ECHO CANCELLATION FOR SPEECH SIGNALS

With rare exceptions, conversations take place in the presence of echoes. We hear echoes of our speech waves as they are reflected from the floor, walls, and other neighboring objects. If a reflected wave arrives a very short time after the direct sound, it is perceived not as an echo but as a spectral distortion, or reverberation. Most people prefer some amount of reverberation to a completely anechoic environment, and the desirable amount of reverberation depends on the application. (For example, much more reverberation is desirable in a concert hall than in an office.) The situation is very different, however, when the leading edge of the reflected wave arrives a few tens of milliseconds after the direct sound. In such a case, it is heard as a distinct echo. Such echoes are invariably annoying, and under extreme conditions can completely disrupt a conversation. It is such distinct echoes that this article discusses.

Echoes have long been the concern of architects and designers of concert halls. However, since the advent of telephony, they have also been the concern of communications engineers, because echoes can be generated *electrically,* due to impedance mismatches at points along the transmission medium. Such echoes are called *line echoes.*

If the telephone connection is between two handsets, the only type of echoes encountered are line echoes. These echoes are not a problem in local telephone calls because the sources of echo are insignificant, and the echoes, if any, occur after very short delays. However, in a long-distance connection in which the end-to-end delay is nonnegligible, the echoes may be heard as distinct echoes. A significant source of line echoes in such circuits is a device called a hybrid, which we discuss briefly in the following section.

Echoes at hybrids have been a potential source of degradation in the telephone network for many decades, and many solutions have been devised to overcome them. Of particular interest to us are devices known as adaptive echo cancelers. Interest in such devices arose during the 1960s, in anticipation of telephone communications via satellites (1,2). As satellite communication gained an ever-increasing share of telephone traffic during the 1970s, considerable development of echo cancelers took place (3–6). Their widespread use began around 1980 with the arrival of a very large scale integration (VLSI) implementation (7). More recently, with the growing use of speech coding in the telecommunications network, delay has again become an issue, thereby further mandating the use of echo cancelers.

When the telephone connection is between hands-free telephones or between two conference rooms, a major source of echoes is the acoustic coupling between the loudspeaker and the microphone at each end. Such echoes have been called *acoustic echoes,* and interest in adaptive cancellation of such echoes has attracted much attention during the past two decades. A caveat to the reader is in order at this point. Although we will be dealing with acoustically generated echoes, we will only consider cancellation of these echoes in the electrical portion of the circuit. We will not discuss the related, but much more difficult, problem of canceling echoes acoustically [i.e., active noise control (8)]. Although single-channel acoustic echo cancelers are in widespread use today, the more difficult problem of multichannel (e.g. stereo) acoustic echo cancellation will doubtlessly arise in future applications involving multiple conference parties and/or superposition of stereo music and other sound effects (for example, in interactive video gaming). We will discuss recently developed methods for echo cancellation in such applications.

In the next two sections we will briefly discuss the problem of line echoes and adaptive cancellation of such echoes. We refer the reader to review articles (9,10) for a more detailed account. Besides introducing the reader to the echo problem, this preliminary discussion will also lay the groundwork for the more modern problem of canceling acoustically generated echoes in both single-channel and multichannel applications, which will be discussed in later sections.

LINE ECHOES

As mentioned in the preceding section, the main source of line echoes is the device known as a hybrid. Figure 1 illustrates, in a highly simplified manner, the function and placement of hybrids in a typical long-distance telephone connection.

Every conventional analog telephone in a given geographical area is connected to a central office by a two-wire line, called the customer loop, which serves for communication in either direction. A local call is set up by simply connecting the two customer loops at the central office. When the distance between the two telephones exceeds about 35 miles, amplification becomes necessary. Therefore, a separate path is needed for each direction of transmission. The device that connects the four-wire part of the circuit to the two-wire portion at each end is known as a hybrid (or a hybrid transformer). With reference to Fig. 1, the purpose of the hybrids is to allow signals from A to go along the path L_1 to B, and to go from B along the path L_2 to A. However, they must prevent signals in path L_1 from returning along the path L_2 back to A. Similarly, the signal in path L_2 is to be prevented from returning along path L_1 back to B.

We do not wish to go into the detailed workings of a hybrid. Further information can be found in Ref. 9 and other references cited therein. Suffice it to say here that a hybrid is a bridge network that can achieve the aforementioned objectives, provided the impedance of the customer loop can be exactly balanced by an impedance located at the hybrid. Unfortunately, this is not possible in practice because there are far fewer four-wire circuits than there are two-wire circuits. Therefore, a hybrid may be connected to any of the customer loops served by the central office. By their very nature, cus-

J. Webster (ed.), *Wiley Encyclopedia of Electrical and Electronics Engineering.* Copyright © 1999 John Wiley & Sons, Inc.

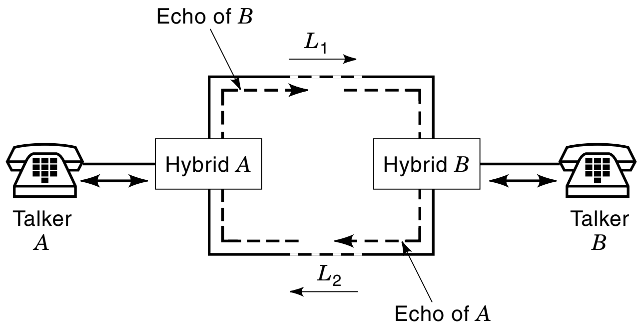

Figure 1. Illustration of a long-distance connection showing local 2-wire loops connected through hybrids to a 4-wire long-line network.

tomer loops have a wide variety of characteristics—various lengths, type of wire, type of telephone, number of extension phones, and so on. It appears, therefore, that the echo at the hybrid cannot be completely eliminated. As a compromise, a nominal impedance is used to balance the bridge, and the average attenuation (in the United States) from input to the return-path output of the hybrid is 11 dB with a standard deviation of 3 dB. This amount of attenuation is not adequate for satisfactory communication on circuits with long delays because the echoes remain audible (9).

The Echo Suppressor

The problem of such echoes has been around ever since the introduction of long-distance communication. On terrestrial circuits, the device most widely used to control line echoes is the echo suppressor (9). Again, we will not describe echo suppressors in detail, but merely mention that they are voice-operated switches whose object is to remove the echo of the talker's speech and yet allow the listener to interrupt, as in normal conversation. The principle of the echo suppressor can be explained by referring to Fig. 2, which shows the end B of the telephone circuit of Fig. 1, with an echo suppressor included. Suppose A has been talking for a while. Based on the level of signals in the paths L_1 and L_2, a decision is made as to whether the signal in L_2 is an interruption by B trying to break into the conversation or an echo of A's speech. If the decision is the latter, then the circuit L_2 is opened (or a large loss is switched in). A similar switch at the other end prevents B's echo from returning to B. During so-called double-talk periods, when both A and B are speaking at the same time, echo suppression is inhibited so that A hears the speech from B superimposed on self-echo from A.

If the decision mechanism were to behave flawlessly, the echo suppressor would be a satisfactory form of echo control. The decision, however, cannot be perfect. The two signals that have to be distinguished are both speech signals, with more or less the same statistical properties. Essentially the only distinguishing property is the level. Therefore, sometimes a high level of echo is returned, and sometimes when the speech level is low (or during initial and final portions of speech bursts) the interrupter's speech is mutilated. However, with considerable ingenuity, echo suppressors have been designed to keep such malfunctions at an acceptable level. Selective echo suppression can also be applied within the structure of subband echo cancelers (11), to be discussed later.

The Line Echo Canceler

Echo suppressors served well for over 70 years on circuits with round-trip delays of less than about 100 ms, corresponding to land line distances of a few thousand miles. With the advent of commercial communications satellites in 1965, however, the situation changed significantly. A synchronous satellite (i.e., one that is stationary with respect to the earth) must have an orbit that is about 23,000 miles above the earth's surface. A telephone connection via such a satellite will have a round-trip echo delay of 500 ms to 600 ms (9). With such long delays, echo suppressors fail to function satisfactorily. The long delay induces a change in the pattern of conversation in a way so as to increase significantly the number of errors. New methods of echo control were proposed for circuits with such long delays. Of these, the most versatile, and the one in widespread use, is the adaptive echo canceler (1). The unique feature that makes it so attractive is that unlike other forms of echo control, the echo canceler does not tamper with the path carrying the echo. Therefore, it never mutilates the speech of the interrupting party. The basic idea of the echo canceler is illustrated in Fig. 3. Again we show only one canceler located at the end B of the telephone circuit of Fig. 1; a similar canceler is symmetrically located at the other end.

As illustrated in Fig. 3, instead of interrupting the path L_2, a synthetic echo is generated from A's speech and subtracted from the signal going out on the path L_2. The syn-

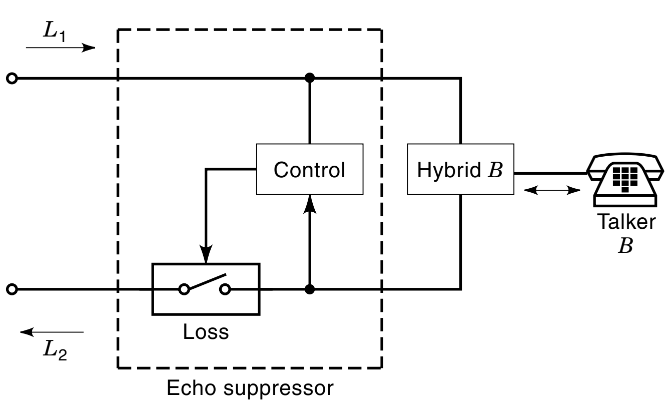

Figure 2. Echo suppressor attempts to remove echo by inserting switched loss when near-end speech (Talker B) is not present.

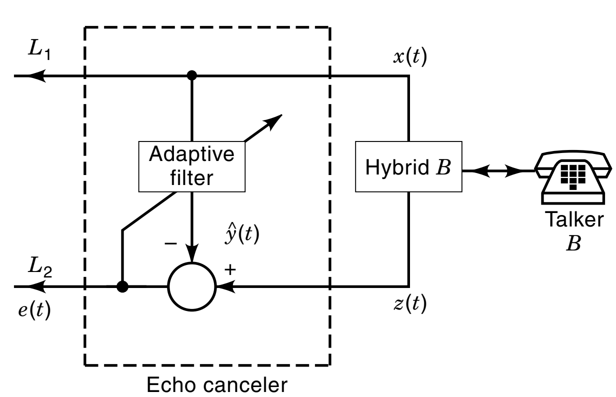

Figure 3. Echo canceler continually removes echo even if near-end talker is active.

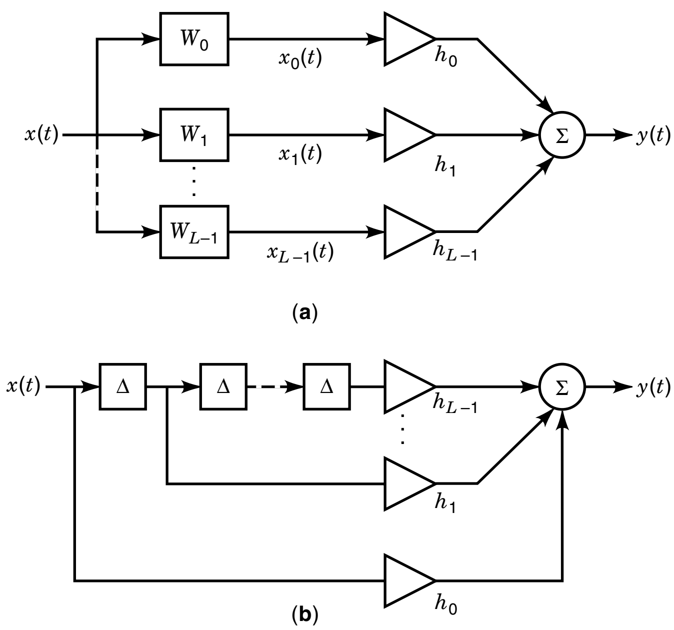

Figure 4. Two methods for synthesizing echoes: using a filter expansion (a) and tapped delay line filter (b).

thetic echo is generated by passing the signal of path L_1 through a filter whose impulse response (or transfer function) matches that of the echo path from $x(t)$ to $z(t)$ via hybrid B.

As mentioned previously, the echo path is highly variable, so the filter in Fig. 3 cannot be a fixed filter. It must be estimated for the particular local loop to which the hybrid gets connected. One simple way to derive the filter is to measure the impulse response of the echo path and then approximate it with some filter structure (e.g., a tapped delay line). However, the echo path is, in general, not stationary. Therefore, such measurements would have to be made repeatedly during a conversation. Clearly this is highly undesirable. To eliminate the need for such measurements, the filter is made adaptive. An algorithm is implemented that uses the residual error to adapt the filter to the characteristics of the local loop, and to track slow variations in these characteristics. In the next section we will discuss several basic adaptation algorithms in some detail.

ADAPTIVE CANCELLATION

To implement a filter that approximates the echo path, the first step is to choose a representation of the filter in terms of a finite number of parameters. Assuming the echo path to be linear, this can be achieved by finding an expansion of the impulse response of the echo path in terms of a set of basis functions. The problem then reduces to the estimation of the expansion coefficients. If $w_l(t)$, $l = 0, 1, 2, \ldots, L-1$ is the (truncated) set of basis functions, then the expansion can be implemented by the set of L filters illustrated in Fig. 4(a). The output of the filter bank, $y(t)$, is related to the input $x(t)$ by the relation

$$\begin{aligned} y(t) &= x(t) * \sum_{l=0}^{L-1} h_l w_l(t) \\ &= \sum_{l=0}^{L-1} h_l x_l(t) \\ &= \boldsymbol{h}^T \boldsymbol{x} \end{aligned} \tag{1}$$

Here $*$ indicates convolution, $x_l(t)$ is the output of the lth filter component, and h_l is the lth expansion coefficient. In the last line of Eq. (1) we have introduced matrix notation, which will be useful later. The boldface quantities $\boldsymbol{h}$ and $\boldsymbol{x}$ are column vectors with dimension $L \times 1$, and the superscript T denotes matrix transpose. Also, for simplicity of notation, we will suppress the dependence of quantities on the time t, except where it helps avoid confusion.

In the special case when $w_l(t) = \delta(t - l\Delta)$, the filter becomes an L-tap transversal filter (tapped delay line) with a delay Δ between taps, as illustrated in Fig. 4(b). This is the most commonly used filter structure, although other structures [e.g., when the $w_l(t)$'s are Laguerre functions or truncated (or damped) sinusoids] have been tried (1). In the discrete-time case, this structure is known as a finite impulse response (FIR) filter.

The general properties of the adaptation algorithms that we shall discuss presently arise mainly from the fact that the output depends linearly on the parameters h_l. Therefore, our discussion will apply for any choice of functions $w_l(t)$ (although, of course, the rate of convergence will depend strongly on that choice). The general features will, in fact, be valid even if the $x_l(t)$'s are nonlinearly filtered versions of $x(t)$. This fact allows one to handle a class of nonlinear echo paths by the same methods. A proposal to do this appears in Ref. 12 but, to our knowledge, has never been used in echo cancellation for speech signals.

The Stochastic Gradient Algorithm

By far the most popular algorithm for adapting the filter structure of Eq. (1) to the echo path is the stochastic gradient algorithm. It is now popularly known as the least mean square (LMS) algorithm and was first introduced around 1960 for adaptive switching (13). The LMS algorithm was initially used for echo cancelers (1) and adaptive antenna arrays (14) in the mid-1960s. Since then, its use has expanded to the general field of adaptive signal processing (15,16), finding applications in many other areas, such as interference cancellation, equalization, and system identification.

The basic idea of the stochastic gradient algorithm is quite simple. Suppose $z(t)$ is the hybrid return signal in Fig. 3. Let us assume that

$$z(t) = y(t) + v(t) \tag{2}$$

where $y(t)$ is an echo of the input signal $x(t)$ and $v(t)$ is an added noise component that may include talker B's speech. We will assume that $y(t)$ has the representation given in Eq. (1) for some (unknown) coefficient vector $\boldsymbol{h}$. If this is not strictly true, then $v(t)$ will include the residual modeling error as well.

Suppose an estimate of the echo

$$\hat{y}(t) = \hat{\boldsymbol{h}}^T \boldsymbol{x} \tag{3}$$

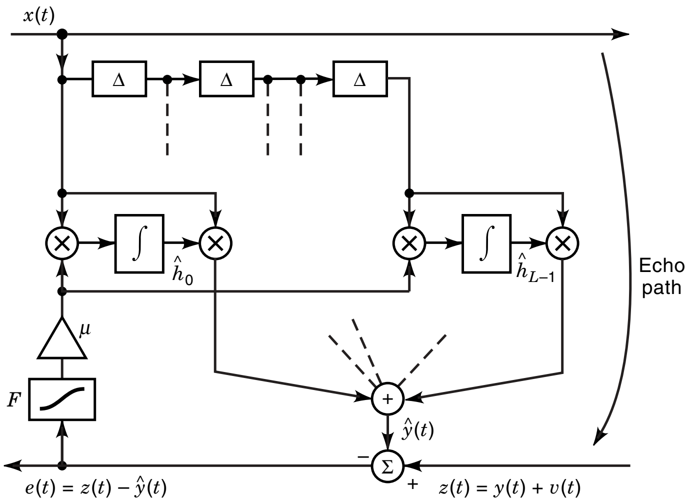

Figure 5. An echo canceler utilizing the stochastic gradient technique, also known as the LMS algorithm.

is formed with a trial coefficient vector $\hat{\boldsymbol{h}}$. We wish to implement an algorithm to improve $\hat{\boldsymbol{h}}$ (i.e., bring it closer to the vector $\boldsymbol{h}$). Since $\boldsymbol{h}$ is unknown, we must evaluate the goodness of $\hat{\boldsymbol{h}}$ indirectly. One measure of the performance of $\hat{\boldsymbol{h}}$ is the error

$$e(t) = z(t) - \hat{y}(t) \tag{4}$$

Since the objective is to make the vector $\hat{\boldsymbol{h}}$ approximate the vector $\boldsymbol{h}$, one might search for the vector $\hat{\boldsymbol{h}}$ that minimizes the expected value of the squared error $e^2(t)$. A natural way is to move $\hat{\boldsymbol{h}}$ in the direction opposite to the gradient of this expected error. Thus one might try the algorithm

$$\begin{aligned}\frac{d\hat{\boldsymbol{h}}}{dt} &= -\frac{\mu}{2}\nabla E\{[z(t) - \hat{y}(t)]^2\} \\ &= -\frac{\mu}{2}\nabla E[e^2(t)]\end{aligned} \tag{5}$$

where μ is a parameter that controls the rate of change, E denotes mathematical expectation, and ∇ is the gradient with respect to $\hat{\boldsymbol{h}}$. Equation (5) is just one form of gradient search for the location of the minimum of a function of several variables. What the stochastic gradient algorithm does is to replace the *expected value* of the squared error by the *instantaneous value.* As we shall see, even such a crude estimate of the gradient is adequate, under certain reasonable conditions, to make $\hat{\boldsymbol{h}}$ approach $\boldsymbol{h}$.

The stochastic gradient version of Eq. (5) is

$$\begin{aligned}\frac{d\hat{\boldsymbol{h}}}{dt} &= -\frac{\mu}{2}\nabla[e^2(t)] \\ &= -\mu e(t)\nabla[e(t)] \\ &= \mu e(t)\boldsymbol{x}(t)\end{aligned} \tag{6}$$

Figure 5 illustrates the block diagram of a circuit to implement the adaptation according to Eq. (6). The circuit shows an analog implementation. All current implementations are digital and are obtained by sampling all the functions at the appropriate (Nyquist) rate and replacing the derivative by a first difference. We will, however, start with the analog representation in this section. This is partly for historical reasons [the earliest echo canceler was implemented as an analog device (1)] but also because the basic properties are easiest to describe in the continuous version. Necessary modifications for the discrete-time case will be added later.

The circuit of Fig. 5 includes a function F that equals the identity function when implementing Eq. (6). The introduction of F allows one to handle a more general criterion than the squared error. For instance, if the expectation of the *magnitude* of the error is to be minimized, then $F(\cdot) = \text{sign}(\cdot)$ must be chosen. This choice of F has been used in some implementations of the algorithm, (see, e.g., Ref. 17). Another choice that has been recently shown to be useful is the ideal limiter (18).

Convergence in the Ideal Case. Suppose first that the echo path is perfectly stationary, the model represents the echo path exactly, and there is no noise or interrupting speech. Under these ideal conditions we have $z(t) = \boldsymbol{h}^T\boldsymbol{x}(t)$, and the error $e(t)$ is given by

$$\begin{aligned}e(t) &= (\boldsymbol{h} - \hat{\boldsymbol{h}})^T\boldsymbol{x} \\ &= \boldsymbol{\epsilon}^T\boldsymbol{x}\end{aligned} \tag{7}$$

where $\boldsymbol{\epsilon}$ is the misalignment vector. Since $\boldsymbol{h}$ is assumed constant, the time derivative of $\hat{\boldsymbol{h}}$ and $\boldsymbol{\epsilon}$ are identical except for the sign. Therefore, Eq. (6) can be rewritten as

$$\frac{d\boldsymbol{\epsilon}}{dt} = -\mu e(t)\boldsymbol{x} \tag{8}$$

Pre-multiplying both sides of Eq. (8) by $2\boldsymbol{\epsilon}^T$, and noting that $2\boldsymbol{\epsilon}^T(d\boldsymbol{\epsilon}/dt) = d\|\boldsymbol{\epsilon}\|^2/dt$, we get

$$\frac{d}{dt}\|\boldsymbol{\epsilon}\|^2 = -2\mu e^2(t) \tag{9}$$

which shows that the length of the misalignment vector $\boldsymbol{\epsilon}$ is nonincreasing. It is strictly decreasing as long as there is an uncanceled echo. Another important piece of information that can be gathered from Eq. (9) is that $e^2(t)$ eventually goes to zero. This is seen by integrating both sides of Eq. (9) with respect to t from 0 to T, which yields

$$\|\boldsymbol{\epsilon}(0)\|^2 - \|\boldsymbol{\epsilon}(T)\|^2 = 2\mu\int_0^T e^2(t)\,dt \tag{10}$$

Since the left-hand side is bounded by the initial value of $\|\boldsymbol{\epsilon}\|^2$, it follows that $e^2(T)$ must, in the limit, go to zero.

We cannot, however, be satisfied with the error going to zero; we want the error to be zero not only for the signal history up to the present, but for *all* subsequent signals. Hence, what we require is that the misalignment vector $\boldsymbol{\epsilon}$ should go to zero. Unfortunately, that is not provable even in the ideal situation considered in this section without imposing conditions on the input signal $x(t)$. The reason is that $e(t) = 0$ does not imply that $\boldsymbol{\epsilon} = 0$, but only that $\boldsymbol{\epsilon}$ is orthogonal to $\boldsymbol{x}$.

Sufficient conditions for the convergence of $\boldsymbol{\epsilon}$ to zero are derived in Ref. 19, and we will not discuss them here. However, intuitively speaking, the conditions assure that the time-varying vector $\boldsymbol{x}(t)$ does not stay confined to a subspace of dimension less than L for too long (i.e., $\boldsymbol{x}$ should evolve in time in such a way as to cover the entire L-dimensional space). In modern terminology, this is referred to as "persistent excitation" (Ref. 16, pp. 690–692).

Even when the persistent excitation condition is satisfied, it is a difficult matter to get accurate estimates of the *rate* of convergence of $\boldsymbol{\epsilon}$. Suppose, for instance, that $x(t)$ is a member of a stationary ergodic process. One would expect that in this case the expected convergence rate could be easily computed. This is not the case. If, for instance, the expectation of both sides of Eq. (9) is taken, it does not help because the right-hand side depends on $\boldsymbol{\epsilon}$ itself. However, if μ is very small, then one can assume that on the right-hand side, $\boldsymbol{\epsilon}$ and $\boldsymbol{x}$ are independent. (For small μ, $\boldsymbol{\epsilon}$ changes slowly, and one may assume that the expectation on the $\boldsymbol{x}$ ensemble, E_x, can be taken with $\boldsymbol{\epsilon}$ assumed quasi-constant.) This is known as the *independence assumption* (1), which can be justified rigorously as a first-order perturbation approximation (20). Under this assumption we see that, using Eq. (7),

$$E_x[e^2(t)] = E_x[\boldsymbol{\epsilon}^T \boldsymbol{x}\boldsymbol{x}^T \boldsymbol{\epsilon}] \\ = \boldsymbol{\epsilon}^T R \boldsymbol{\epsilon} \tag{11}$$

where $R \equiv E[\boldsymbol{x}\boldsymbol{x}^T]$ is the correlation matrix of $\boldsymbol{x}$. Then the expected value of Eq. (9) gives the exponentially decaying upper and lower bounds

$$\exp(-2\mu\lambda_{\max}t) \le E\|\boldsymbol{\epsilon}\|^2 \le \exp(-2\mu\lambda_{\min}t) \tag{12}$$

where $\lambda_{\max}$ and $\lambda_{\min}$ are, respectively, the maximum and minimum eigenvalues of R.

Fortunately, for convergence rates of interest in general, these bounds are useful. Nevertheless, it is important to remember that the bounds are *not* valid for large μ. For instance, if $\lambda_{\min} > 0$, the upper bound shown implies that $\boldsymbol{\epsilon}$ can be made to go to zero as fast as desired by merely increasing μ. This is not the case. In fact, the following simple argument shows that the convergence rate must start *decreasing* when μ is *increased* beyond a certain value. Note from Eq. (8) that $\boldsymbol{\epsilon}$ changes in a direction such as to make it more orthogonal to $\boldsymbol{x}$. If μ is so large that $\boldsymbol{\epsilon}$ can change much faster than $\boldsymbol{x}$, it is intuitively clear from Eq. (7) that $\boldsymbol{\epsilon}$ rapidly becomes perpendicular to $\boldsymbol{x}$. From there on, it stays perpendicular to $\boldsymbol{x}$ and hence does not change in length appreciably. (If $\boldsymbol{x}$ were a strictly constant vector, $\boldsymbol{\epsilon}$ would not change at all once it became perpendicular to $\boldsymbol{x}$.)

The argument of the last paragraph shows that the convergence rate goes to zero as $\mu \to \infty$, and it obviously goes to zero as $\mu \to 0$. Therefore, there is some optimum value of μ that gives the most rapid convergence. There is no known way to derive this optimum even for a simple (e.g., stationary ergodic) input signal $x(t)$, let alone a speech signal. A good setting can only be found experimentally. However, some theoretically derived bounds, and a more rigorous derivation of the intuitive arguments presented, may be found in Ref. 19.

Although the convergence rates are difficult to estimate, it is clear from Eq. (12) that the convergence rate can fluctuate quite a lot if the spread of eigenvalues of the correlation matrix R is large. To reduce these fluctuations, one would ideally want to "whiten" the speech signal (i.e., make all the eigenvalues equal and constant). Speech is a nonstationary signal whose spectral properties change much more rapidly than does the echo path. Whitening it, therefore, requires a fast adaptation, in addition to the adaptation to the echo path. However, one source of variability of the eigenvalues can be eliminated rather easily. This is the variability due to the change in signal level. Since the eigenvalues are proportional to the variance (or power) of the input signal, this objective can be accomplished by dividing the right-hand side of Eq. (6) by a local estimate of power. One simple way is to modify Eq. (6) to

$$\frac{d\hat{\boldsymbol{h}}}{dt} = \mu \frac{e(t)}{\boldsymbol{x}^T\boldsymbol{x}} \boldsymbol{x}(t) \tag{13}$$

All line echo cancelers in use today implement a discrete-time version of Eq. (13); that is,

$$\hat{\boldsymbol{h}}_{n+1} = \hat{\boldsymbol{h}}_n + \mu \frac{e_n}{\boldsymbol{x}_n^T\boldsymbol{x}_n} \boldsymbol{x}_n \tag{14}$$

where μ is a new constant and the subscript n indicates the value of a quantity at time t equal to n times the sampling interval. Because of the division by the input power, this algorithm is called the *normalized* LMS (NLMS) algorithm.

The discrete-time formulation also introduces a new element into the convergence problem: Due to the one-sample update delay, the algorithm can go unstable if the step size μ is increased beyond a certain value. Analysis of this stability condition is facilitated by again invoking the independence assumption. For the LMS algorithm [Eq. (14) without the normalizing denominator], making the independence assumption shows that convergence of the adaptive weight vector in mean is assured if $\mu < 2/\lambda_{\max}$ (15). However, convergence of the mean-square weight and mean-square error is somewhat more restrictive, requiring $\mu < (2/3)/\mathrm{tr}(R)$ (21). The NLMS algorithm of Eq. (14) can be interpreted as a projection that solves an underdetermined least mean square problem (Ref. 16, pp. 352–356), and for a stationary process, convergence in the first and second moment is guaranteed for $\mu < 2$ (22). This result has also been shown to hold for a (nonstationary) spherically invariant process (23), which has been suggested as a model for speech signals. For both LMS and NLMS, a good rule of thumb for achieving fastest convergence is to set the step size to about half of its maximum stable value [i.e., $\mu \approx (1/3)/\mathrm{tr}(R)$ for LMS and $\mu \approx 1$ for NLMS]. However, in practice, even smaller values are usually used to ensure stability in the presence of transient disturbances.

Convergence in the Nonideal Case. The convergence process, in practice, is even more complicated than described in the previous section. Detailed discussion of the nonideal case is beyond the scope of this article. However, in Refs. 19 and 24 it is shown that under essentially the same restrictions on $x(t)$, theoretical bounds can be derived in the nonideal case as well. If the only perturbation is an additive noise, then the vector $\boldsymbol{\epsilon}$ converges to lie within a sphere around the origin, whose radius is proportional to the root mean square (rms) value of the noise. If the echo path is not constant, then the radius of the sphere is also proportional to the rate of change of the impulse response.

The most severe situation arises during intervals of double talking (i.e., intervals during which the speech from speakers A and B is present simultaneously at the echo canceler). If the echo canceler has converged to a small misalignment, the interfering speech signal from B can be much louder than the uncanceled echo and can completely misalign the canceler in a very short time. About the only effective way of dealing with this problem is to use a system similar to the echo suppressor to detect the occurrence of double talking. However, instead of breaking the return path, just the adaptation loop is temporarily disabled during these intervals.

One of the most widely used double-talk detectors is the so-called Geigel algorithm (25), which declares the presence of near-end speech whenever

$$|y(n)| > \beta \max_{n-L \le m < n} |x(m)| \tag{15}$$

where β is a suitably chosen constant [e.g., 1/2 (-6 dB)]. The Geigel algorithm works fairly reliably for line echo cancelers where the hybrid echo return loss is reasonably well defined. However, for acoustic echo cancelers, selection of the threshold value β is more problematic because, for example, the return loss may be negative (a gain). Some recent advances in double-talk detection additionally make use of the correlation (26) or coherence (27) between x and y or between x and e. A proof of the equivalence of these techniques as well as a new normalized correlation technique appears in Ref. 28.

Other Algorithms

Before turning to a discussion of acoustic echo cancelers, let us briefly discuss four algorithms that attempt to improve on the simple stochastic gradient algorithm discussed so far: (1) a canceler based on two echo path models, (2) the least squares (LS) algorithm, (3) the recursive least squares (RLS) algorithm, and (4) the affine projection (AP) algorithm. None of these is in common use for line (or acoustic) echo cancellation, although prototypes have been implemented. The main reason why the two-path approach has not found wide application is the difficulty in designing a decision algorithm needed for its implementation, as well as the additional memory requirements. As for the LS and RLS algorithms, the main reason is the added computational complexity. The AP algorithm attempts to bridge the gap in complexity between the LS and RLS algorithms and the much simpler LMS and NLMS algorithms. Also, there are now fast recursive least-square (FRLS) and fast affine projection (FAP) algorithms. In view of the rapid advancement of digital technology, these and other more complex algorithms will, no doubt, be used in the near future.

Two Echo Path Models. In the discussion of the stochastic gradient algorithm, we defined the error signal $e = z - \hat{\boldsymbol{h}}^T\boldsymbol{x}$, which was used to control the adaptation of the cancelling filter $\hat{\boldsymbol{h}}$. In the usual implementation of the echo canceler, this same error signal is the one sent to the remote station as the echo-canceled signal. However, this is not imperative. The returned signal could be derived differently, say as $\tilde{e} = z - \tilde{\boldsymbol{h}}^T\boldsymbol{x}$, where the filter $\tilde{\boldsymbol{h}}$ is derived from, but not identical to, $\hat{\boldsymbol{h}}$. In Ref. 29, it is argued that this observation can be used to advantage. Since the gradient used in the LMS algorithm is only a very crude estimate of the gradient of the mean squared error, every adaptive step does not necessarily improve $\hat{\boldsymbol{h}}$. The authors' suggestion is to monitor both e and $\tilde{e}$ and to copy the coefficients of $\hat{\boldsymbol{h}}$ into $\tilde{\boldsymbol{h}}$ whenever $\hat{\boldsymbol{h}}$ performs consistently better than $\tilde{\boldsymbol{h}}$ over some specified time interval. The decision when to transfer coefficients may be based on observation of e, $\tilde{e}$, $\hat{\boldsymbol{h}}$, $\tilde{\boldsymbol{h}}$, $\boldsymbol{x}$, and z. The added memory and computational requirements have discouraged use of this algorithm in the past, although it has been incorporated into a few products. Continued reduction in the cost of digital signal processing should make this algorithm find widespread application.

The Least Squares Algorithm. Previously we noted that the impulse response $\hat{\boldsymbol{h}}$ can be estimated by minimizing the expectation of the squared error, e^2. The stochastic gradient algorithm sidesteps the problem of estimating the *expected* value of e^2 by taking an incremental step in the direction that reduces its *instantaneous* value. Instead of this, the LS algorithm minimizes a better deterministic approximation to the expected value. Specifically, considering the discrete-time formulation, the LS algorithm computes the vector $\hat{\boldsymbol{h}}$ that minimizes the arithmetic mean of e_m^2 for some range of sampling instants m. This is equivalent to minimizing

$$\xi_n = \sum_{m=n-M+1}^{n} e_m^2 \tag{16}$$

Upon substituting for e in terms of $\hat{\boldsymbol{h}}$ and $\boldsymbol{x}$, the problem reduces to minimizing

$$\xi_n = \hat{\boldsymbol{h}}_n^T X_n^T X_n \hat{\boldsymbol{h}}_n - 2\boldsymbol{p}_n^T \hat{\boldsymbol{h}}_n + \sum_{m=n-M+1}^{n} z_m^2 \tag{17}$$

Here X_n is an $M \times L$ matrix whose mth row ($1 \le m \le M$) is $\boldsymbol{x}_{n-M+m}^T$, and the vector

$$\boldsymbol{p}_n \equiv \sum_{m=n-M+1}^{n} z_m \boldsymbol{x}_m \tag{18}$$

Setting the gradient of ξ_n in Eq. (17) to zero shows that $\hat{\boldsymbol{h}}_n$ is the solution of

$$X_n^T X_n \hat{\boldsymbol{h}}_n = \boldsymbol{p}_n \tag{19}$$

and can be computed by a single matrix inversion. Since the matrix to be inverted is of size $L \times L$, its inversion would ordinarily require $O(L^3)$ computations. (Recall that L is the length of the vector $\hat{\boldsymbol{h}}_n$.) However, if the adaptive structure is a transversal filter, then the matrix X_n is Toeplitz (i.e., all entries on any diagonal are identical). Taking advantage of this property, the inversion can be performed in $O(L^2)$ operations (30).

The solution depends on both n and M. The matrix $X_n^T X_n$ is invertible if and only if X_n has independent columns. If this is not the case, then the solution is not unique, and a pseudoinverse of $X_n^T X_n$ can be used to select the minimum norm solution. If the input and the echo path were stationary, the dependence on n would be eliminated, and a single matrix inversion would be required. However, because of the time variations of $\boldsymbol{h}_n$, the solution $\hat{\boldsymbol{h}}_n$ must be updated as often as the available hardware allows.

The Recursive Least Squares Algorithm. The least squares algorithm is a *block* processing algorithm. An optimum estimate of $\boldsymbol{h}$ is derived from a block of data (of length M in the preceding description). This optimum estimate is assumed to be valid until the next block of data is processed to give a new estimate of $\boldsymbol{h}$, and so on. There is an alternative algorithm in which an optimal estimate of $\boldsymbol{h}$ is obtained recursively at *every* time instant. The algorithm is a deterministic version of the Kalman filter. At every instant, the estimate $\hat{\boldsymbol{h}}_n$ minimizes a weighted sum of the squared errors at all past instants of time. To be able to track slowly varying impulse responses, the weighting is chosen such that errors in the remote past do not affect the current estimate. Recursive algorithms can be derived for several weighting functions that achieve this objective. One convenient error measure is

$$\xi_n = \sum_{m=-\infty}^{n} \lambda^{n-m} e_m^2 \tag{20}$$

with λ chosen in the range $0 \leq \lambda < 1$. The value chosen for λ determines the effective duration of the past input that is used to derive the current estimate $\hat{\boldsymbol{h}}_n$.

In terms of $\boldsymbol{x}$ and $\hat{\boldsymbol{h}}_n$, Eq. (20) can be rewritten as

$$\xi_n = \hat{\boldsymbol{h}}_n^T R_n \hat{\boldsymbol{h}}_n - 2\boldsymbol{p}_n^T \hat{\boldsymbol{h}}_n + \sum_{m=-\infty}^{n} \lambda^{n-m} z_m^2 \tag{21}$$

where the matrix

$$R_n \equiv \sum_{m=-\infty}^{n} \lambda^{n-m} \boldsymbol{x}_m \boldsymbol{x}_m^T \tag{22}$$

and the vector

$$\boldsymbol{p}_n \equiv \sum_{m=-\infty}^{n} \lambda^{n-m} z_m \boldsymbol{x}_m \tag{23}$$

Thus at time n, the optimal impulse response vector $\hat{\boldsymbol{h}}_n$ is the solution of

$$R_n \hat{\boldsymbol{h}}_n = \boldsymbol{p}_n \tag{24}$$

From the definitions of R_n and $\boldsymbol{p}_n$ it is straightforward to show that they satisfy the recursions

$$R_n = \lambda R_{n-1} + \boldsymbol{x}_n \boldsymbol{x}_n^T \tag{25}$$

and

$$\boldsymbol{p}_n = \boldsymbol{p}_{n-1} + z_n \boldsymbol{x}_n \tag{26}$$

Because of the recursion in Eq. (25), R_n^{-1} can be obtained by updating R_{n-1}^{-1} through use of the matrix inversion lemma (Ref. 16, p. 480). The optimal estimate $\hat{\boldsymbol{h}}_n$ is thus obtained recursively from $\hat{\boldsymbol{h}}_{n-1}$. The update algorithm is rather cumbersome, although simple in principle. We refer the reader to Ref. 16, Chapter 13 for details.

Recursion based on Eqs. (25) and (26) requires $O(L^2)$ operations per iteration. Although much less than the $O(L^3)$ computations that would be required without the recursions, the computational load is still much more than the $2L$ multiplications per iteration required by the LMS algorithm. The advantage gained by the extra computations is the highly improved convergence. We note, however, that under certain circumstances, the tracking performance of the RLS algorithm may not be improved over that the LMS algorithm (31).

As in the case of the LS algorithm, the computational requirements can be reduced dramatically if the adaptive structure is a transversal filter. Algorithms that accomplish this, known as fast RLS or fast transversal filter algorithms, have been developed in Refs. 32 through 34 and others. They achieve the good convergence properties of the RLS algorithm at a computational requirement that grows linearly with L.

The main limitation of these fast algorithms is that they tend to be numerically unstable unless multiple precision arithmetic is used. In practical implementations the algorithms have to be periodically reset. Nevertheless, as mentioned later, the fast transversal filter algorithm has recently been successfully implemented for subband acoustic cancellation. Further progress on stabilization of the fast RLS algorithm has been reported in Ref. 35.

Affine Projection Algorithms. An algorithm was introduced in Ref. 36 based on affine projections of the most recent M data vectors and is the basis for algorithms that converge rapidly for autoregressive (AR) processes of order less than or equal to M and with numerical complexity ML, which is intermediate between that of LMS and RLS. The affine projection algorithm can be viewed as a generalization of the NLMS algorithm ($M = 1$) and can be embellished and interpreted from various viewpoints (37). A numerically efficient implementation of the algorithm appears in Refs. 38 and 39.

SINGLE-CHANNEL ACOUSTIC ECHO CANCELLATION

The problem of canceling acoustic echoes in hands-free telephony and teleconferencing differs from the cancellation of line echoes mainly because of the different nature of the echo paths (40). Instead of the mismatch of the hybrid, a loudspeaker-room-microphone system needs to be modeled in these applications (Fig. 6). As with line echoes, echo suppressors can be employed, but for reasons discussed later they are even less satisfactory in this regime. Accordingly, this section will focus mainly on acoustic echo cancellation.

Comparing typical impulse responses of echo paths for line echoes and acoustic echoes, it becomes obvious that acoustic echo cancellation is a far more challenging task than line echo cancellation. The duration of the impulse response of the acoustic echo path is usually several times longer (100 ms to 400 ms) and it may change rapidly at any time (e.g., due to an opening door or a moving person). Achieving even a modest

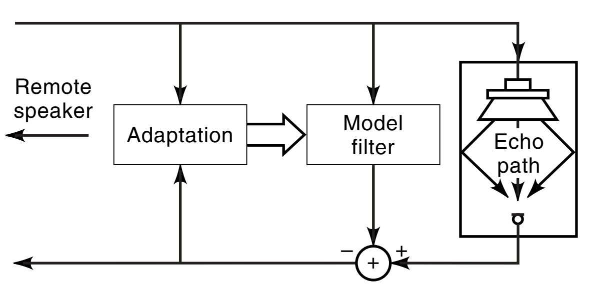

Figure 6. An acoustic echo canceler is used to cancel echoes that arise from coupling between a loudspeaker and microphone.

improvement (say, a misalignment 20 dB below the uncanceled room response) requires a transversal filter with over 1000 taps at an 8 kHz sampling rate for a typical office (40). Some commercial products foresee requiring as many as 4000 taps at a 16 kHz sampling rate. In Ref. 41, a method is proposed for reducing the number of parameters by modeling the acoustic echo paths as a recursive, or infinite impulse response (IIR), filter with common acoustical poles and zeros. However, this appears to be useful only at low frequencies below about 1 kHz. Therefore, adaptive transversal, or finite impulse response (FIR), filters are still the preferred choice.

For line echo cancellation, it is sometimes possible to switch to a slower-converging tracking mode after an initial rapid convergence (with, for example, a training signal). This is because for certain circuits it is possible to assume that the echo path varies only very slowly. As mentioned previously, this is not a good assumption for acoustic echo cancellation, so one cannot switch to a slowly converging algorithm even in the tracking mode. Indeed, some theoretical models cast doubt as to whether the LMS algorithm can ever satisfactorily track typical acoustic changes in a room (42), and as previously mentioned, the RLS algorithm may fare even worse in this respect (31).

We see, therefore, that acoustic echo cancelers require more computing power than line cancelers for two reasons: first, just because of the length of the impulse response to be compensated, and second, because faster converging algorithms are desirable. Current products (and prototypes) only partly meet these requirements. This is mainly because they use the LMS type of adaptation algorithms. Such algorithms are known to perform poorly for long impulse responses and with speech as the input signal (16).

As mentioned previously, LS and RLS algorithms can increase convergence speed. They have recently been successfully used to implement line echo cancelers. However, they are still infeasible for acoustic echo cancellation because of the long impulse responses involved. Other methods have therefore been explored to improve the convergence speed of LMS-type algorithms. One direction aims at "whitening" the speech signal for the adaptation. A simple way to do this is to employ a continuously updated first-order linear predictor to prewhiten the reference signal x used to update the adaptive weight vector (43), and this leads to faster convergence with little increase in complexity. Another direction foresees time-varying step size factors for the different taps of the LMS-adapted FIR filter. A general method for doing this, not specifically tailored for the acoustic echo cancellation problem, was suggested in Ref. 44. Another approach exploits the structure of the impulse response of the acoustic echo path and assigns different step sizes to different sections of the echo path impulse response (1,45). The idea is that, ideally, the impulse response samples with large values are adapted with large step size while those with small values get a small step size. This should result in a faster overall convergence. Obviously, the efficiency of this method depends highly on the a priori knowledge concerning the current echo path impulse response and the ability to adjust the step size accordingly.

As previously mentioned, echo suppressors by themselves are unsatisfactory for acoustic echoes. However, such techniques can be used to augment acoustic echo cancellation in suppressing residual echo. These techniques include the concept of "center clipping" (11) and frequency-selective suppression using low-order adaptive transversal filters (46,47). Selective echo suppression can also be applied within the structure of subband acoustic echo cancelers (11), to be discussed later.

All of the aforementioned approaches attempt to improve convergence speed without adding too much computational complexity. Recently, subband techniques have been developed to reduce the computational complexity of acoustic echo cancelers with long impulse responses, while at the same time providing more favorable circumstances for fast convergence.

Subband Approach

Subband structures were first proposed in 1984 independently in Ref. 48 for teleconferencing purposes, and in Ref. 49 as a general framework for acoustic echo cancellation. The fundamental structure is depicted in Fig. 7. With reference to that figure, it is seen that the input and output signals of the echo path are passed through identical analysis filter banks A, producing vectors of M subband signals that are sampled at a reduced rate. The cancellation structure C forms a vector of subband signals $\hat{\boldsymbol{y}}$ to approximate the corresponding subband echo signals $\boldsymbol{y}$. The resulting subband errors $\boldsymbol{e}$ are passed through a synthesis system S to give a full-band signal e, which is then transmitted back to the remote loudspeaker. The adaptation block utilizes the vector of subband error signals and input signals to adjust the input-output characteristics of the cancellation unit so as to drive the vector of error signals toward zero.

Intuitively, the promise of the subband structure is twofold: On the one hand, the computational complexity is reduced because of the downsampling, and on the other hand, the convergence of the adaptation algorithms is expected to speed up for LMS-type algorithms because of the decomposition of the input speech signal. To illustrate the advantage in computational complexity, consider the case in which the canceler has M parallel adaptive filters, and all subband signals are downsampled by the same factor R. Assume that the impulse response of each of the subbands has the same length

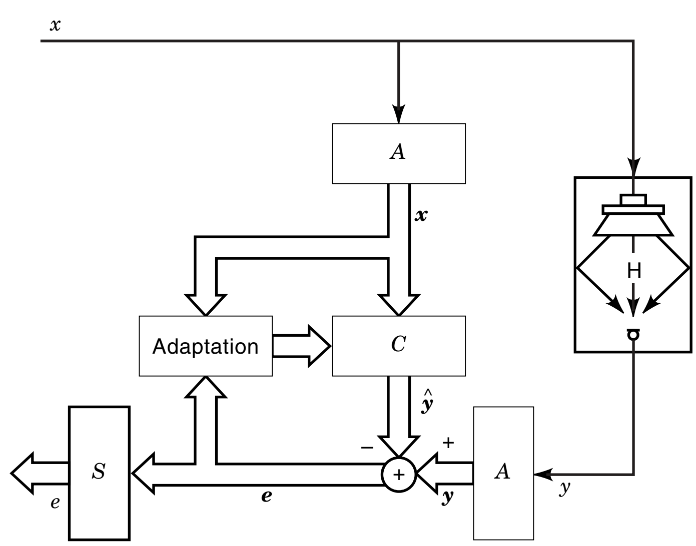

Figure 7. Block diagram of a subband echo canceler showing analysis (A) and synthesis (S) filter banks, subband cancellation unit (C), and adaptation control.

as the full-band impulse response. Then the number of coefficients in each band needed for its representation is fewer by a factor R compared to the full-band representation because of the downsampling. Further, filtering and adaptation of the subband cancelers is performed at the reduced sampling rate. Therefore, the computational complexity (measured as computations per second) for one subband canceler is $1/R^2$ that of the full-band canceler. Taking into account all M subbands, the complexity can be expected to be reduced by a factor of approximately R^2/M, assuming that the computational load for the analysis and synthesis systems is negligible. The reduction in computational complexity can be exploited in several ways: The overall system bandwidth or the duration of the impulse response to be modeled can be increased, more complex adaptation algorithms can be employed, or, most obviously, hardware can be saved to reduce cost. The price paid for these advantages can also be inferred from a comparison of Figs. 6 and 7: In the subband canceler the microphone signal is delayed before it is sent to the far end by the group delay of the cascade of analysis and synthesis filters. Means for dealing with this problem of delay will be described later.

Before discussing possible choices for the components of the subband echo canceler, let us temporarily set aside the problem of adaptation algorithms and concentrate on defining structures that are capable of performing the required cancellation. Once we have identified appropriate structures, we will take up the question of adaptation algorithms suitable in the context of acoustic echo cancellation.

Assuming that all systems involved are linear and considering a fixed but arbitrary instant in time, a matrix notation in the z-transform domain can be used to describe the subband structure conveniently (50). For analysis and synthesis we allow the general class of systems that can be represented as filter banks with time-invariant filtering and downsampling or upsampling, respectively (see, e.g., Refs. 51 and 52). Although extension of the description to nonuniform downsampling factors is possible (53), we use here the same downsampling factor R for all subbands for the sake of simplicity. Then the general cancellation condition for nulling the local signals,

$$\boldsymbol{e}(t) \equiv \boldsymbol{0} \tag{27}$$

implies that (50)

$$S(z)C(z^R)A(z)\boldsymbol{x}(z) = S(z)A(z)H(z)\boldsymbol{x}(z) \tag{28}$$

where the matrices $A(z)$, $S(z)$, $H(z)$, $C(z^R)$ correspond to analysis, synthesis, echo path, and cancellation unit, respectively, and the vector $\boldsymbol{x}(z)$ corresponds to the input signal. For signal-independent solutions we must have

$$S(z)C(z^R)A(z) = S(z)A(z)H(z) \tag{29}$$

Solutions satisfying Eq. (29) are called *synthesis dependent* because, in general, the fact that the full-band error is zero does not imply that the vector of subband errors is also zero. The latter condition is met by the subset of solutions of Eq. (29) that satisfy the stricter condition

$$C(z^R)A(z) = A(z)H(z) \tag{30}$$

Solutions of Eq. (30) are called *synthesis independent.*

Synthesis-dependent implementations can be very efficient when the impulse responses are long since orthogonal transforms can be utilized (54). However, a major shortcoming of this technique is that the error information is available to the adaptation algorithm only after having been delayed by the synthesis filter bank. This has a deleterious effect on convergence, especially when the echo path response is changing rapidly. Another disadvantage of this configuration is apparent when we observe that the cancellation unit is a set of M parallel adaptive filters each with a single input and a single output, and the adaptation algorithm must adapt all these filters on the basis of a common error signal. The components of the error signal outside the frequency range of each filter thus act as a noise on the adaptation process for that filter. Therefore, the synthesis-independent solution of Eq. (30) is the most useful structure in practice.

Much work has been carried out to develop cancelers using only two or four bands (43,48,55,56). However, other authors have tried to explore the concept to a greater extent (50,53,57–63). The previously mentioned whitening technique can also be applied in subbands (64).

Let us consider some of the implications of solutions to the cancellation condition, Eq. (30). As before, assuming that the cancellation unit consists of a set of M parallel (adaptive) filters, Eq. (30) has analytically well-defined solutions for given analysis/synthesis systems, as derived in Refs. 50 and 53. There, it is shown that solutions exist if and only if the analysis produces aliasing-free subband signals. Obviously, this requires bandpass filters with infinite stop-band attenuation. Also, the downsampling factor R must be chosen such that the passband and transition region of the modulated versions of the analysis bandpass do not overlap in the frequency domain after downsampling. If this requirement is met, each subband canceler has to model an ideally bandpass-filtered and downsampled version of the full-band echo path impulse response. The passband of the ideal bandpass should cover the transition and passband region of the corresponding analysis bandpass. However, this solution can obviously not be realized in the strict sense. First, bandpass filters with infinite stop-band attenuation can only be approximated and, second, filtering a finite impulse response with an ideal bandpass leads to an impulse response that extends from $-\infty$ to $+\infty$ in the time domain and, therefore, is not realizable either. Thus, two approximation problems are linked to the realization of the frequency-subband concept, and it is important to consider the extent to which the approximation errors impair its usefulness. This is briefly outlined in the next section, where the minimum misalignment that can be obtained for a given frequency subband configuration is considered (for further details see Refs. 53 or 60).

Approximation of the Subband Solution. Let us consider the approximation of Eq. (30) that is possible with ideal settings of the cancelers. In the next section we will consider the adaptation aspects. We will use a quadratic distance measure to estimate system misalignment. This may be interpreted as the transmission loss factor of the corresponding system with white noise as the input signal. The misalignment of the whole structure has two components: the misalignment due to the residual aliasing within the subband signals and the misalignment of the subband cancelers due to the truncation of their impulse responses. Numerical investigation shows that the overall system misalignment is dominated by the truncation effect until it reaches a certain threshold (e.g., about -25 to -30 dB for $M = 16$ and $R = 12$) (60). For smaller misalignment, the truncation errors of the subband cancelers have little impact and the misalignment of the whole system is determined mainly by the aliasing caused by the finite stop-band attenuation of the analysis filter. The reason for this becomes obvious from the following consideration of the spectral distribution of the truncation error.

If only a small number of coefficients is used to approximate the response of the ideal subband canceler, the truncation will cause more or less uniformly distributed deviations from the ideal frequency response of the subband canceler. Beyond a certain number of coefficients, the truncation error in the inner region of the subband will be small and the misalignment of the subband canceler is mainly determined by the amount of error concentrated at the band edges, where the steep slope has to be approximated. This contribution to the misalignment of the subband canceler, however, has little influence on the misalignment of the overall system. This is because the misalignment due to the subband canceler is heavily weighted down at the band edges by the analysis and synthesis bandpass filters. Thus, once a certain level of truncation error at the band edges has been achieved, there is little to be gained by increasing the length of the subband canceler. This observation is important for an efficient design of the canceler: It implies that the delay needed to model noncausal coefficients of the subband canceler impulse response can be kept small. Once the truncation error has been reduced to a low value, the stop-band attenuation controls the residual misalignment. The stop-band attenuation must be chosen sufficiently high so as to keep the misalignment due to aliasing small.

The effect of the aliasing within the subband signals also explains the problems with the otherwise attractive choice of critical sampling ($R = M$) (62). Perfect reconstruction filter banks with critical sampling cause the transition regions of the analysis filters to be aliased into the subband signals. This causes severe misalignment of the overall subband structure (62). For this reason it is necessary to choose $R < M$ to decrease the aliasing of the transition regions, even though this sacrifices some computational efficiency.

Adaptation of the Subband Structure. As pointed out earlier, the canceler in each subband is essentially independent of the others. Therefore, the subband structure can utilize any of the adaptation algorithms developed for full-band cancelers. We will discuss a few of them, emphasizing the differences compared to their application in full-band systems.

To date, the normalized LMS algorithm is the most widespread adaptation algorithm for both full-band and subband systems. For this algorithm, it has been observed that the initial convergence is faster in the subband canceler compared to the full-band implementation. As discussed previously, the convergence speed of the LMS algorithm depends directly on the eigenvalue spread of the autocorrelation matrix of the input signal. Therefore, it is often concluded that the eigenvalue spread of the subband signals must be smaller than that of the full-band signal. However, examining the actual eigenvalues shows that their spread is in fact larger for the subband signals than for the full-band signal. This is because the subband signals are not really "whiter." The slopes of the frequency responses of the analysis filters cause notches in the subband spectra at the band edges, thus creating some very small eigenvalues. Indeed, the convergence behavior of the subband canceler itself is, in general, not better for a subband signal than for the full-band signal. The improvement that is observed for the subband structure as a whole is the result of the same masking property that affects the truncation errors: The overall system misalignment at the band edges of the subband canceler spectra get little weight due to the characteristics of the analysis and synthesis filters. These parts of the subband canceler spectra are the same ones that are affected by the small eigenvalues of the subband signal, and, therefore, correspond to the slower converging modes. But as these modes have little influence on the overall system misalignment, the subband canceler initially converges faster than a full-band system because the eigenvalue spread in the inner region of the subband spectra is indeed smaller than that of the full-band signal. The increased convergence speed for "nonwhite" input signals is, therefore, the result of an only indirect "whitening" effect. The residual band-edge components, however, cause slow asymptotic convergence (65). As a solution to this problem, it was suggested in Ref. 65 that the analysis filter bandwidth be increased so as to push out the band-edge energy beyond the passband of the synthesis filter, thereby eliminating slowly converging components. This idea was subsequently developed and demonstrated in Ref. 66.

Skipping all the incremental improvements that are possible to speed up the normalized LMS algorithm in the subband structure (see the section on full-band approaches), let us mention briefly the most sophisticated adaptation algorithm used so far for acoustic echo cancellation. This is the RLS implementation, as proposed in Ref. 61. In this implementation, the number of subbands is $M = 16$, the decimation rate is $R = 13$, and the subband filters have about 100 taps each. The adaptation algorithm is a modified version of the fast transversal filter algorithm of Ref. 34 mentioned previously. As discussed in that section, such algorithms need to be periodically reinitialized to guarantee numerical stability. To avoid sudden large increases of misalignment, the subband filters are reset one at a time. The computational complexity is only on the order of a full-band implementation of the LMS algorithm, whereas the initial convergence speed is comparable to that of a full-band RLS algorithm.

Extensions of the Subband Structure. There are at least two areas in which subband cancelers could be improved. First, the cancelers discussed previously realize the cancellation condition of Eq. (30) with a bank of adaptive filters, in which aliasing is kept small by choosing a decimation rate, R less than the number of filters M. The choice $R = M$ is not recommended for $M > 4$, because of the spectral gaps needed to avoid aliasing. In an attempt to realize this maximum decimation, a structure is proposed in Ref. 62 that uses adaptive cross filters to cancel the influence of aliasing in each subband. However, it appears doubtful that this structure would have satisfactory convergence properties, except, perhaps, in some special cases. The main subband filter and the corresponding cross filters have to use the same subband error signal to steer their respective adaptation by the LMS algorithm. As such, the adaptation algorithms for the different filters have no indication of the contribution of each to the

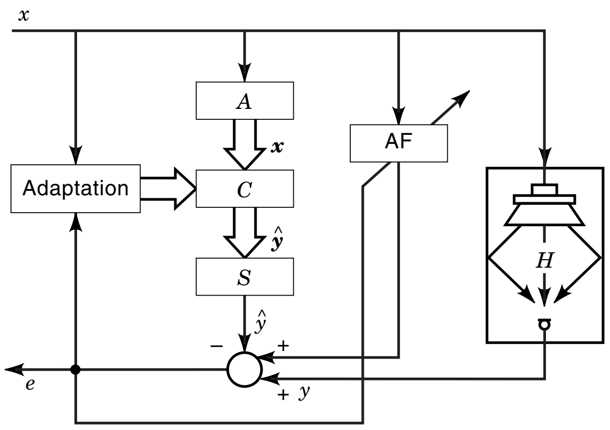

Figure 8. A short wide-band adaptive filter (AF) is used to eliminate delay in the signal path of a subband canceler.

total error. Therefore, it appears to be difficult to achieve stable and fast initial convergence. Also, the fact that additional cross filters have to be adapted reduces the computational efficiency, and the gain over a slightly oversampled system (e.g., $R = 3M/4$) might be marginal.

One undesirable aspect of subband cancelers is the delay introduced into the path from the near-end talker to the remote listener. (The delay in the analysis and synthesis filters shown in the bottom portion of Fig. 7.) This delay may be 10 ms to 20 ms, giving a round-trip delay that is twice that value. In some applications this much delay may not be tolerable. An interesting variant of the subband canceler, which eliminates this delay, has been proposed in Ref. 58 and is depicted in Fig. 8. The analysis and synthesis filters are removed from the return path, and an extra synthesis filter bank is included in the path that generates the cancelling signal. Notice that the delay in the return path has been eliminated; however, because of the delay in the adapted path, it is clear that the initial portion of the impulse response cannot be cancelled. Therefore, an additional full-band adaptive filter (marked AF in Fig. 8) is necessary. With that filter included, the system is, in principle, able to cancel the echoes without introducing delay. Note, however, that the adaptation algorithm is now forced to work with the full-band error signal. As mentioned previously, this structure has certain disadvantages. In particular, an inherent delay is introduced in the feedback loop. This might be a drawback in applications in which the impulse response is changing rapidly.

Another technique to eliminate signal path delay, while retaining the computational advantages of subband processing, was introduced in Ref. 67. A block diagram appears in Fig. 9. Here the adaptive weights are computed in subbands but are then transformed to an equivalent full-band FIR filter, thereby eliminating any delay in the signal path. This configuration is open loop in the sense that the error is derived in subbands independent from the full-band output error. Alternatively, a closed-loop version is also available that converges somewhat slower but is capable of completely eliminating the effects of aliasing (67).

MULTICHANNEL ACOUSTIC ECHO CANCELLATION

Until now, we have only considered single-channel acoustic echo cancellation, which is the most prevalent in current usage. However, there are many new applications in which multichannel sound (e.g., stereo) is envisioned to provide an ever more lifelike and transparent audio/video medium. These applications include multiparty room-to-room conferencing, multiparty desktop conferencing, and interactive video gaming involving multichannel sound. In these multichannel applications, there are multiple acoustic paths from multiple loudspeakers to multiple microphones, and echos arising from these paths must be cancelled for full-duplex communication. As we shall see, there are unexpected complications with multichannel sound that require special treatment.

As an introduction to the fundamental problem of multichannel acoustic echo cancellation, consider the room-to-room stereo conferencing scenario of Fig. 10 (68). A transmission room is depicted on the right, wherein two microphones are employed to pick up signals from a talker via two acoustic paths that are characterized by the impulse responses g_1 and g_2. (For convenience, all acoustic paths are assumed to include loudspeaker and/or microphone responses.) These stereophonic signals are then transmitted to loudspeakers in the receiving room on the left, which in turn are coupled to one of the microphones via the paths indicated with impulse responses h_1 and h_2, producing an outgoing signal y. (Similar paths couple to the other microphone in the receiving room, but for simplicity, only echo cancellation for the one microphone signal will be discussed here; similar remarks will apply to the other microphone signal.)

The previously discussed single-channel acoustic echo canceler is thus generalized using two adaptive FIR filters $\hat{h}_1$ and $\hat{h}_2$ to model the two echo paths in the receiving room. Driving these filters with the loudspeaker signals x_1 and x_2 produces an estimate $\hat{y}$ that is subtracted from the echo signal y to form an error signal e, which is intended to be small in the absence of near-end speech (i.e., speech generated in the receiving room).

To generalize the LMS algorithm for stereo acoustic echo cancellation, let the echo signal be expressed as (68)

$$y(t) = \boldsymbol{h}_1^T \boldsymbol{x}_1 + \boldsymbol{h}_2^T \boldsymbol{x}_2 \tag{31}$$

where $\boldsymbol{h}_1$ and $\boldsymbol{h}_2$ are L-dimensional vectors of the loudspeaker-to-microphone impulse responses in the receiving room, and $\boldsymbol{x}_1 \equiv [x_1(t), x_1(t-1), \ldots, x_1(t-L+1)]^T$ and $\boldsymbol{x}_2 \equiv [x_2(t), x_2(t-1), \ldots, x_2(t-L+1)]^T$ are vectors comprising the L most recent loudspeaker signal samples. The error signal is then written as

$$e(t) = y(t) - \hat{\boldsymbol{h}}_1^T \boldsymbol{x}_1 - \hat{\boldsymbol{h}}_2^T \boldsymbol{x}_2 \tag{32a}$$

where $\hat{\boldsymbol{h}}_1$ and $\hat{\boldsymbol{h}}_2$ are L-dimensional vectors of the adaptive filter coefficients.

The error signal can be written more compactly as

$$e(t) = y(t) - \hat{\boldsymbol{h}}^T \boldsymbol{x} \tag{32b}$$

with

$$y(t) = \boldsymbol{h}^T \boldsymbol{x} \tag{33}$$

where $\hat{\boldsymbol{h}} \equiv [\hat{\boldsymbol{h}}_1^T | \hat{\boldsymbol{h}}_2^T]^T$ is the concatenation of $\hat{\boldsymbol{h}}_1$ and $\hat{\boldsymbol{h}}_2$ and, likewise, $\boldsymbol{h} \equiv [\boldsymbol{h}_1^T | \boldsymbol{h}_2^T]^T$ and $\boldsymbol{x} \equiv [\boldsymbol{x}_1^T | \boldsymbol{x}_2^T]^T$. In terms of $\boldsymbol{h}$, we can

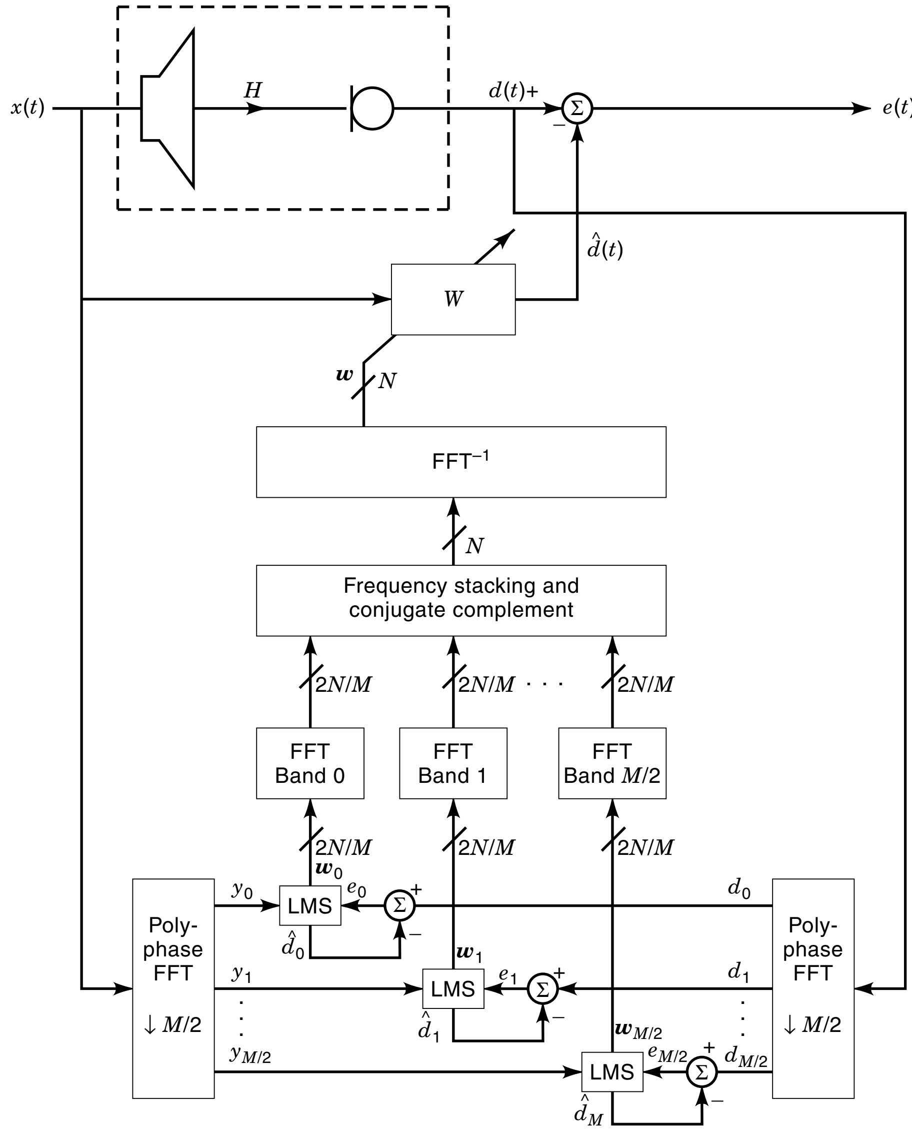

Figure 9. Another technique, called a delayless subband echo canceler, eliminates signal path delay by transforming subband adaptive weights ($\boldsymbol{w}_0$, $\boldsymbol{w}_1$, . . ., $\boldsymbol{w}_{M/2}$) to equivalent wideband filter weights ($\boldsymbol{w}$).

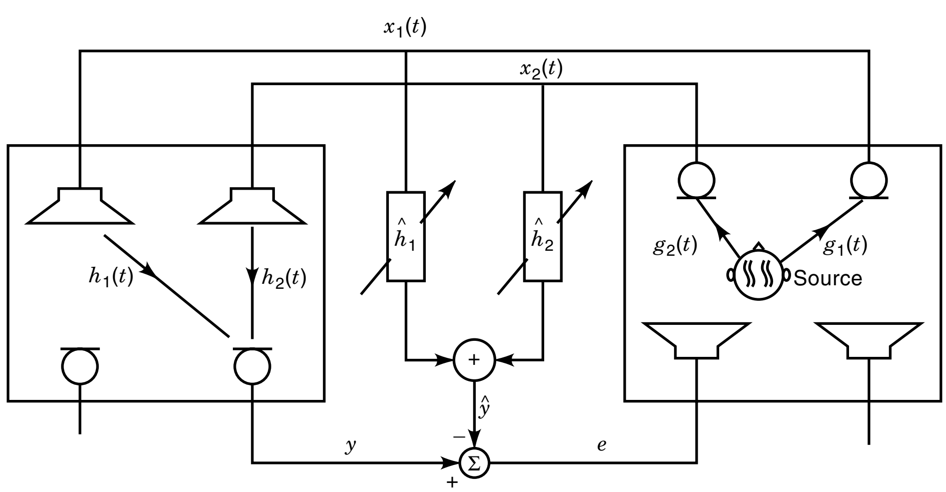

Figure 10. Schematic diagram of stereophonic echo cancellation, showing the use of adaptive filters $\hat{h}_1$ and $\hat{h}_2$ to cancel echo y arising from echo paths h_1 and h_2 from two loudspeakers on the left to one of the microphones.

rewrite Eq. (32b) as

$$e(t) = (\boldsymbol{h} - \hat{\boldsymbol{h}})^T \boldsymbol{x} = \boldsymbol{\epsilon}^T \boldsymbol{x} \tag{34}$$

where

$$\boldsymbol{\epsilon} \equiv \boldsymbol{h} - \hat{\boldsymbol{h}} \tag{35}$$

is the composite misalignment vector. With this notation, and making the transition to discrete time, the two-channel NLMS algorithm can be expressed as

$$\hat{\boldsymbol{h}}_{n+1} = \hat{\boldsymbol{h}}_n + \mu \frac{e_n}{\boldsymbol{x}_n^T \boldsymbol{x}_n} \boldsymbol{x}_n \tag{36}$$

where μ is the adaptive step size.

For the adjustment of the adaptive filter, Eq. (36) is formally identical to the adjustment of the single-channel echo canceler discussed previously. It therefore follows that the error $e(t)$ eventually converges to zero for suitable choice of the step size μ. However, $e(t) \to 0$ does not necessarily imply that $\|\boldsymbol{\epsilon}(t)\| \to 0$ (i.e., $\hat{\boldsymbol{h}} \to \boldsymbol{h}$), which is the primary goal of the adaptive filter. The importance of requiring that $\hat{\boldsymbol{h}} \to \boldsymbol{h}$ was discussed previously for the single-channel case; as we will see, this becomes even more critical in the case of multichannel cancelers.

Misalignment: The Nonuniqueness Problem

The main new feature that distinguishes stereo echo cancelers from conventional single-channel cancelers can be explained even without considering the "control" aspects of the adaptation algorithm. Therefore, setting aside the important question of *how* convergence is achieved, let us for the moment just assume that $e(t)$ has been driven to be identically zero. From Eq. (34), it follows that

$$\epsilon_1 * x_1 + \epsilon_2 * x_2 = 0 \tag{37}$$

where ϵ_1 and ϵ_2 are components of the misalignment corresponding to $h_1 - \hat{h}_1$ and $h_2 - \hat{h}_2$, respectively. For the single-talker situation depicted in Fig. 1, this further implies

$$[\epsilon_1 * g_1 + \epsilon_2 * g_2] * s(t) = 0 \tag{38}$$

where $s(t)$ is the acoustic signal generated by the talker. In the frequency domain, Eq. (38) becomes

$$[\varepsilon_1(j\omega)G_1(j\omega) + \varepsilon_2(j\omega)G_2(j\omega)]S(j\omega) = 0 \tag{39}$$

where the Fourier transforms of time functions are denoted by corresponding uppercase letters.

Consider first a single-channel situation, say $G_2 = 0$. In that case, except at zeros of G_1S, Eq. (39) yields $\varepsilon_1 = 0$. Thus, complete alignment ($\hat{h}_1 = h_1$) is achieved by ensuring that G_1S does not vanish at any frequency.

In the stereophonic situation, on the other hand, even if S has no zeroes in the frequency range of interest, the best that can be achieved is

$$\varepsilon_1 G_1 + \varepsilon_2 G_2 = 0 \tag{40}$$

This equation *does not* imply $\varepsilon_1 = \varepsilon_2 = 0$, which is the condition of complete alignment. The problem with stereo echo cancelers is apparent from Eq. (40): Even if the receiving room impulse responses h_1 and h_2 are fixed, any change in G_1 or G_2 requires adjustment of ε_1 and ε_2, except in the unlikely condition where $\varepsilon_1 = \varepsilon_2 = 0$. Thus, not only must the adaptation algorithm track variations in the receiving room, it must also track variations in the transmission room. The latter variations are particularly difficult to track; for if one talker stops talking and another starts talking at a different location, the impulse responses g_1 and g_2 change abruptly and by very large amounts. The difficult challenge, then, is to devise an algorithm that (as in the case of a single-channel canceler) converges independently of variations in the transmission room.

We note that the fundamental problem is not resolved even if we know exactly the impulse responses g_1 and g_2, as in desktop conferencing, where x_1 and x_2 are synthesized (e.g., synthesized stereo) by appropriate choice of g_1 and g_2. This is because there is still no way to identify h_1 and h_2 uniquely.

If two or more independent and spatially separated sources are active in the transmission room, then the nonuniqueness problem essentially disappears because Eq. (40) cannot be simultaneously satisfied for two linearly independent choices of the vector (G_1, G_2) unless $\varepsilon_1 = \varepsilon_2 = 0$. Similarly, if the transmission room frequency responses G_1 and G_2 vary more rapidly over frequency than ε_1 and ε_2, this tends to force the misalignment to zero because Eq. (40) cannot otherwise be simultaneously satisfied over a frequency range for which ε_1 and ε_2 are constrained to be nearly identical while G_1 and G_2 change appreciably. This situation actually occurs in room-to-room conferencing because the impulse responses g_1 and g_2 are generally longer than L and the number of taps used to model h_1 and h_2 and their frequency responses vary rapidly. Thus, the "tails" of the transmission room impulse responses theoretically resolve the nonuniqueness problem; however, solutions so obtained are very poorly conditioned and useless in practice due to the tails of the receiving room impulse responses (69).

Search for Solutions

There have been several partially successful attempts to solve the nonuniqueness problem in stereo acoustic echo cancellation. These include the use of a single adaptive filter and various linear signal decorrelation techniques (68,69). Single adaptive filters, which attempt to estimate the echo using either x_1 or x_2 alone, are unsuitable for practical stereo echo cancellation because such a filter still depends strongly on the responses G_1 and G_2 of the transmission room and room responses do not, in general, have stable inverses. Linear signal decorrelation techniques that have not proved useful include addition of independent random noise to each channel, which is ineffective even if noise shaping is used to exploit masking; use of interchannel decorrelation filters, which only resolve ambiguity at frequencies where G_1 or G_2 (but not both) is zero; frequency shifting, which causes the apparent direction of the sound to oscillate, thereby totally destroying the stereophonic effect; and interleaving comb filters, which have acceptable psychoacoustic degradation only above about 1 kHz.

One solution that has proven effective for speech is the use of nonlinear distortion in each channel, which has the effect of reducing the coherence between the signals x_1 and x_2 (69). This distortion is purposely created by adding to the signal a fraction of its nonlinearly distorted version. Thus, modified signals are formed as

$$x_i'(t) = x_i(t) + \alpha f[x_i(t)], \quad i = 1, 2 \tag{41}$$

where $f(\cdot)$ is a nonlinear function and α is a constant. A choice of f that has proved effective and simple to implement is the half-wave rectifier

$$f(x) = \begin{cases} x, & x \geq 0 \\ 0, & x < 0 \end{cases} \tag{42}$$

The modification Eq. (41) using the half-wave nonlinearity can also be interpreted as the addition of a full-wave nonlinearity, after making suitable scaling changes. The distortion introduced by Eq. (42), even for values of α as large as 0.3, is hardly noticeable for speech. This is at first surprising because, usually, such high distortion is objectionable in high-fidelity audio systems. One possible explanation for why speech is not greatly degraded is that the distortion for vowel-like sounds is comprised of harmonics that tend to be masked by corresponding harmonics of the original signal. Masking is a well-known psychoacoustic phenomenon by which one sound covers up another, and is also used to advantage in perceptual audio coding.

Even with the use of nonlinear signal transformations, the stereo acoustic echo cancellation problem remains difficult. First, two adaptive filters must be used on each end of the communication link. Moreover, because the coherence is reduced to only slightly below unity by the nonlinear distortion, the signal components driving convergence to the true solution are relatively small and therefore convergence of the misalignment is very slow with the LMS or NLMS algorithm. For this reason, RLS or other rapidly converging algorithms (as discussed previously) must be employed for this application. For acoustic echo cancellation involving hundreds or even thousands of taps, such algorithms can consume a great deal of computational power.

One compromise solution to the computational problem has been suggested (70), whereby stereo is employed only below a crossover frequency of about 1 kHz (where stereo is most effective) and mono above that frequency. In this way, computationally intensive fast-converging algorithms are reduced to operating at a low sampling rate, while the more computationally efficient NLMS algorithm is conventionally employed at higher frequencies. A slight loss of stereo localization is suffered. However, the processing load and hence cost are greatly reduced.

Other techniques have also been suggested for solving the stereo acoustic echo cancellation problem (71–73), which remains an area of active research. Thus, at present, the technology is still evolving and it will be some time before it settles out.

CONCLUDING REMARKS

When adaptive echo cancelers were first proposed about 30 years ago, many people expressed doubts about their economic feasibility and about the feasibility of cancelers with more than 50 or 100 adapted parameters. Advances in digital technology have proven these doubts to be unfounded. Since the appearance of the first VLSI implementation of cancelers in 1980, line echo cancelers have become ubiquitous on the telephone network. Several millions of these devices have now been deployed. Cancelers based on similar principles have also found widespread use in data communication (although we have not dealt with that application in this article). The most modern application of voice echo cancelers is to the cancellation of acoustic echoes (e.g., for hands-free conference telephony). In this article, we have described several approaches to this problem. Hardware implementations based on these proposals have been in use since the mid-1980s. Given the pace of development of digital technology, the next decade may well see widespread use of acoustic echo cancelers for offices and larger conference rooms, for both monophonic and stereophonic sound.

ACKNOWLEDGMENTS

The authors would like to thank Eric Diethorn for meticulously reading the text and providing many useful comments. Portions of this article appeared in Ref. 54 and are included here with permission of the publisher.

BIBLIOGRAPHY

1. M. M. Sondhi, An adaptive echo canceller, *Bell Syst. Tech. J.,* **46** (5): 497–511, 1967.
2. A. W. Thies and R. B. Zmood, New ways of echo suppression, *Austral. Telecomm. Rev.,* **1**: 14–19, 1967.
3. S. J. Campanella, H. G. Suyderhoud, and M. Onufry, Analysis of an adaptive impulse response echo canceller, *Comsat Tech. Rev.,* **2**: 1–36, 1972.
4. N. Demytko and L. K. Mackechnie, A high speed digital adaptive echo canceller, *Austral. Telecomm. Rev.,* **7**: 20–28, 1973.
5. Y. Kato et al., A digital adaptive echo canceller, *NEC Res. Dev.,* **31**: 32–41, 1973.
6. O. A. Horna, Echo canceller with adaptive transversal filter utilizing pseudo-logarithmic coding, *Comsat Tech. Rev.,* **7**: 393–428, 1977.
7. D. L. Duttweiler and Y. S. Chen, A single chip VLSI echo canceller, *Bell Syst. Tech. J.,* **59** (2): 149–160, 1980.
8. S. M. Kuo and D. R. Morgan, *Active Noise Control Systems: Algorithms and DSP Implementations,* New York: Wiley, 1996.
9. M. M. Sondhi and D. A. Berkley, Silencing echoes on the telephone network, *Proc. IEEE,* **68**: 948–963, 1980.
10. D. G. Messerschmitt, Echo cancellation in speech and data transmission, *IEEE J. Selected Areas Commun.,* **SAC-2**: 283–297, 1984.
11. O. M. M. Mitchell and D. A. Berkley, A full-duplex echo suppressor using center clipping, *Bell Syst. Tech. J.,* **50**: 1619–1630, 1971.
12. E. J. Thomas, Some considerations on the application of Volterra representation of nonlinear networks to adaptive echo cancellers, *Bell Syst. Tech. J.,* **50** (8): 2797–2805, 1971.
13. B. Widrow and M. E. Hoff, Jr., Adaptive switching circuits, in *IRE Wescon Conc. Rec.,* part 4, 1960, pp. 96–104.
14. S. P. Applebaum, *Adaptive arrays,* Syracuse Univ. Res. Corp., Rep. SPL-709, June 1964; reprinted in *IEEE Trans. Antennas Propagat.,* **AP-24**: 573–598, 1976.
15. B. Widrow and S. D. Stearns, *Adaptive Signal Processing,* Englewood Cliffs, NJ: Prentice-Hall, 1985.
16. S. Haykin, *Adaptive Filter Theory,* Englewood Cliffs, NJ: Prentice-Hall, 1991.

17. M. M. Sondhi and A. J. Presti, A self-adaptive echo canceller, *Bell Syst. Tech. J.,* **45** (14): 1851–1854, 1966.
18. T. Gänsler, A double-talk resistant subband echo canceller, *Signal Process.,* **65** (1): 89–101, 1998.
19. M. M. Sondhi and D. Mitra, New results on the performance of a well known class of adaptive filters, *Proc. IEEE,* **64**: 1583–1597, 1976.
20. J. E. Mazo, Analysis of decision-directed equalizer convergence, *Bell Syst. Tech. J.,* **59**: 1857–1876, 1979.
21. A. Feuer and E. Weinstein, Convergence analysis of LMS filters with uncorrelated Gaussian data, *IEEE Trans. Acoust. Speech Signal Process.,* **ASSP-33**: 222–230, 1985.
22. D. T. M. Slock, On the convergence behavior of the LMS and the normalized LMS algorithm, *IEEE Trans. Signal Process.,* **41**: 2811–2825, 1993.
23. M. Rupp, The behavior of LMS and NLMS algorithms in the presence of spherically invariant processes, *IEEE Trans. Signal Process.,* **41**: 1149–1160, 1993.
24. A. Weiss and D. Mitra, Digital adaptive filters: Conditions for convergence, rates of convergence, effects of noise and errors arising from the implementation, *IEEE Trans. Inf. Theory,* **IT-25**: 637–652, 1979.
25. D. L. Duttweiler, A twelve-channel digital echo canceler, *IEEE Trans. Commun.,* **26**: 647–653, 1978.
26. H. Ye and B.-X. Wu, A new double-talk detection algorithm based on the orthogonality theorem, *IEEE Trans. Commun.,* **39**: 1542–1545, 1991.
27. T. Gänsler et al., A double-talk detector based on coherence, *IEEE Trans. Commun.,* **44**: 1421–1427, 1996.
28. J. Benesty, D. R. Morgan, and J. H. Cho, A new class of doubletalk detectors based on cross-correlation, *IEEE Trans. Speech Audio Process.,* submitted.
29. K. Ochiai, T. Araseki, and T. Ogihara, Echo canceller with two echo path models, *IEEE Trans. Commun.,* **COM-25**: 589–595, 1977.
30. S. Marple, Efficient least-squares FIR system identification, *IEEE Trans. Acoust. Speech Signal Process.,* **ASSP-29**: 62–73, 1981.
31. E. Eweda, Comparison of RLS, LMS, and sign algorithms for tracking randomly time-varying channels, *IEEE Trans. Signal Process.,* **42**: 2937–2944, 1994.
32. L. Ljung, M. Morf, and D. D. Falconer, Fast calculation of gain matrices for recursive estimation schemes, *Int. J. Control,* **27**: 1–19, 1978.
33. D. D. Falconer and L. Ljung, Application of fast Kalman estimation to adaptive equalization, *IEEE Trans. Commun.,* **COM-26**: 1439–1446, 1978.
34. J. M. Cioffi and T. Kailath, Fast, recursive-least-squares transversal filters for adaptive filtering, *IEEE Trans. Acoust. Speech Signal Process.,* **ASSP-32**: 304–337, 1984.
35. D. T. M. Slock and T. Kailath, Numerically stable fast transversal filters for recursive least squares adaptive filtering, *IEEE Trans. Signal Process.,* **39**: 92–114, 1991.
36. K. Ozeki and T. Umeda, An adaptive filtering algorithm using an orthogonal projection to an affine subspace and its properties, *Electron. Commun. Japan,* **67-A**: 19–27, 1984.
37. D. R. Morgan and S. G. Kratzer, On a class of computationally-efficient, rapidly-converging, generalized NLMS algorithms, *IEEE Signal Process. Lett.,* **3**: 245–247, 1996.
38. S. L. Gay and S. Tavathia, The fast affine projection algorithm, *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.,* 1995, pp. 3023–3026.
39. M. Tanaka et al., Fast projection algorithm and its step size control, *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.,* 1995, pp. 945–948.
40. E. Hänsler, The hands-free telephone problem—An annotated bibliography, *Signal Process.,* **27**: 259–271, 1992.
41. Y. Haneda, S. Makino, and Y. Kaneda, Common acoustical pole and zero modeling of room transfer functions, *IEEE Trans. Speech Audio Process.,* **2**: 320–328, 1997.
42. L. M. van de Kerkhof and W. J. W. Kitzen, Tracking of a time-varying acoustic impulse response by an adaptive filter, *IEEE Trans. Signal Process.,* **40**: 1285–1294, 1992.
43. H. Yasukawa, I. Furukawa, and Y. Ishiyama, Acoustic echo control for high quality audio teleconferencing, *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.,* 1989, pp. 2041–2044.
44. R. W. Harris, D. M. Chabries, and F. A. Bishop, A variable step (VS) adaptive filter algorithm, *IEEE Trans. Acoust. Speech Signal Process.,* **ASSP-34**: 309–316, 1986.
45. S. Makino, Y. Kaneda, and N. Koizumi, Exponentially weighted stepsize NLMS adaptive filter based on the statistics of a room impulse response, *IEEE Trans. Speech Audio Process.,* **1**: 101–108, 1993.
46. R. Martin and S. Gustafsson, The echo shaping approach to acoustic echo control, *Speech Commun.,* **20**: 181–190, 1996.
47. V. Turbin, A. Gilloire, and P. Scalart, Comparison of three post-filtering algorithms for residual acoustic echo reduction, *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.,* 1997, pp. 307–310.
48. I. Furukawa, A design of canceller of broad band acoustic echo, *Int. Teleconf. Symp.,* 1984, pp. 1/8–8/8.
49. W. Kellermann, Kompensation akustischer Echos in Frequenzteil-bändern, *Aachener Kolloquium 1984,* Aachen, FRG (in German), 1984, pp. 322–325.
50. W. Kellermann, Analysis and design of multirate systems for cancellation of acoustical echoes, *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.,* 1988, pp. 2570–2573.
51. M. Vetterli, A theory of multirate filter banks, *IEEE Trans. Acoust. Speech Signal Process.,* **ASSP-35**: 356–372, 1987.
52. P. P. Vaidyanathan, *Multirate Systems and Filter Banks,* Englewood Cliffs, NJ: Prentice-Hall, 1992.
53. W. Kellermann, *Zur Nachbildung physikalischer Systeme durch parallelisierte digitale Ersatzsysteme im Hinblick auf die Kompensation akustischer Echos,* PhD thesis, Technische Hochschule Darmstadt, Darmstadt, FRG (in German), 1989.
54. M. M. Sondhi and W. Kellermann, Adaptive echo cancellation for speech signals, in S. Furui and M. M. Sondhi (eds.), *Advances in Speech Signal Processing,* New York: Marcel Dekker, 1992, chap. 11.
55. Y. Itoh et al., An acoustic echo canceller for teleconference, *Proc. Int. Conf. Commun.,* 1987, pp. 46.6-1–46.6-5.
56. H. Yasukawa, S. Shimada, and I. Furukawa, Acoustic echo canceller with high speech quality, *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.,* 1987, pp. 49.8-1–49.8-4.
57. A. Gilloire, Experiments with sub-band acoustic echo cancellers for teleconferencing, *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.,* 1987, pp. 49.2-1–49.2-4.
58. J. Chen et al., A new structure for sub-band acoustic echo canceler, *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.,* 1988, pp. 2574–2577.
59. A. Gilloire and M. Vetterli, Adaptive filtering in sub-bands, *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.,* 1988, pp. 1572–1575.
60. W. Kellermann, Some aspects of the frequency-subband approach to the cancellation of acoustical echoes, *IEEE Conf. Rec. ASSP Workshop Appl. Digital Signal Process. Audio Acoust.,* 1989.

61. B. Hätty, Block recursive least squares adaptive filters using multirate systems for cancellation of acoustical echoes, *IEEE Conf. Rec. ASSP Workshop Appl. Digital Signal Process. Audio Acoust.,* 1989.
62. A. Gilloire and M. Vetterli, Adaptive filtering in subbands with critical sampling: Analysis, experiments, and application to acoustic echo cancellation, *IEEE Trans. Signal Process.,* **40**: 1862–1875, 1992.
63. B. Farhang-Boroujeny and Z. Wang, Adaptive filtering in subbands: Design issues and experimental results for acoustic echo cancellation, *Signal Process.,* **61**: 213–223, 1997.
64. H. Yasukawa and S. Shimada, An acoustic echo canceller using subband sampling and decorrelation methods, *IEEE Trans. Signal Process.,* **41**: 926–930, 1993.
65. D. R. Morgan, Slow asymptotic convergence of LMS acoustic echo cancelers, *IEEE Trans. Speech Audio Process.,* **3**: 126–136, 1995.
66. P. L. DeLeon and D. M. Etter, Experimental results with increased bandwidth analysis filters in oversampled, subband acoustic echo cancelers, *IEEE Signal Process. Lett.,* **2**: 1–3, 1995.
67. D. R. Morgan and J. Thi, A delayless subband adaptive filter, *IEEE Trans. Signal Process.,* **43**: 1819–1830, 1995.
68. M. M. Sondhi, D. R. Morgan, and J. L. Hall, Stereophonic acoustic echo cancellation—An overview of the fundamental problem, *IEEE Signal Process. Lett.,* **2**: 148–151, 1995.
69. J. Benesty, D. R. Morgan, and M. M. Sondhi, A better understanding and an improved solution to the specific problems of stereophonic acoustic echo cancellation, *IEEE Trans. Speech Audio Process.,* **6**: 156–165, 1998.
70. J. Benesty, D. R. Morgan, and M. M. Sondhi, A hybrid mono/stereo acoustic echo canceler, *IEEE Trans. Speech Audio Process.,* **6**: 468–475, 1998.
71. J. Benesty et al., Adaptive filtering algorithms for stereophonic acoustic echo cancellation, *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.,* 1995, pp. 3099–3102.
72. J. Benesty, P. Duhamel, and Y. Grenier, A multi-channel affine projection algorithm with applications to multi-channel acoustic echo cancellation, *IEEE Signal Process. Lett.,* **3**: 35–37, 1996.
73. S. Makino et al., Subband stereo echo canceller using the projection algorithm with fast convergence to the true echo path, *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.,* 1997, pp. 299–302.

M. MOHAN SONDHI
DENNIS R. MORGAN
Bell Laboratories, Lucent Technologies

ECONOMICS OF SOFTWARE DEVELOPMENT. See SOFTWARE COST ESTIMATION.