A family of 2-wire, duplex modems operating at data signalling rates of up to 9600 bit/s for use on the general switched telephone network and on leased telephone-type circuits," *CCITT Red Book*, Fasicicle VIII.1, pp. 221-238, 1984.

[2] J. Salz, "On the start-up problem in digital echo cancellers," *Bell Syst. Tech. J.*, vol. 62, pp. 1353-1364, July-Aug. 1983.

[3] M. L. Honig, "Echo cancellation of voice-band data signals using RLS and stochastic-gradient algorithms," *IEEE Trans. Commun.*, vol. COM-33, pp. 65-73, Jan. 1985.

[4] J. M. Cioffi and T. Kailath, "An efficient, RLS, data-driven echo canceller for fast initialization of full-duplex data transmission," *IEEE Trans. Commun.*, vol. COM-33, pp. 601-611, July 1985.

[5] J. J. Werner, "An echo-cancellation-based 4800 bit/s full-duplex DDD modem," *IEEE J. Select. Areas Commun.*, vol. SAC-2, pp. 722-730, Sept. 1984.

[6] J. M. Cioffi and J. J. Werner, "The tap-drifting problem in digitally implemented data-driven echo cancellers," *AT&T Bell Lab. Tech. J.*, vol. 64, pp. 115-138, Jan. 1985.

[7] J. G. Proakis, *Digital Communications*. New York: McGraw-Hill, 1983.

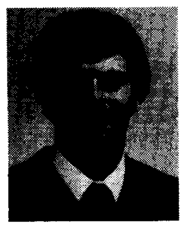
[8] J. J. Werner, "Effects of channel impairments on the performance of an in-band data-driven echo canceller," *AT&T Bell Lab. Tech. J.*, pp. 91-108, Jan. 1985.

[9] R. E. Crochiere and L. R. Rabiner, *Multirate Digital Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1983.

[10] J. M. Cioffi, "Memory efficient use of block adaptive algorithms in data driven echo cancellers," in *Proc. IEEE Int. Conf. Commun.*, Philadelphia, PA, June 1988.

[11] G. Long and F. Ling, "A new complex system identification method and its application to echo canceller fast initialization," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing,* Albuquerque, NM, Apr. 1990, pp. 1671-1674.

[12] F. Ling and G. Long, "Correlation-based fast training of data-driven Nyquist in band echo cancellers," in *Proc. IEEE Int. Conf. Commun.*, Atlanta, GA, Apr. 1990, pp. 1280-1284.



John M. Cioffi (S'77-M'79) was born in Harvey, IL, on November 7, 1956. He received the B.S. degree from the University of Illinois, Urbana, in 1978, the M.S. degree from Stanford University, Stanford, CA, in 1979, and the Ph.D. degree, also from Stanford University, in 1984, all in electrical engineering.

He has been an Assistant Professor of Electrical Engineering at Stanford University since 1986. His interests are in digital communication and storage. He was a member of the Technical Staff at Bell Laboratories, Holmdel, NJ, from 1978 to 1984 and a Research Staff Member at IBM Research, San Jose, CA, from 1984 to 1986.

Dr. Cioffi has been an Associate Editor for the IEEE TRANSACTIONS ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING since 1984, and is an NSF Presidential Investigator (1987-1992).