FEEDBACK IN DATA TRANSMISSION

P. M. EBERT
Bell Telephone Laboratories
Holmdel, New Jersey

Summary  A survey of the possible gains to be realized by the use of various feedback techniques is given. Noiseless information feedback is considered in detail, and a transmission system for this latter case is given and analyzed. This system is shown to achieve a transmission rate very close to the largest rate possible.

Introduction  In the transmission of data numerous feedback systems have been used for many years. Without exception these have been decision feedback systems, implemented by variations of the repeat request strategy. The object of these systems is to achieve low probability of error without excessive equipment complexity. The normal repeat request strategy requires that the incoming data be buffered or stored during repeats of previous data. This results in a random distribution of the amount of data in storage at any given time. This distribution is invariably exponential and therefore implies a probability of buffer overflow which is an exponential function of the amount of storage.

The primary advantage of information feedback over decision feedback is that the probability of error can be made to go to zero much faster than exponentially. It should be noted that this gain is not due only to the use of information feedback. It is not possible to make the probability of error go to zero faster than exponentially for most discrete channels, and Pinsker\(^1\) has shown that the same result holds for continuous channels if there is a peak power constraint at the transmitter. Consequently the phenomenon of faster than exponential decrease in probability of error occurs only in some rare discrete channels and in continuous channels when the peak power constraint is replaced by an average power constraint. It is the latter case which we shall consider in this paper.

Practical Advantages of Information Feedback  There are three practical considerations which make information feedback attractive, some of which are shared by decision feedback. First, since the goal is a small probability of error goes to zero with blocklength is important. The faster than exponential (in fact exponentially exponential)

\(^1\) M. S. Pinsker, “Error Probability for Block Transmission on a Gaussian Memoryless Channel with Feedback.” Problemy Peredachi Informatsii 4, 3-19 (1968).
decrease may allow a significantly shorter blocklength for a given probability of error. This ties in with the second gain; simplification of decoding complexity. Both decision feedback and information feedback result in simple decoding. With decision feedback the receiver just detects errors and requests a repeat, and with information feedback the receiver consists of an analog to digital converter which converts an estimate of a real number into its binary representation. However, a shorter blocklength will naturally produce a decoding simplification. In addition the buffering problem is somewhat simplified by the operation of information feedback. Decision feedback detects errors and institutes a retransmission, whereas information feedback tends to prevent errors from occurring (by increased transmitter power) in the first place.

The third advantage of information feedback is the possible increase in the capacity of the channel. This is more of a theoretical advantage because the gain is not very large and because, as yet, no one has succeeded in transmitting anywhere near the channel capacity. In certain cases it can be shown that the proposed information feedback system can come within .4% of doubling the capacity of the channel, which is a surprising result in that it has been shown that feedback cannot increase the capacity by more than a factor of two for this type of channel.

The major limitation on information feedback, at the present time, is how best to accommodate the noise in the reverse channel. Butman has concluded that the probability of error cannot be driven to zero with linear feedback techniques in the presence of noisy feedback. However, if the goal is simply to transmit as reliably on the forward channel as on the reverse channel (a reasonable goal if the reverse channel has a much higher capacity) the noise in the feedback takes on secondary importance. One can assume that the feedback is error free and thus errors in the feedback simply cause errors in the transmitted data.

A. Information Feedback Technique for Time Continuous Channels. The scheme described and analyzed here is a modification of the Schalkwijk, Kailath scheme for time discrete channels. The ground rules are that the system has the time interval [0,T] to transmit one of exp [RT] signals. The interfering noise is Gaussian with covariance R(τ). The spectrum of the noise does not go to zero at infinite frequency, and the eigenvalues of the kernel R(τ), over the range [0,t], are bound away from zero. We will calculate the probability or error and average transmitter power required for the rate R.

---

2 P. M. Ebert, “The Capacity of the Gaussian Channel With Feedback,” to appear in BSTJ.


The data to be transmitted is represented as a point, \( x \), on the interval \([-1,1]\).

Since there are \( \exp[RT] \) possible signals, the signal points are separated by \( \exp[-RT] \). Now the problem has been reduced to transmitting an accurate estimate of a real number in \( T \) seconds. The transmission philosophy is that the receiver forms an estimate of \( x \), call it \( \hat{x}(t) \), and returns this estimate to the receiver via the noiseless feedback. The transmitter then transmits an error signal, \( x - \hat{x}(t) \). However, since the receiver estimate is getting better as time progresses \( x - \hat{x}(t) \) gets smaller and the transmitter can insert an increasing amplification factor. The increasing amplification factor which produces a constant average transmitter power is exponential, and the transmitted signal is:

\[
\left[ x - \hat{x}(t) \right] \exp[\alpha t] .
\]

The estimator at the receiver forms a minimum variance unbiased estimate of \( x \) given by:

\[
\hat{x}(t) = \frac{1}{A(t)} \sum_{i} \frac{e_i}{\lambda_i} \int_{0}^{t} \left[ r(\tau) e^{-\alpha \tau} + \int_{0}^{\tau} \hat{x}(z) h(\tau-z) e^{-\alpha(\tau-z)} dz \right] d\tau
\]

where \( \lambda_i \) and \( \theta_i(\tau) \) are the eigenfunctions and eigenvalues of \( R(\tau) \) over the interval \([0,t] \). The coefficients \( e_i \) are defined by

\[
e_i = \int_{0}^{t} e^{-\alpha(\tau-\tau)} \theta_i(\tau) \int_{0}^{\tau} h(x) e^{-\alpha x} dx \ d\tau ,
\]

\( h(z) \) is the impulse response of the channel, \( r(\tau) \) is the channel output, and \( A(t) \) is defined by

\[
A(t) = \sum_{i} \frac{e_i^2}{\lambda_i}
\]

The expression for \( \hat{x}(t) \) is quite complicated, but it leads to a simple expression for probability of error and power. Before we proceed, it should be pointed out that whenever \( h(\tau) \) has a delay, there is a period of time for which it is impossible to form an unbiased estimate. This is indicated by \( e_i \) and \( A(t) \) both taking on the value zero. This problem is overcome by defining \( \hat{x}(t) \) to be zero when \( A(t) = 0 \).

At time \( T \) a decision is made by choosing as the data that value of \( x \) closest to \( \hat{x}(T) \). The probability of error is simply the probability that \( | x - \hat{x}(\tau) | \geq \exp[RT] \), which can be calculated exactly because the quantity \( x - \hat{x}(T) \) is Gaussian.
\[ P_e = \text{erfc} \left[ \sqrt{\frac{A(T)}{2}} \exp \left[ T(\alpha-R) \right] \right] \leq (2\pi)^{-1/2} A(T)^{-1/4} \exp \left[ -\frac{A(T)}{2} \exp \left[ 2T(\alpha-R) \right] \right]. \]

This goes to zero doubly exponentially in \( T \) as long as \( R < \alpha \).

The average transmitter power is \( 1/A(t) \) which for large \( t \), becomes:

\[
P = 2\alpha \exp[\alpha D] \left( \int_0^\infty h(\tau) \exp[-\alpha \tau] d\tau \right)^2 \left( \int_0^\infty g^{-1}(\tau) \exp[-\alpha \tau] d\tau \right)^2
\]

where \( g^{-1}(\tau) \) is the realizable inverse of the noise filter, and \( D \) is the round trip excess delay not included in \( h(\tau) \).

**Performance for Several Noise Spectra**

There are a few noise spectra for which the performance functions can be computed in terms of elementary functions. First let \( N(w) = N_0 \delta(w) \) and \( h(\tau) = \delta(\tau) \). Now the power is just \( 2N_0\alpha \). The capacity of this channel (either with or without feedback) is \( P/2N_0 \) and consequently we can transmit at any rate less than the capacity.

Another example is noise with the spectrum:

\[
N(w) = \exp \left[ -\frac{c}{a^2 + w^2} \right].
\]

Tedious calculations show that

\[
g^{-1}(\tau) = \delta(\tau) + \sqrt{c/2a} \exp[-a \tau] \text{ I. } (\sqrt{2c \tau/a})
\]

and that the power is

\[
P = 2\alpha \exp \left[ -\frac{c}{a(a+\alpha)} \right]
\]

on the other hand a relation between rate and power without feedback is given by Shannon’s formula\(^5\)

\[
R = - \frac{1}{4\pi} \int_{-\infty}^{\infty} \ln N(w) \, dw = c/4a.
\]

---

and the power

$$P = \frac{1}{2\pi} \int_{-\infty}^{\infty} \left[ 1 - N(\omega) \right] d\omega = \frac{c}{2a} \exp \left[ \frac{-c}{2a^2} \right] \left[ I_0 \left( \frac{c}{2a^2} \right) + I_1 \left( \frac{c}{2a^2} \right) \right].$$

In order to compare the rates we equate the powers and write $\alpha$ as $kc/4a$. Now $k$ represents the ratio of maximum rate with feedback to the maximum rate without feedback for the same average power. The resulting relation is:

$$k = \left[ I_0(u) + I_1(u) \right] \exp \left[ u \frac{2 - ku}{2 + ku} \right],$$

where $u = c/2a^2$. For every value of $u$ one can calculate a corresponding value of $k$. First we see that as $u$ goes to zero, $k$ goes to one, which is just the white noise case of the first example. For all values of $u$ less than 1900, $k$ is larger than one and $k$ takes on its maximum of 1.996 at $u = 2.27$. This maximum is rather broad and $k$ remains near two for a large range of $u$. For example when $u = 1$, $k = 1.885$ and when $u = 5$, $k = 1.925$. The significance of these calculations lies in the fact that $k$ can never be larger than two for any transmission system operating on a channel with additive Gaussian noise.

**Implementation**  The description and analysis of the system have been in terms of a block coded system because such system are conceptually simpler. However, one would probably build the system in a convolutional (or recurrent) form. If one considers $x$ to be made up of an infinite number of components:

$$x = \sum_{i=1}^{\infty} x_i 2^{-i}, \quad x_i = \pm 1$$

it is clear that those components of $x$, which are small compared to $e^{-\alpha t_n}$, are unimportant. Therefore the transmitter need not know all the components of $x$ at the start of transmission, but only needs to know a linearly increasing fraction of them. This is the essence of a convolutional system. The feedback implementation is also simplified by the convolutional approach. $\hat{x}$ now consists of an infinite set of binary components. At any given time some of the components are known quite well and some are not known at all. Again, as time passes the number of known components increases linearly with time. Thus the feedback channel need only consist of a binary channel communicating the binary components of $\hat{x}$ as they become known, plus additional information indicating corrections to past binary components which have to be changed due to the updated estimate. In this system the achievable error rate would be limited by the error rate in the feedback channel.
Conclusions  We have examined in detail one possible way to use a high capacity or highly reliable feedback channel in order to achieve a low probability of error on the noisy forward channel. We have found that two of the defects of decision feedback (buffering of the input data, and low rate) can be eliminated by the use of information feedback. In addition we have found that the proposed coding scheme has a remarkable ability to achieve large transmission rates.