THE BELL SYSTEM TECHNICAL JOURNAL
Vol. 50, No. 3, March, 1971
Printed in U.S.A.

Performance of an Adaptive Echo Canceller Operating in a Noisy, Linear, Time-Invariant Environment

By J. R. ROSENBERGER and E. J. THOMAS

(Manuscript received October 8, 1970)

In this paper we derive equations describing the performance of various adaptive echo canceller configurations operating in a linear, time-invariant environment. We relate the parameters in these equations to measurable environmental factors, discuss their effect on performance, and verify the results empirically.

In general, the performance of an echo canceller cannot be exactly predicted for speech inputs. Therefore, the derived equations assume a stationary constant power random input. However, it is shown that the results obtained in this manner give useful estimates of the performance to be expected with speech inputs. The similarities and differences of the results for a constant power random input and speech input are discussed in detail.

I. INTRODUCTION

The echo problem in the telephone network is caused by the interaction of the following three factors: (*i*) The impedance mismatches that exist at hybrid junctions cause reflections of incident electrical waves. (*ii*) The existence of a bi-directional transmission medium permits the reflected signal to reach the talker as echo. (*iii*) Time delay due to the finite propagation time of a signal makes the echo annoying. Historically the problem has been alleviated by increasing trunk loss, balancing hybrids, applying four-wire circuits where practical, and providing echo suppressors.[1]

Echo suppressors are used when the echo delay exceed about 45 ms. An echo suppressor is a voice-operated device which switches a large loss in the echo path, as shown in Fig. 1a. This loss blocks the echo effectively but also tends to block speech from the near-end customer

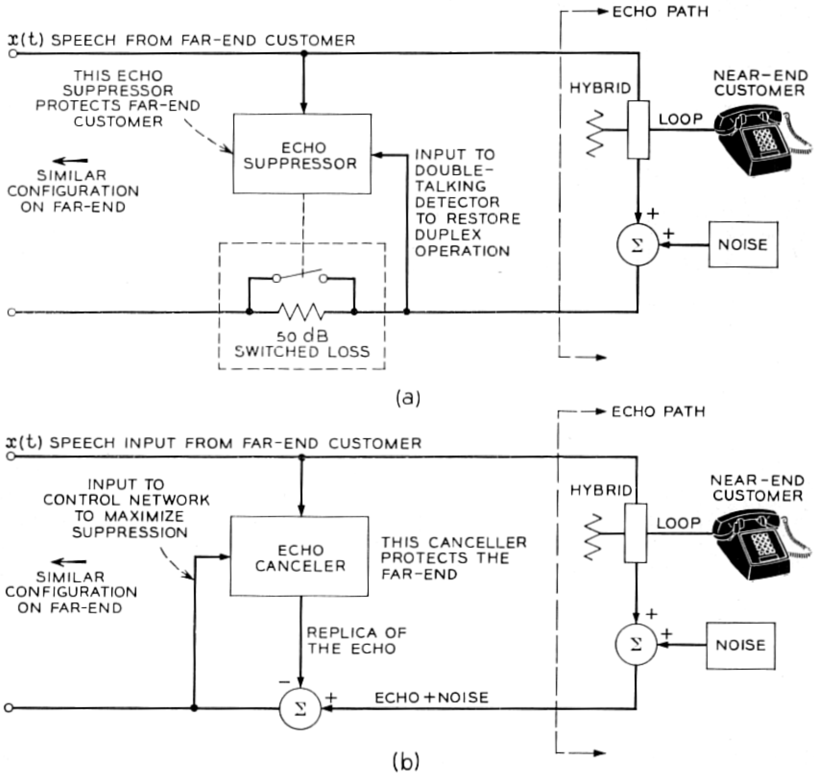

Fig. 1—Block diagrams showing how (a) an echo suppressor is applied in a telephone connection and (b) an echo canceller could be applied in a telephone connection.

when he wishes to interrupt the far-end customer. This situation is known as double-talking. During double talking it is necessary to restore the connection to full duplex. Some speech mutilation (called chopping) and echo occur during these double-talking periods. It has been shown that these degradations become increasingly disturbing as the echo delay increases.[2,3]

The performance of echo suppressors on synchronous satellite circuits is less than satisfying due to the very long delay of such circuits.[4] A new approach to the echo problem, called adaptive echo cancellation,[5–7] has been suggested as a possible alternative. In an echo canceller an approximation of the echo signal is automatically constructed and subtracted from the actual echo with no impairment to duplex operation.

In this paper, our aim is to analyze the operation of an adaptive echo canceller in a linear time-invariant environment. Our hope is to give the reader insight into the parameters which affect performance and their interrelatedness. Our method of presentation is as follows. In Section II we discuss the environmental factors affecting performance and we derive equations describing the operation of various echo canceller configurations in a linear time-invariant environment. Since we were unable to characterize an echo canceller for speech inputs, the derived equations assume a stationary random input. However, in Section III we interpret the equations and show empirically that the results obtained give useful estimates of the performance to be expected with speech inputs.

II. MATHEMATICAL DESCRIPTION OF THE ENVIRONMENT AND THE ADAPTIVE ECHO CANCELLER

The echo paths that we will consider are assumed to be linear, time-invariant channels, not necessarily band-limited and otherwise general. This is not to imply that all real echo paths can be so characterized. In fact, time-varying echo paths have been observed and others are suspected of being significantly nonlinear. These deviations from the conditions assumed above may result in serious performance limitations. References 8 and 9 describe the effect on performance when the environment is either nonlinear or time-variant.

A digital echo canceller, having filters with bandwidth, B, determined by the sampling interval, T, can be used with all echo paths so long as the filter bandwidths are at least as wide as that of the input signal, $x(t)$, i.e., $T \leqq 1/2B$. The same can be said for the bandwidths of the filters of an analog canceller. We will assume that these are also bandlimited to B Hz.

Assuming $x(t)$ and the echo signal, $y(t)$, are bandlimited to B Hz, we can equivalently represent them as sequences of the sampled values at times $t = nT$ where $n = 0, 1, 2, \cdots$. Similarly other signals pertinent to the echo canceller are discrete or continuous and have the independent variable nT or t, respectively. For the sake of brevity we will adopt a common notation, letting ζ denote t or nT. Also the convolution operation will be denoted as

$$\alpha(\zeta) * \beta(\zeta)$$

which corresponds to

$$\int_{-\infty}^{\infty} \alpha(\tau)\beta(\zeta - \tau)\, d\tau$$

in the analog case and to

$$\sum_{k=-\infty}^{\infty} \alpha(kT)\beta(\zeta - kT)$$

in the digital case. Where other differences occur the specific variable will be used.

Whether the echo canceller is digital or analog, it would be inserted into the connection as shown in Fig. 1b. The customer on the left (far end) is being protected from echo by the canceller shown. The customer on the right (near end) is being protected by a similar echo canceller on the far end of the connection.

Three factors that affect the performance of an echo canceller are the types of signals used, echo paths encountered, and echo canceller configuration. We will consider these three points separately. There are three different signals present in the echo canceller environment:

(*i*) The speech of the far-end customer, called the input signal $x(t)$.
(*ii*) The speech of the near-end customer. When the near-end customer and far-end customer speak simultaneously, we have double-talking. This constitutes an interference to the echo canceller.
(*iii*) Interfering circuit noise which is inherent to the echo path.

The echo canceller must perform satisfactorily when these signals are present in all possible combinations. Circuit noise, denoted as $\rho(t)$, is assumed to be a zero mean random process with variance σ_ρ^2 band-limited to B Hz.

A block diagram of the canceller circuit used is shown in Fig. 2. The basic components of this canceller are:

(*i*) A set of M filters having orthonormal impulse responses which are the first M members of a complete basis set.
(*ii*) A control network which automatically weights and sums the outputs of the M filters to generate an approximation of the echo.
(*iii*) Devices to couple the canceller to the telephone plant. The A-D and D-A converters are required for an analog canceller operating in a digital plant or vice versa. The set of M filters have impulse responses $\lambda_1(\zeta)$, $\lambda_2(\zeta)$, $\cdots$, $\lambda_M(\zeta)$. The output of each filter, denoted as $w_m(\zeta)$ and given by

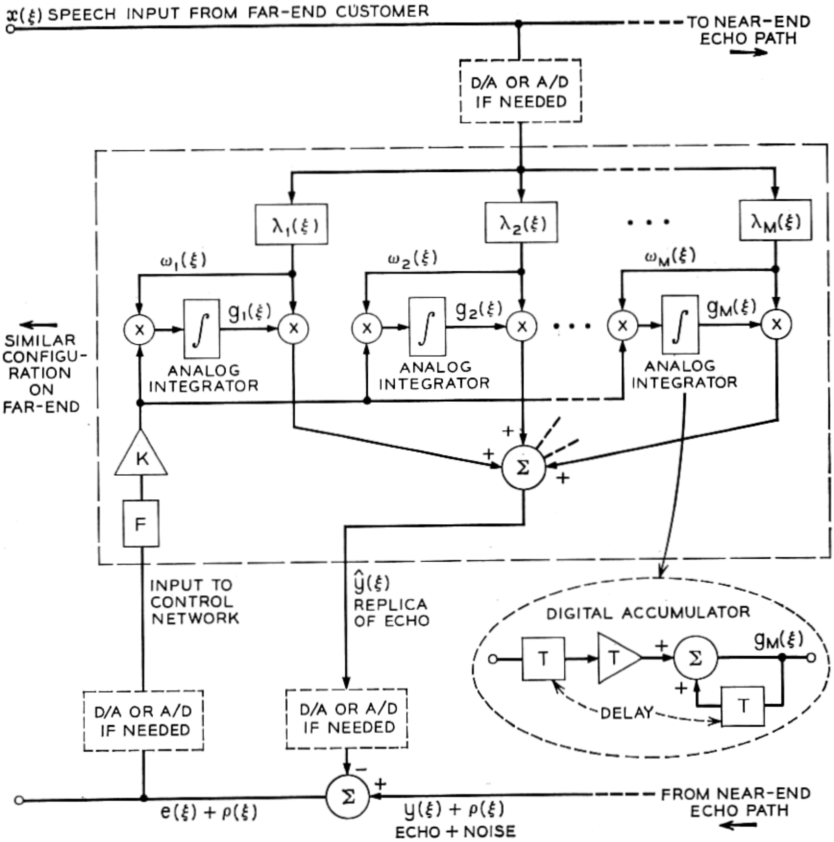

Fig. 2—Block diagram of the structure of an echo canceller.

$$w_m(\zeta) = \lambda_m(\zeta) * x(\zeta), \tag{1}$$

is weighted by the value of the tap gain $g_m(\zeta)$. At $\zeta = 0$ the tap gain is set to some initial value (usually assumed to be zero). The sum of these weighted outputs is the approximation of the echo and is denoted as $\hat{y}(\zeta)$. Thus we have

$$\hat{y}(\zeta) = \sum_{m=1}^{M} g_m(\zeta) w_m(\zeta) \tag{2}$$

which is subtracted from $y(\zeta)$ to give the cancelled echo denoted as $e(\zeta)$. The cancelled echo plus noise $\rho(\zeta)$ is operated on by a function F and then multiplied by a positive factor K. F may be any odd non-decreasing function.

We will consider the two cases:

$$(i)\quad F[\cdot] = [\cdot] \quad \text{and}$$

$$(ii)\quad F[\cdot] = \operatorname{sgn}[\cdot] = \begin{cases} +1 & \text{if } [\cdot] > 0, \\ -1 & \text{if } [\cdot] < 0. \end{cases}$$

The resulting signal is multiplied by $w_m(\zeta)$ and integrated to yield the value of each $g_m(\zeta)$.

The analog network is governed by the set of differential equations

$$\frac{d}{dt} g_m(t) = KF[e(t) + \rho(t)]w_m(t), \qquad m = 1, 2, \cdots, M. \tag{3}$$

The digital network is governed by the set of difference equations

$$\begin{aligned} g_m(nT) = g_m(nT - T) + KTF[e(nT - T) \\ + \rho(nT - T)]w_m(nT - T), \qquad m = 1, 2, \cdots, M. \end{aligned} \tag{4}$$

We can write the error, $e(\zeta)$, as

$$\begin{aligned} e(\zeta) &\equiv y(\zeta) - \hat{y}(\zeta) \\ &= \sum_{m=1}^{M} [c_m - g_m(\zeta)]w_m(\zeta) + q(\zeta), \end{aligned} \tag{5}$$

where

$$q(\zeta) \equiv \sum_{m=M+1}^{\infty} c_m \lambda_m(\zeta) * x(\zeta). \tag{6}$$

The coefficients c_m, $m = 1, 2, \cdots$ are the generalized Fourier coefficients of the echo path transfer function $H(f)$ over the bandwidth $|f| \leqq B$ relative to the complete basis set. They are given by the equation

$$c_m = \int_{-B}^{B} H(f)\overline{\Lambda_m}(f),^{\dagger} \qquad m = 1, 2, \cdots \tag{7}$$

where $\Lambda_m(f)$ is the Fourier transform of $\lambda_m(\zeta)$. The term $q(\zeta)$ is called the uncancellable part of $e(\zeta)$.

Echo suppression achieved ζ seconds after the start of canceller operation is defined as

$$S(\zeta) \equiv -10 \log \frac{E[e^2(\zeta)]^{\ddagger}}{E[y^2(\zeta)]}. \tag{8}$$

† The overbar denotes complex conjugation.
‡ E denotes ensemble average.

Maximum achievable suppression is denoted as S_{max} and equals $\lim_{\zeta\to\infty} [S(\zeta)]$. We define the average settling time t_s to be the time in seconds required for the suppression $S(\zeta)$ to each 98 percent of S_{max} in decibels.

We will now derive equations for maximum achievable suppression and average settling time for the two cases $F[\cdot] = [\cdot]$ and $F[\cdot] = \text{sgn}[\cdot]$. To facilitate the derivations we define the column matrices

$$W(\zeta) \equiv \begin{bmatrix} w_1(\zeta) \\ \vdots \\ w_M(\zeta) \end{bmatrix}$$

and

$$R(\zeta) \equiv C - G(\zeta) \equiv \begin{bmatrix} c_1 \\ \vdots \\ c_M \end{bmatrix} - \begin{bmatrix} g_1(\zeta) \\ \vdots \\ g_M(\zeta) \end{bmatrix}.$$

Using these matrices, we may write $e(\zeta)$ as

$$e(\zeta) = R'(\zeta)\cdot W(\zeta) + q(\zeta)^{\dagger} \tag{9}$$

and $y(\zeta)$ as

$$y(\zeta) = C'\cdot W(\zeta) + q(\zeta). \tag{10}$$

In the derivations which follow we will assume:

(*i*) The input signal, $x(t)$, is a stationary random process having a rectangular power density spectrum

$$P_x(f) = \begin{cases} \sigma_x^2/2B; & |f| \leqq B; \\ 0 \;; & |f| > B; \end{cases}$$

(*ii*) The circuit noise, $\rho(t)$, is a stationary zero mean random process bandlimitted to B Hz and independent of $x(t)$;

(*iii*) For the case $F[\cdot] = \text{sgn}\,[\cdot]$ we will further assume that $x(t)$ and $\rho(t)$ are gaussian with zero mean;

(*iv*) $R(\zeta)$ is independent of both $x(\zeta)$ and $\rho(\zeta)$.

With regard to the last assumption, it is clear that, since $R(\zeta)$ is a function of $x(\zeta)$ and $\rho(\zeta)$, it cannot truly be independent of $x(\zeta)$ and $\rho(\zeta)$. For reasonable values of the feedback factor K, the rate of change

† An apostrophe denotes matrix transpose and a dot denotes scalar multiplication.

of $R(\zeta)$ will be much slower than that of $x(\zeta)$ or $\rho(\zeta)$, so that the assumption is justified for a wide range of operating conditions.

Substituting equations (9) and (10) into equation (8), it can be shown that

$$S(\zeta) = -10 \log \left[\frac{\int_{-B}^{B} |H(f)|^2 \, df - C' \cdot C + E[R'(\zeta) \cdot R(\zeta)]}{\int_{-B}^{B} |H(f)|^2 \, df} \right]. \tag{11}$$

The integral in the denominator of equation (11) is defined as the echo path energy which will be denoted as ψ. We see from equation (11) that $S(\zeta)$ is maximized for a given echo path if $E[R'(\zeta) \cdot R(\zeta)] = 0$. The maximum value is

$$S(\zeta) = -10 \log \left[\frac{\psi - C' \cdot C}{\psi} \right]. \tag{12}$$

Actually this suppression may not be achieved because $E[R'(\zeta) \cdot R(\zeta)]$ may not vanish. However, equation (12) gives a theoretical limit on suppression as defined—this limit being a function of the basis set used and the filter set truncation. We will define an incompleteness (truncation) factor I as

$$I \equiv \frac{\psi - C' \cdot C}{\psi}. \tag{13}$$

Note that I is a nonseparable function of the environment and the echo canceller. That is to say, to calculate the incompleteness factor one must know the echo-path transfer function over the bandwidth of the input signal, the filter set used in the echo canceller and the number of taps employed in the canceller.

To find maximum achievable suppression and average settling time, we must evaluate the term $E[R'(\zeta) \cdot R(\zeta)]$ for the digital and analog case under the two conditions of the function F.

2.1 *Evaluation of $E[R'(t) \cdot R(t)]$ for the Analog Case to Yield S_{max} and t_s*

Using the definition of $R(\zeta)$ and equation (3), we may write

$$\frac{d}{dt} [R'(t) \cdot R(t)] = -2KF[R'(t) \cdot W(t) + q(t) + \rho(t)]R'(t) \cdot W(t). \tag{14}$$

2.1.1 *The Case $F[\cdot] = [\cdot]$*

For this case, we can write equation (14) as

$$\frac{d}{dt}[R'(t)\cdot R(t)] = -2K\{[R'(t)\cdot W(t)]^2 + [R'(t)\cdot W(t)][q(t) + \rho(t)]\}. \quad (15)$$

Solving equation (15) for the expectation of $R'(t)\cdot R(t)$ gives

$$E[R'(t)\cdot R(t)] = R'(0)\cdot R(0) \exp\left(-\frac{K\sigma_x^2 t}{B}\right) \qquad t \geqq 0. \quad (16)$$

Substituting equation (16) into equation (11) yields

$$S(t) = -10 \log\left[I + \frac{R'(0)\cdot R(0)}{\psi} \exp\left(-\frac{K\sigma_x^2}{B}\, t\right)\right]. \quad (17)$$

As $t \to \infty$, we have

$$S_{\max} = -10 \log I. \quad (18)$$

For the given assumptions, the maximum achievable suppression is not a function of circuit noise and is limited only by the incompleteness of the filter set. Of course, strictly speaking, $S_{\max}$ would be less than this limit by an amount depending on the correlation existing between $R(t)$ and $\rho(t)$.

Defining the term

$$s(t_s) \equiv \log^{-1}\left(-\frac{0.98 S_{\max}}{10}\right), \quad (19)$$

the antilog of the suppression at the settling time t_s, substituting this into equation (17) and rearranging we get

$$t_s = \frac{B \log\left[\dfrac{R'(0)\cdot R(0)}{\psi(s(t_s) - I)}\right]}{0.434 K\sigma_x^2}. \quad (20)$$

2.1.2 *The Case* $F[\cdot] = \operatorname{sgn}[\cdot]$ *(hard limiter)*

As above, $S_{\max}$ and t_s may be derived yielding the following two equations:

$$S_{\max} = -10 \log I \quad (21)$$

and

$$t_s = \frac{B}{K\sigma_x^2\sqrt{\frac{2}{\pi}}}\left[2(\gamma - \mu) + 2.3\theta \log\left(\left[\frac{\gamma - \theta}{\gamma + \theta}\right]\left[\frac{\mu + \theta}{\mu - \theta}\right]\right)\right] \quad (22)$$

where

$$\mu = \left(\frac{\sigma_\rho^2 + \psi\sigma_x^2 s(t_s)}{2B}\right)^{\frac{1}{2}}.$$

$$\theta = \left(\sigma_\rho^2 + \frac{\sigma_x^2(\psi - C'\cdot C)}{2B}\right)^{\frac{1}{2}},$$

$$\gamma = \left(\sigma_\rho^2 + \frac{\sigma_x^2[\psi - C'\cdot C + R'(0)\cdot R(0)]}{2B}\right)^{\frac{1}{2}}.$$

If the circuit noise $\rho(t)$ can be neglected and if $I \approx 0$ then equation (22) can be written as

$$t = \frac{1}{\sigma_x K}\,[\pi B R'(0)\cdot R(0)]^{\frac{1}{2}}[1 - \sqrt{s(t)}] \tag{23}$$

where $s(t)$ is the antilog of the suppression at time t.

2.2 *Evaluation of $E[R'(nT)\cdot R(nT)]$ for the Digital Case to Yield $S_{\max}$ and t_s*

Using the definition of $R(\zeta)$ and equation (4), we can write

$$\begin{aligned} R'(nT)\cdot R(nT) - R'(nT - T)\cdot R(nT - T) \\ = -2KTF[R'(nT - T)\cdot W(nT - T) + q(nT - T) \\ + \rho(nT - T)]R'(nT - T)\cdot W(nT - T) \\ + K^2T^2F^2[R'(nT - T)\cdot W(nT - T) + q(nT - T) \\ + \rho(nT - T)]W'(nT - T)\cdot W(nT - T). \end{aligned} \tag{24}$$

2.2.1 *The Case $F[\cdot] = [\cdot]$*

Solving equation (24) for the expectation of $R'(t)\cdot R(t)$, it may be shown that $S(nT)$ is given by

$$\begin{aligned} S(nT) = -10 \log \Bigg[I\left(1 + MK^2T^2\sigma_x^4\left[\frac{\alpha^n - 1}{\alpha - 1}\right]\right) \\ + \frac{R'(0)\cdot R(0)}{\psi}\,\alpha^n + \frac{MK^2T^2\sigma_x^4}{\nu}\left(\frac{\alpha^n - 1}{\alpha - 1}\right)\Bigg] \end{aligned} \tag{25}$$

for $n = 0, 1, 2, \cdots$ and where

$$\begin{aligned} \alpha &= 1 - KT\sigma_x^2[2 - KT(M + 2)\sigma_x^2], \qquad 0 < \alpha < 1; \\ \beta &= MK^2T^2\sigma_x^4(\psi - C'\cdot C) + MK^2T^2\sigma_x^2\sigma_\rho^2. \end{aligned} \tag{26}$$

Note that the limits on α are necessary to yield a convergent system. We define the signal to noise ratio ν at the output of the echo path as

$$\nu = \frac{\psi \sigma_x^2}{\sigma_\rho^2}, \tag{27}$$

and

$$\frac{\mathrm{S}}{\mathrm{N}} = 10 \log \nu \text{ dB}.$$

The maximum suppression is given from the limiting value of $S(nT)$ as $n \to \infty$ as

$$S_{\max} = -10 \log \left[I\left(1 - \frac{MK^2T^2\sigma_x^4}{\alpha - 1}\right) - \frac{MK^2T^2\sigma_x^4}{\nu(\alpha - 1)} \right], \tag{28}$$

and t_s is given by

$$t_s = \frac{T \log \left[\dfrac{s(t_s) - s_{\max}}{\dfrac{R(0)\cdot R(0)}{\psi} + \dfrac{IM(KT\sigma_x^2)^2}{\alpha - 1} + \dfrac{M(KT\sigma_x^2)^2}{\nu(\alpha - 1)}} \right]}{\log [\alpha]}. \tag{29}$$

2.2.2 *The Case* $F[\cdot] = \mathrm{sgn}\,[\cdot]$

With the hard limiter we can write equation (24) as

$$\begin{aligned} R'(nT)\cdot R(nT) &- R'(nT - T)\cdot R(nT - T) \\ &= -2KT \,\mathrm{sgn}\, [R'(nT - T)\cdot W(nT - T) + q(nT - T) \\ &\quad + \rho(nT - T)] W'(nT - T)\cdot W(nT - T) \\ &\quad + K^2T^2W'(nT - T)\cdot W(nT - T). \end{aligned} \tag{30}$$

Using the same assumptions and analysis technique as used for the analog case, we can derive the average value of equation (30) obtaining

$$\begin{aligned} E[R'(nT)\cdot R(nT)] &- E[R'(nT - T)\cdot R(nT - T)] \\ &= -2KT\sigma_x \sqrt{\frac{2}{\pi\psi}} \frac{E[R'(nT - T)\cdot R(nT - T)]}{\sqrt{\dfrac{E[R'(nT - T)\cdot R(nT - T)]}{\psi} + I + \dfrac{1}{\nu}}} \\ &\quad + MK^2T^2\sigma_x^2. \end{aligned} \tag{31}$$

Rather than attempt a solution of this nonlinear difference equation, we will find only the limiting value of $E[R'(nT)\cdot R(nT)]$. For

$E[R'(nT) \cdot R(nT)]$ to converge, we require that the right hand side of equation (31) be nonpositive. Using this inequality and solving for $E[R'(nT) \cdot R(nT)]$ gives

$$S_{\max} \leqq -10 \log \left\{ I + \frac{\pi (MKT\sigma_x)^2}{16\psi} + \left[\frac{\pi^2 (MKT\sigma_x)^4}{256\psi^2} + \frac{2\pi (MKT\sigma_x)^2}{16\psi} \left(I + \frac{1}{\nu} \right) \right]^{\frac{1}{2}} \right\}. \tag{32}$$

An explicit expression for the average settling time t_s for this case is not available. As an alternative, the analog equation for settling time t_s, equation (22), with K replaced by KT may be used to predict settling time for this digital case. When using equation (22), we should use equation (32) to calculate $s(t_s)$.

III. EMPIRICAL RESULTS AND INTERPRETATIONS OF THE THEORETICAL RESULTS

In this section we will compare the theoretical results with empirical findings. We will then discuss the performance predicted by the equations as several of the parameters are varied. Finally we will demonstrate that although the equations were derived for a noise input, they yield useful information about the operation of the canceller for speech inputs provided that the echo path is linear and time-invariant. The empirical results tabulated below pertain only to digital implementations, since an analog system was not available for testing.

The echo canceller shown in Fig. 2 was simulated on a digital computer. Also an echo path, chosen to have characteristics which are similar to those of real echo paths which have been observed, was simulated on the computer. Experiments have also been performed incorporating various analog echo paths with the results in general agreement with predicted performance. For the sake of brevity the latter results are omitted.

The measure of echo canceller suppression which we use to monitor canceller performance is defined by the equation

$$S(\zeta) = -10 \log \frac{\int_0^B \left| H(f) - \sum_{m=1}^{M} g_m(\zeta) \Lambda_m(f) \right|^2 df}{\int_0^B |H(f)|^2 \, df}. \tag{33}$$

Equation (33) yields a measure of the goodness of fit across the entire band at time ζ. It is not necessarily equivalent to the subjective echo reduction which a listener would perceive. In fact subjective testing on a very limited basis indicates that perceived loss may be greater than is indicated by this measure. It is important that this subjective factor be considered when comparing the predicted suppression of an echo canceller with what is considered to be necessary for adequate echo reduction. It is clear that equation (33) would give values of suppression identical to the values given by the previously derived equations if the input to the canceller is noise with a rectangular power density spectrum.

3.1 *Random Noise Input Results*

3.1.1 $F[\cdot] = [\cdot]$

In Fig. 3 we plot suppression for a typical simulation. The choice of parameters for the simulation is listed on the figure. The crosses indicate the results of the computer simulation [equation (33)] which is compared against the values of settling time, t_s, and maximum suppression, S_{max}, as predicted by equations (29) and (28) respectively. We also compare the results of the simulation against the suppression as a function of time predicted by equation (25). We see from the figure that a high value of suppression is obtainable.

In Fig. 4, we allowed the echo canceller to reach its maximum suppression with no circuit noise present. Then we introduced a high noise level, S/N = −18 dB, for approximately 1.5 seconds and then removed it, simulating doubletalking. It is clear from the figure that the results are in very good agreement with the equations.

From these results and numerous others using different echo paths and basis sets we draw the following conclusions:

> The assumption of the independence of $R(\zeta)$ from $x(\zeta)$ and $p(\zeta)$ is quite reasonable for suppressions of up to 40 dB, S/N ratios as low as −20 dB, and settling times of 0.3 second and greater. Therefore for random noise inputs we conclude that the derived equations are very accurate predictors of the performance of an echo canceller.

We now focus our attention on the nature of the equations, and discuss the effects of various environmental factors upon them.

In Figs. 5 through 7, we plot maximum suppression (28) versus $KT\sigma_x^2$ for 100 taps and the incompleteness factors $I = 0.01, 0.001, 0.0001$. In all three figures S_{max} decreases as the S/N decreases. Note that for small values of $KT\sigma_x^2$, as the incompleteness factor I decreases,

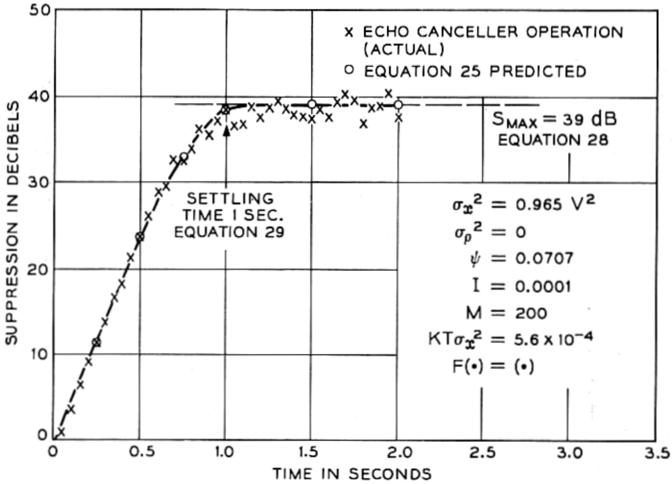

Fig. 3—Comparison of the results of a computer simulation of an echo canceller with the results predicted by the equations. The crosses indicate the simulation results and the circles indicate the predicted results.

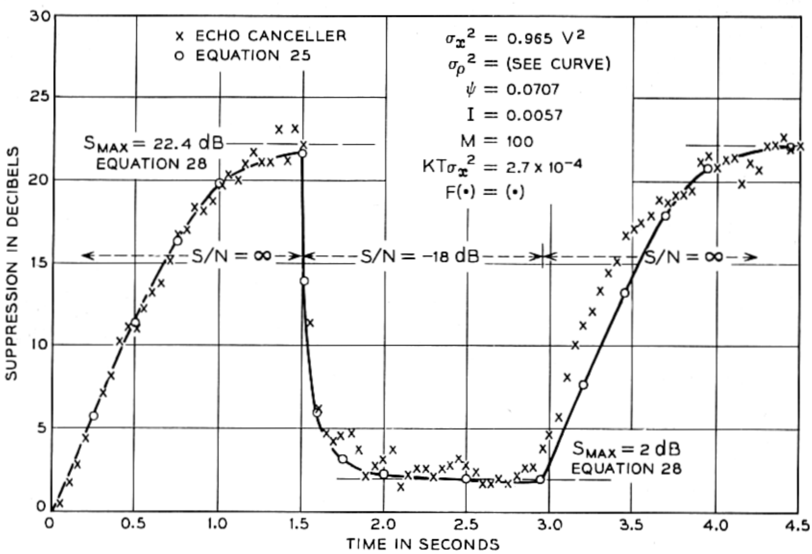

Fig. 4—The same as Fig. 3 except for a different choice of parameters.

the increase in suppression becomes larger for equal increments in S/N. That is, circuit noise (and double-talking) becomes more troublesome as the echo canceller is designed to give higher and higher values of S_{max}.

Note that for fixed S/N, K, and T, an increase in the input signal power reduces the maximum achievable suppression. On the other hand, for fixed σ_ρ^2 and $KT\sigma_x^2 < 10^{-3}$, a change in signal has no appreciable affect upon maximum suppression. However from equation (25), we see that convergence is assured if and only if $0 < \alpha < 1$. This in turn, implies that $KT\sigma_x^2 < 2/(M + 2)$. Therefore, we cannot make $KT\sigma_x^2$ arbitrarily large.

Figures 5 through 7 were calculated for $M = 100$. In order to investigate the sensitivity of S_{max} to M we have plotted S_{max} versus M for $KT\sigma_x^2 = 0.0001$ and $I = 0.001$ and 0.01 in Fig. 8. Observe that S_{max} is a weak function of M. Thus, Figs. 5 through 7 can be used to predict S_{max} for given I and $KT\sigma_x^2$ with little regard for M.

In Figs. 9 through 11, we plot settling time versus $KT\sigma_x^2$ for various choices of I and S/N. Note that a decreasing S/N results in a decreasing settling time. This could lead to the erroneous conclusion that high noise levels help convergence. We find, however, that as the noise level increases, S_{max} decreases. In some cases of very high noise levels, the echo canceller could even provide a net gain. Intuitively it is clear that, starting with zero suppression, it should take less time to settle to the lower level of suppression. For example, consider Fig. 9, with $KT\sigma_x^2 =$

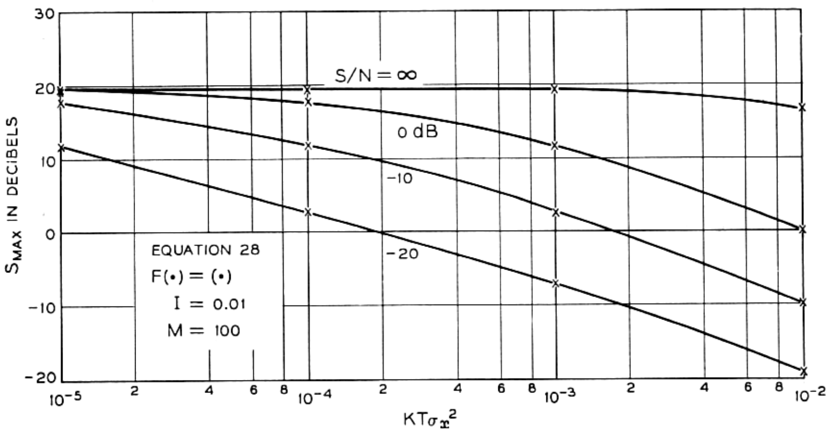

Fig. 5—Theoretically maximum suppression versus $KT\sigma_x^2$ for an incompleteness factor of 0.01 and various echo-to-noise ratios.

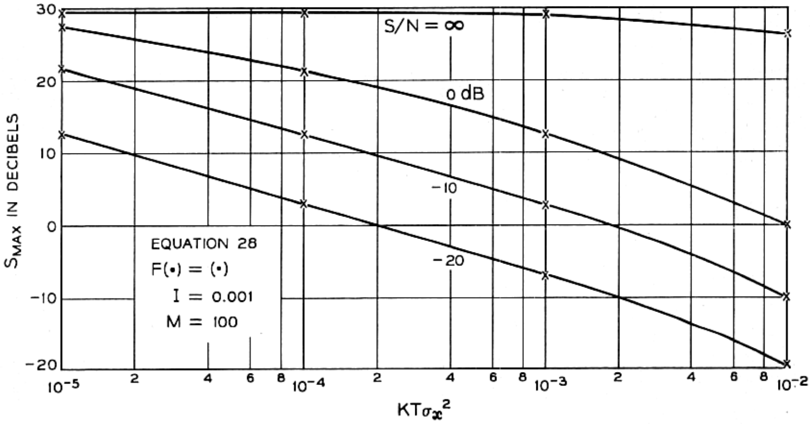

Fig. 6—The same as Fig. 5 but with an incompleteness factor of 0.001.

10^{-3}. It shows that for S/N = ∞ the settling time is approximately 3.6×10^3 iterations versus 1.6×10^3 for S/N = −20 dB. However, Fig. 5 shows that $S_{\max}$ is 19 dB and −7 dB respectively. This situation also illustrates what may happen when a strong interference such as double-talking occurs. The interference will cause divergence to a reduced suppression and may even cause a net gain. We also conclude

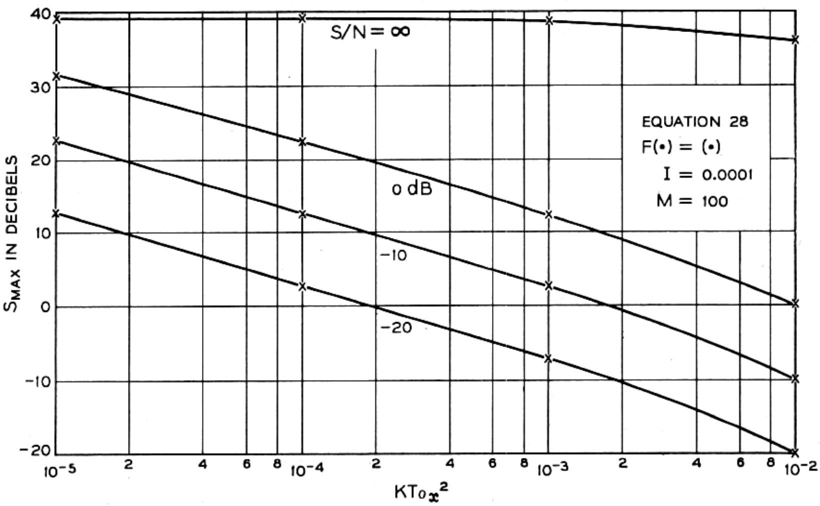

Fig. 7—The same as Figs. 5 and 6 but with an incompleteness factor of 0.0001.

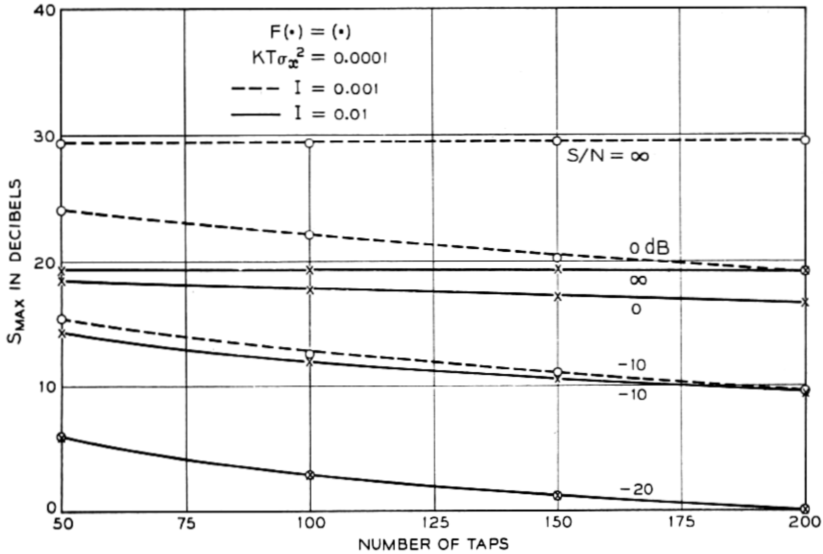

Fig. 8—Theoretically maximum suppression versus number of taps for various echo-to-noise ratios and incompleteness factors.

that for fixed S/N and KT, the larger the input signal power, the faster the settling time will be. However, for the reasons explained previously, the signal power cannot be made arbitrarily large. For constant noise power and fixed KT, it can be seen that settling time is decreased and the rate of increase of suppression is made larger with increased input signal power.

In Fig. 12 we plot settling time as a function of the number of filters M for several values of S/N and I. We see that settling time is relatively insensitive to M and that Figs. 9 through 11 may be used to estimate settling time irrespective of M.

3.1.2 $F[\cdot] = \text{sgn}\,[\cdot]$

We now consider the echo canceller with a hard limiter in the feedback loop. We cannot predict the exact temporal performance of this canceller configuration because we have no solution to the governing difference equation (31). However, we may estimate it by using the solution of the analog differential equation and replacing K with KT. Since this imposes no limit on maximum suppression, we must combine this with the limiting value of S_{max} given by equation (32). This technique yields a reasonable prediction of the operation. For this case

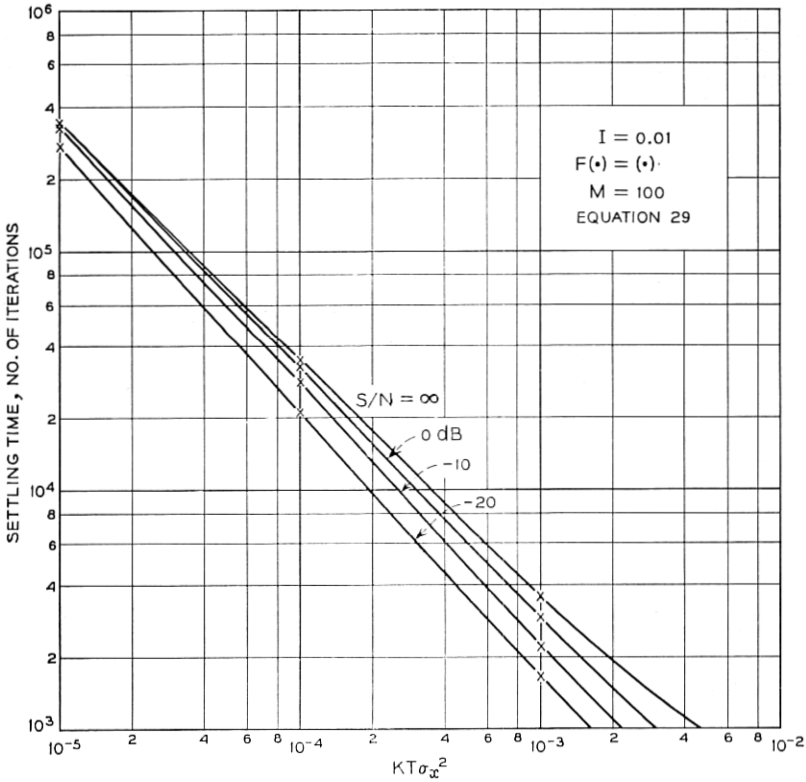

Fig. 9—Settling time versus $KT\sigma_x^2$ for an incompleteness factor of 0.01 and various echo-to-noise ratios.

there are too many variables to present easily a set of curves which describes the operation quantitatively. Therefore we will make some qualitative observations which are generally true. We will use Figs. 13 and 14 as typical examples but we emphasize that these curves are only quantitatively valid for the particular choice of parameters given.

We observe that S_{max} and settling time are inversely proportional to σ_x. For fixed $KT\sigma_x$, S_{max} decreases as S/N decreases. For constant signal to noise ratio, S_{max} decreases with increasing $KT\sigma_x$. However, for constant noise level, S_{max} is relatively insensitive to changes in $KT\sigma_x$ over a wide range. Note too that for $KT\sigma_x$ sufficiently large the canceller may introduce a net gain. For fixed S/N, settling time is a decreasing function of $KT\sigma_x$. For fixed $KT\sigma_x$ a decrease in S/N produces an increase in settling time, while for the case $F[\cdot] = [\cdot]$ the opposite was true. We will now turn our attention to the operation of the echo canceller with speech; we will attempt to interpret the equations in this new light. We will also attempt to show empirically that the results we obtain give useful estimates of performance.

3.2 *Operation With Speech*

The fundamental differences between noise and speech are:

(*i*) The short time (50 ms) average power of speech is erratic from time interval to time interval whereas by comparison it is relatively constant for the random noise.

(*ii*) The spectral density of speech is nonuniform, and depends on

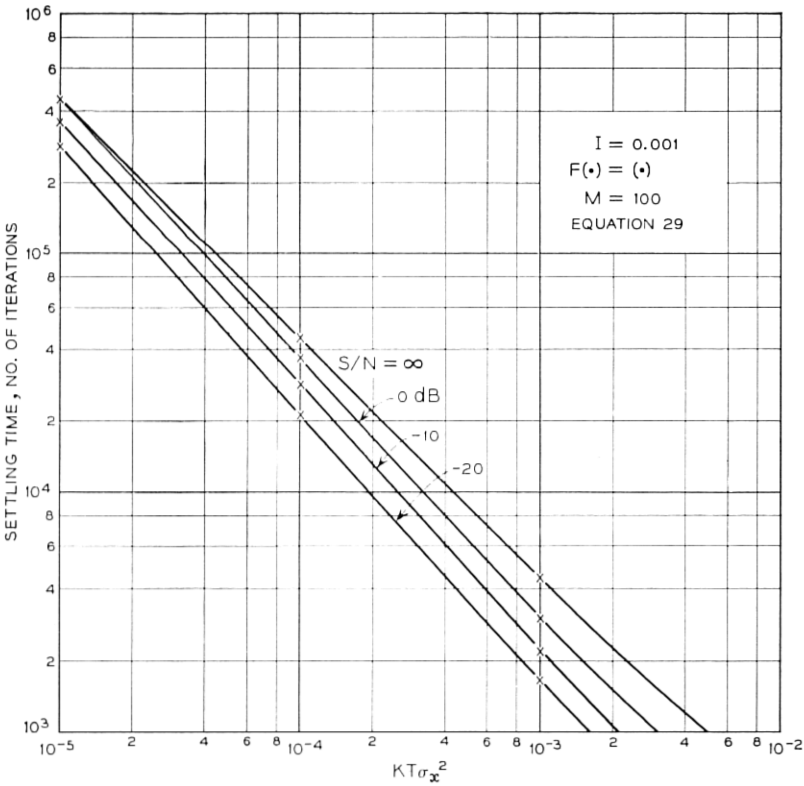

Fig. 10—The same as Fig. 9 but with an incompleteness factor of 0.001.

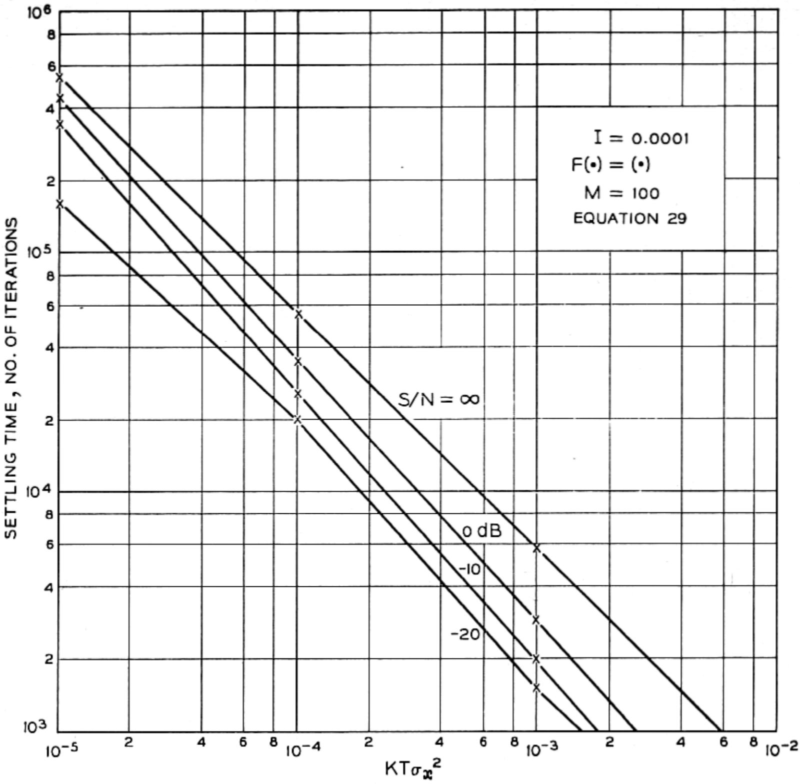

Fig. 11—The same as Figs. 9 and 10 but with an incompleteness factor of 0.0001.

which phoneme is spoken and who speaks it. In fact, the speech power is usually concentrated only a few narrow frequency bands at a time. However, if enough time is allowed to elapse, it is reasonable to assume that the speech power* will eventually scan the entire available bandwidth.

At present, no adequate statistical description of a speech signal accounting for the above properties is available. Using the long-time (several seconds) estimate of average speech power, we have found that the results derived in this paper for random noise may be used as an estimate of the performance which can be expected with speech

* Strictly speaking, this is also true for the random-noise case. However, for noise, the power density spectrum may be considered uniform for a shorter period of elapsed time.

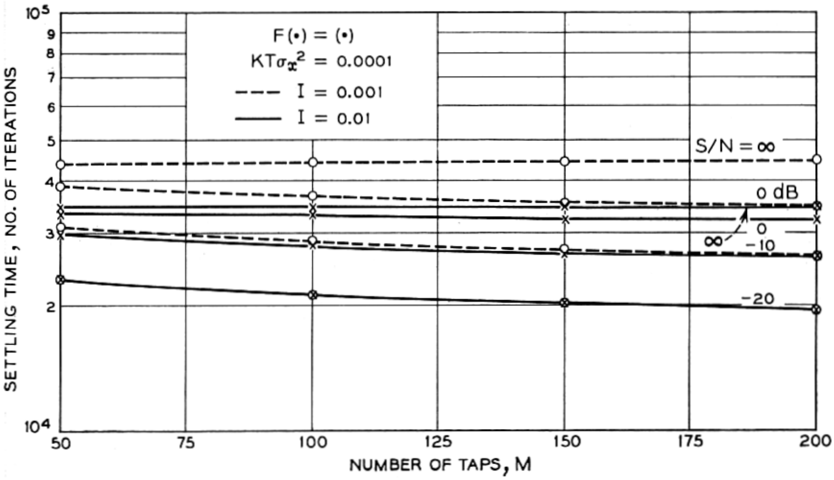

Fig. 12—Settling time versus number of taps for incompleteness factors of 0.01 and 0.001.

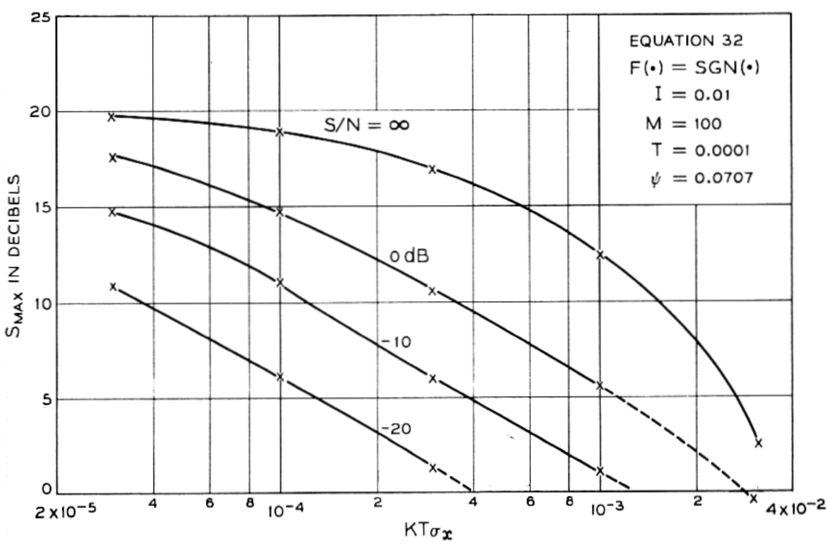

Fig. 13—Theoretically maximum suppression versus $KT\sigma_x$ for an incompleteness factor of 0.01 and various echo-to-noise ratios.

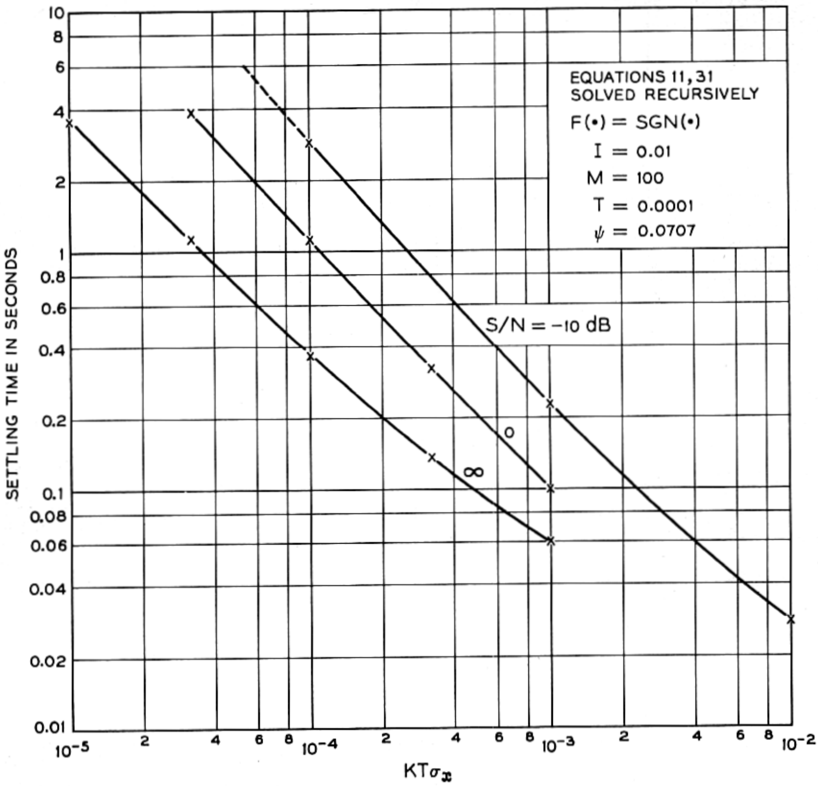

Fig. 14—Settling time versus $KT\sigma_x$ for an incompleteness factor of 0.01 and various echo-to-noise ratios.

inputs. However, the blind use of the equations may give erroneous and optimistically deceiving estimates. We show below how the equations should be used to predict the maximum achievable suppression obtained from the echo canceller with a speech input, and discuss the significance of the settling time estimates.

Because of the variation in speech power level on a short-time (50 ms) basis, we find that the rate of convergence of an echo canceller is erratic. To illustrate this, assume for the moment that we have available speech with a uniform spectral density but with short-time power level variations. For such an input signal, we would find that the operation would be as predicted by the equations with σ_x^2 (short-time power estimate) considered a function of time.

The time variation of the speech spectrum causes individual tap settings of the echo canceller to become correlated. This effect is not taken into account in the equations. This correlation causes the settling-time to be longer than the equations predict. Because of the variations in the spectral density of speech over time, we find that the echo canceller converges on a frequency selective basis. Figure 15 shows the plot of suppression versus time for the echo canceller with a random noise input and $F[\cdot] = \text{sgn}[\cdot]$. The suppression was measured in 20 adjacent frequency bands approximately 200 Hz wide from 0 to 4000 Hz. The suppression in each band was computed by integrating only over that band using equation (33). Two of these bands are shown in Fig. 15. Note that each one converges at approximately the same rate to a limiting value where it then begins to oscillate. Note also that for each band the limiting suppression is reached very close to the predicted settling time of 0.3 second. The other 18 bands behaved similarly. Similar results were obtained with $F[\cdot] = [\cdot]$.

Figure 16 demonstrates what happens when speech is used instead

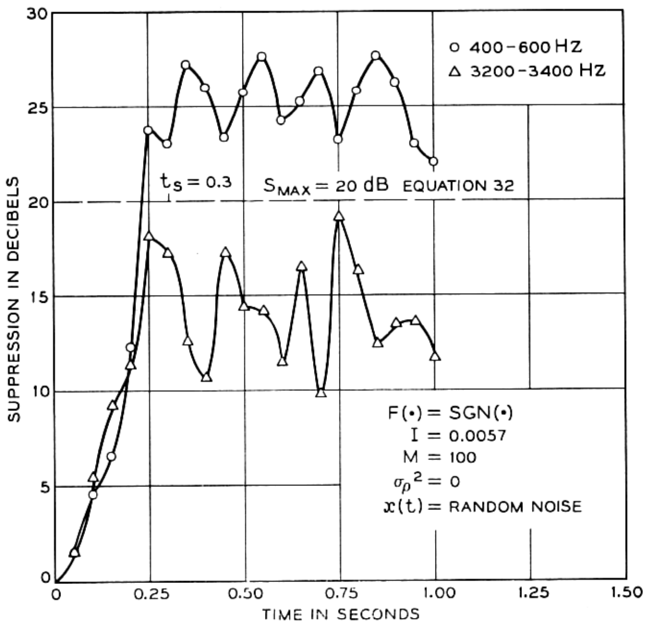

Fig. 15—Suppression as a function of time in the 400- to 500-Hz and 3200- to 3400-Hz frequency bands for a random noise input signal.

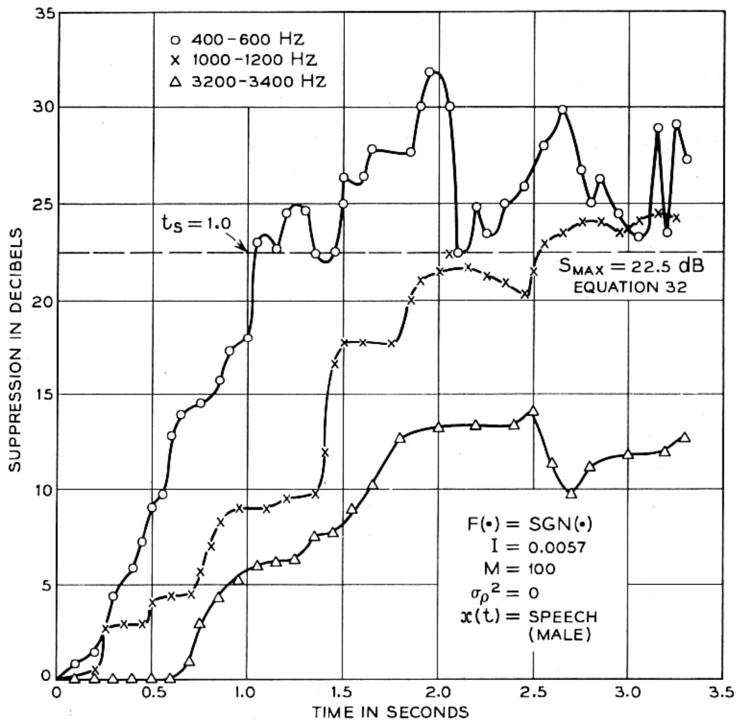

Fig. 16—Suppression as a function of time in the 400- to 600-Hz, 1000- to 1200-Hz and 3200- to 3400-Hz bands with a speech input signal.

of random noise. The designated value of settling time in this experiment was $t_S = 1$ second, using the long-time average (5 seconds) speech power as σ_x^2. Note that the suppression increases at different rates. For example, between $t = 0.75$ second and $t = 1.25$ seconds, the suppression in the 400- to 600-Hz band increases 9 dB while the suppression in the other 2 bands increases 4 dB at most. Between $t = 1.25$ seconds and $t = 1.5$ seconds, however, the rate of convergence becomes most rapid in the 1000–1200 Hz band. This frequency selective convergence is undoubtedly due to the variation in the spectral distribution of speech power. One result of this is a longer overall settling time based on our measure of suppression. The experiment indicates that although the overall settling time may be longer, the echo canceller converges in some frequency bands more rapidly than average. The bands where this speedy convergence takes place are those where the speech power is greatest at a given time. Because of this, we believe that the perceived settling time would be shorter than is indicated by our measure of suppression, equation (33).* Over a long time (several seconds), this selective convergence results in a fit almost equivalent to that of the random noise across the bandwidth of the speech input. We find that the maximum suppression S_{max} achieved with a speech input is very nearly equal to that given by the equations when the long-time average speech power is used for σ_x^2.

Figure 17 shows some typical results which were obtained for a

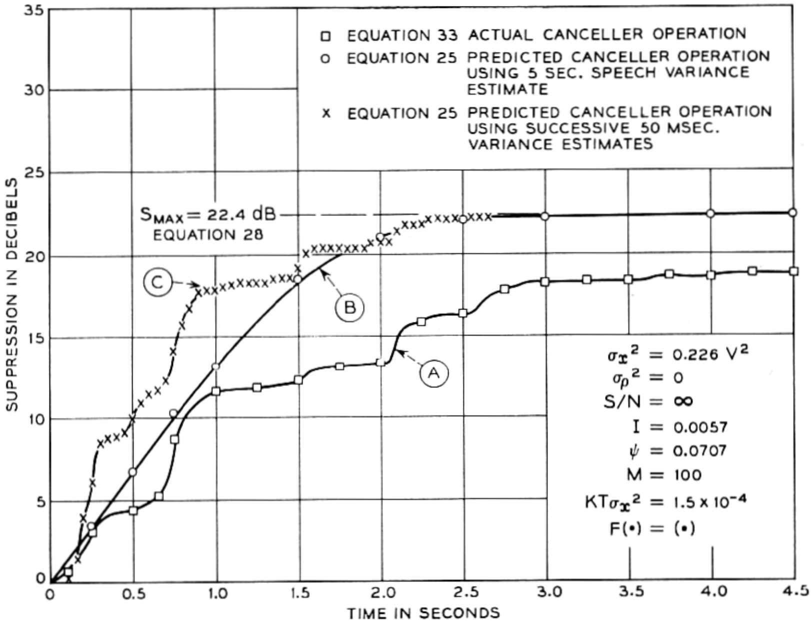

Fig. 17—Comparison of the results of a computer simulation of an echo canceller for a speech input with those predicted by the equations.

simulation with a speech input. Curve (A) shows the results of simulations where $S(nT)$ is calculated every 50 ms using equation (33). Curve (B) shows the performance as predicted using equation (25) and a long-time average (5 seconds) speech power for σ_x^2. The settling time was 2.5 seconds. Curve (C) shows the resulting prediction when

* A need for subjective tests which relate suppression (Equation 44) to perceived suppression exists.

σ_x^2 is replaced by a function of time $\sigma_x^2(\zeta)$, which in this case is the successive short-time (50 ms) average speech power. Note that Curves (A) and (C) are almost identical in shape. However, (C) settles more rapidly to its final value as expected. As explained, the nonuniformity of the speech spectrum causes the longer settling time. Had the simulation been plotted for time longer than 4.5 seconds, we would observe that (A) would converge to its limiting value near $S_{max} = 22.4$ dB.

Another typical case is shown in Fig. 18. A hard limiter was used in the feedback loop of the canceller. The same segment of speech was used here as used in Fig. 17. The value of K was chosen to give a settling time for random noise of 1 second. Note that the echo canceller converged to within 1 dB of S_{max} in 3 to 4 seconds.

The effects of high noise levels are shown in Fig. 19. With the S/N ratio computed to be -10 dB, we see that the echo canceller converged in approximately 1.5 seconds to $S_{max} = 11$ dB, where the suppression then tended to vary around this value. This demonstrates that the canceller converges to S_{max} as predicted in the presence of high levels of noise. Such a strong noise simulates the effect of double-talking. Had the

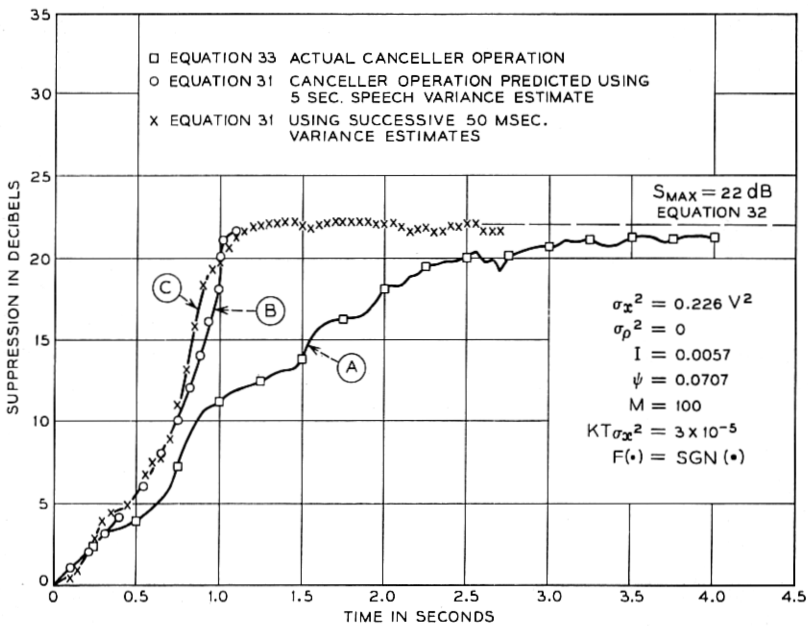

Fig. 18—The same as Fig. 17 with $KT\sigma_x^2 = 3.5 \times 10^{-5}$.

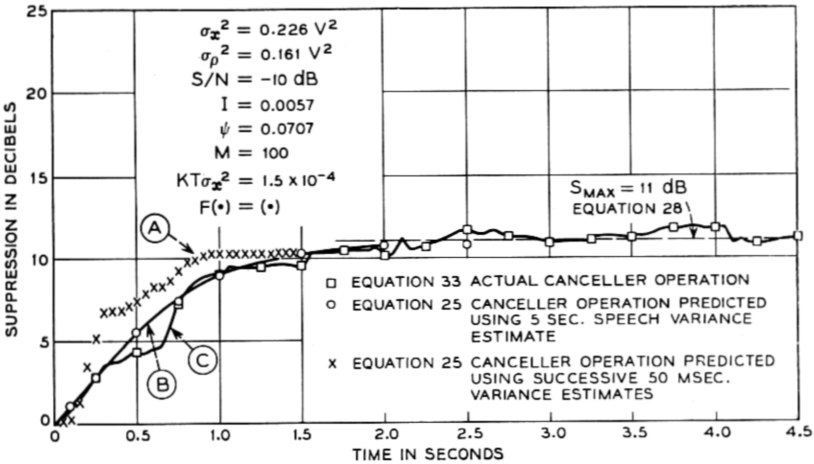

Fig. 19—The same as Fig. 17 with S/N = −10 dB.

noise been introduced after the canceller had converged to some higher S_{max}, the canceller would begin to diverge to the limiting value of $S = 11$ dB. The rate of divergence with speech double-talking is slower than that with random noise. Thus we would find that the effect of double-talking as predicted by the equations would be more severe than it actually is. Also, such an interfering signal causes frequency selective divergence of the echo canceller's transfer function. The divergence is greatest where the interfering signal spectrum is largest. This is not necessarily where the input signal spectrum is largest.

In summary, we see that a long time estimate of speech power can be used in the equations to give a good estimate of the limiting value of suppression S_{max}. Also we see that the canceller performs generally as predicted with a speech input, and in the presence of strong interfering noise. In general, the settling time of the canceller is longer than that predicted by the equations. The settling time may be reduced by increasing K but this must be weighed against the resulting decrease in S_{max}. Also if K is made too large convergence may not take place at all.

IV. SUMMARY

We have described the performance of an adaptive echo canceller operating in a linear, time-invariant, noisy environment. Both digital and analog implementations were considered. In both cases the echo cancellers were assumed to consist of a set of M filters having orthonormal impulse responses which were selected as the first M member of a complete basis set. A weighted sum of the filter outputs approximates the echo signal. The approximation is substracted from the real echo and the difference signal is used to continually improve the approximation so that the cancelled echo power tends toward a minimum. We have used the mean-squared value of the difference between the transfer functions of the echo canceller and the echo path uniformly weighted over the bandwidth of the input signal as a measure of suppression. This measure is not necessarily equivalent to the subjective echo loss (apparent loss perceived by listeners) other than on a relative basis. Sets of equations were derived giving maximum achievable suppression and settling time of the echo canceller. We have shown that despite certain simplifying assumptions made in their derivations, the equations accurately describe the performance for a random noise input.

Families of curves—Figs. 5 through 7, 9 through 11, 13, and 14 show maximum suppression and average settling time for a range of incompleteness factors I, S/N, and a factor related to the input power. The results of simulations are shown to be in close agreement with the predictions.

For a speech input we have found that the equations for maximum suppression can be used to predict performance. The long-time (several seconds) average speech power is used in the equations. The short-time variability of speech power and spectral variations of the speech signal cause the settling time of the echo canceller to be longer than that given by the equations. We have found that during convergence a speech input causes the transfer function of an echo canceller to converge on a frequency selective basis—the fit being best where the power spectrum of the input is greatest. We find that, given enough time, the transfer function of the echo canceller converges to essentially a uniform fit of the echo path transfer function over the bandwidth of the input signal. We have also found that an interfering speech signal (such as exists during double-talking) will cause the echo canceller to diverge on a frequency selective basis. The rate of divergence with speech interference is less than that for the random noise interference.

Before concluding, two final points should be reemphasized. All the previous analysis is only valid when the environment is linear and time-invariant. At present, we suspect that certain systems (compandored systems for example) exhibit non-negligible nonlinearities.

For such systems, the previous analysis may not suffice depending on the type and magnitude of the nonlinearities. In any case it becomes extremely dangerous for compandor type nonlinearities to attempt to relate the performance of an echo canceller to a speech input from the white noise equations given previously.

Also, it should be stressed, that the measure of performances we have defined are objective in nature. These measures are not necessarily equivalent to the subjective echo reduction which a listener will preceive. A need exists to relate the objective and subjective.

REFERENCES

1. Unrue, J. E., "Controlling Echo in the Bell System," Bell Labs Record, *47,* No. 7 (August 1969), pp. 233–238.
2. Emling, J. W., and Mitchell, D., "Effects of Time Delay and Echos on Telephone Conversations," B.S.T.J., *42,* No. 6 (November 1963), pp. 2869–2891.
3. Klemmer, E. T., "Subjective Evaluation of Transmission Delay in Telephone Conversation," B.S.T.J., *46,* No. 6 (July–August 1967), pp. 1141–1147.
4. Helder, G. K., "Customer Evaluation of Telephone Circuits with Delay," B.S.T.J., *45,* No. 7 (September 1966), pp. 1157–1191.
5. Sondhi, M. M., "An Adaptive Echo Canceller," B.S.T.J., *46,* No. 3 (March 1967), pp. 497–511.
6. Sondhi, M. M., and Presti, A. J., "A Self Adaptive Echo Canceller," B.S.T.J., *45,* No. 10 (December 1966), pp. 1851–1854.
7. Becker, F. K., and Rudin, H. R., "Application of Automatic Transversal Filters to the Problem of Echo Suppression," B.S.T.J., *45,* No. 10 (December 1966), pp. 1847–1850.
8. Thomas, E. J., "An Adaptive Echo Canceller in a Non-Ideal Environment," unpublished work.
9. Thomas, E. J., "Some Considerations of the Application of the Volterra Representation of Nonlinear Networks to Adaptive Echo Cancellers and Control Systems," unpublished work.