

# The Role of Feedback in Communication

Thomas M. Cover\*  
Stanford University

## 1 Introduction.

Most of us have a reasonably good idea of the role of feedback in control systems. One can drive from Boston to New York with one's eyes open, but not with them closed. Feedback not only makes driving simpler, it makes it possible. Recursive corrections are easier than open loop corrections.

We wish to ask the same questions about the role of feedback in communication. Some possible roles that feedback might play in communication include the following: correct receiver misunderstanding; predict and correct the noise; cooperate with other senders; determine properties of the channel; reduce communication delay; reduce computational complexity at the transmitter or the receiver and improve communication rate.

Information theory will be the primary tool in the analysis because information theory establishes the boundaries of reliable communication. Shannon proved the first shocking result about feedback, which we shall treat in the next section.

## 2 Feedback for memoryless channels.

Consider the following setup. One has a transmitter  $X$ , a receiver  $Y$ , and a conditional probability mass function  $p(y|x)$ . If one wishes to use this channel  $n$  times, we shall define a  $(2^{nR}, n)$  feedback code as a collection of code words  $X(W, Y)$  of block length  $n$  in which the  $i$ -th transmission  $X_i(W, Y_1, \dots, Y_{i-1})$  depends on the message  $W \in \{1, 2, \dots, 2^{nR}\}$  as well as the previous received  $Y$ 's available through feedback. The

---

\*This work was partially supported by NSF Grant ECS82-11568, DARPA Contract N00039-84-C-0211, and Bell Communications Research.

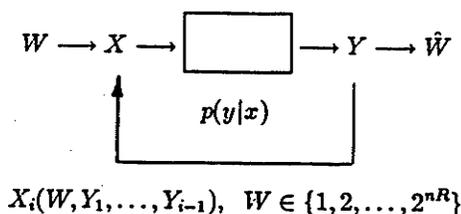


Figure 1: Channel with Feedback.

capacity  $C$  of such a channel is the maximum rate  $R$  at which one can find  $(2^{nR}, n)$  codes such that the probability of decoding error  $P_e = \Pr\{\hat{W}(Y) \neq W\}$  tends to zero asymptotically as  $n \rightarrow \infty$ . The first major surprise in the role of feedback in communication is that feedback does not improve the capacity of memoryless channels.

We first recall that the channel capacity for a discrete memoryless channel without feedback is given by

$$C = \max_{p(x)} I(X; Y) \quad (1)$$

(The codewords for such a channel depend on  $W$  only and not on the previous  $Y$ 's). Thus communication rates up to  $C$  bits per channel use can be achieved. We now have the following result due to Shannon.[8]

**Theorem 1** *Feedback does not improve the capacity of memoryless channels.*

**Proof:** We take a feedback code, as described above, put a uniform probability distribution on the indices  $W \in \{1, 2, \dots, 2^{nR}\}$  and then observe the following chain of inequalities:

$$\begin{aligned} nR &\stackrel{(a)}{=} H(W) \\ &\stackrel{(b)}{\leq} H(W) - H(W|Y^n) + n\epsilon_n \\ &\stackrel{(c)}{=} I(W; Y^n) + n\epsilon_n \\ &\stackrel{(d)}{=} H(Y^n) - H(Y^n|W) + n\epsilon_n \\ &\stackrel{(e)}{=} H(Y^n) - \sum H(Y_i|W, Y^{i-1}) + n\epsilon_n \\ &\stackrel{(f)}{=} H(Y^n) - \sum H(Y_i|W, Y^{i-1}, X_i) + n\epsilon_n \\ &\stackrel{(g)}{=} H(Y^n) - \sum H(Y_i|X_i) + n\epsilon_n \end{aligned}$$

$$\begin{aligned}
& \stackrel{(h)}{\leq} \sum H(Y_i) - \sum H(Y_i|X_i) + n\epsilon_n \\
& \stackrel{(i)}{=} \sum I(X_i, Y_i) + n\epsilon_n \\
& \stackrel{(j)}{\leq} nC + n\epsilon_n
\end{aligned}$$

where  $\epsilon_n \rightarrow 0$ , as  $n \rightarrow \infty$ . Here (a) is the entropy of a uniformly distributed random variable, (b) is Fano's inequality, (c) is by definition of  $I$ , (d) is an expansion of  $I$ , (e) is the chain rule, (f) uses the fact that  $X_i$  is a function of  $W, Y^{i-1}$  via feedback, (g) follows from the fact that we have a discrete memoryless channel, (h) is a well known upper bound, (i) is an identity, and (j) follows from the definition of  $C$  in (1). Thus we have shown that feedback does not increase capacity.

On the other hand, there are numerous results that indicate that feedback may reduce the coding complexity. The simplest example of this might be the binary erasure channel in which one sends  $X \in \{0, 1\}$  and receives  $Y \in \{0, 1, E\}$ , where  $E$  denotes erasure. Suppose  $Y = X$  with probability  $1 - \alpha$ , and  $Y = E$ , with probability  $\alpha$ . The capacity of the channel is  $C = 1 - \alpha$ , but is difficult to achieve. On the other hand, with feedback, there is no problem. Merely retransmit the intended symbol whenever an erasure is received. The expected number of transmissions required to reveal a given symbol  $X$  is  $1/(1 - \alpha)$ . Thus  $(1 - \alpha)n$  transmitted bits require  $n$  transmissions and the resulting capacity is  $C = 1 - \alpha$ . In fact, from this example, we see that not only is encoding and decoding simple with feedback, but that the delay in the received transmission is finite and small.

### 3 Gaussian channels with feedback.

We will now take a look at the special case of Gaussian channels. At this time, we will remove the memoryless condition from the channel and allow the noise to be time-dependent and correlated. For the Gaussian additive noise channel, that simply means that the additive  $Z$  process is normal with mean 0 and covariance matrix  $K$ . Thus the channel is of the form  $Y_i = X_i + Z_i$ . The  $X_i$ 's have the previous defined codebook structure, where  $X_i$  depends on the index  $W$  and the previous  $Y$ 's. In addition, for the Gaussian channels, there is a power constraint

$$\frac{1}{n} \sum X_i^2(W, Y_1, \dots, Y_{i-1}) \leq P \quad (2)$$

We immediately have a problem with this because if  $X_i$  depends only on  $W$ , we can guarantee that its power is limited to  $P$ , but if  $X_i$  also depends on the received  $Y$ 's in a nontrivial way, some  $Y$  sequences may cause  $X$  to burst the power constraint. So in general, we can either talk about an expected power constraint on  $X$  averaged over  $Y$  or we shall talk about making an error whenever  $Y$  causes the power constraint to be

violated. Both approaches will yield the same answers below. Let  $C_{FB}$  be the feedback channel capacity for the Gaussian channel with a given covariance structure and  $C_{NFB}$  be the associated capacity when feedback is not allowed. It is shown in Cover and Pombra [5] that

$$C_{FB} = \lim_{n \rightarrow \infty} \max_{\frac{1}{n} \text{tr}(K_X) \leq P} \frac{1}{2n} \log \frac{|K_{X+Z}^{(n)}|}{|K_Z^{(n)}|}$$

$$C_{NFB} = \lim_{n \rightarrow \infty} \max_{\frac{1}{n} \text{tr}(K_X) \leq P} \frac{1}{2n} \log \frac{|K_X^{(n)} + K_Z^{(n)}|}{|K_Z^{(n)}|}$$

The maximization in the definition of  $C_{FB}$  is over  $K_{X+Z}$  in which  $X_i$  and  $Z_i$  are conditionally independent given the past  $X^{i-1}, Z^{i-1}$ . Here it is true that  $C_{FB} > C_{NFB}$ , i.e. feedback *strictly* increases capacity. The reason is due solely to the fact that if the transmitter knows  $Y_i$ , he can determine  $Z_i = Y_i - X_i$  and therefore all the previous terms in the noise process, and, because the noise is not independent, he can predict where the noise is going and combat it.

Incidentally, the notion of combating the noise is somewhat misleading. In fact, many considerations show that one should not fight the noise, but join it. There is more space between the quills of a porcupine if they point out than if they point in. Thus if one is constrained to add some signal power to noise one should add it in the same direction rather than inwards.

The following two theorems limit the effect of feedback for Gaussian channels.

**Theorem 2 (Pinsker and Ebert).**

$$C_{FB} \leq 2C_{NFB} \quad (3)$$

**Proof:** Pinsker stated the result and didn't publish it. Ebert [11] published the result in B.S.T.J. A new proof of the result can be found in Cover and Pombra [5].

**Theorem 3**

$$C_{FB} \leq C_{NFB} + \frac{1}{2} \text{ bits per transmission}$$

**Proof:** (See Cover and Pombra [5].)

Theorem 2 says feedback at most doubles the capacity, and Theorem 3 says that one can at most add one half bit per transmission to the communication. Feedback helps, but not much.

## 4 Unknown channels.

In the previous section, we showed how feedback is used to predict the noise. Here we will discuss a more extreme case, one in which feedback helps by an arbitrary amount.

Consider the following channel discussed by Dobrushin [10]. The input alphabet is  $\mathcal{X} = \{1, 2, \dots, m\}$  and the output alphabet is  $\mathcal{Y} = \{0, 1\}$ . All but the  $i$ -th input symbol is crushed into  $y = 0$ , and the  $i$ -th symbol is received as  $y = 1$ . The transmitter does not know which symbol  $i$  survives. This channel stays constant over many uses. (The channel is stationary but not ergodic.)

First we find the capacity with feedback. This is quite simple. The transmitter first cycles through the integers  $1, 2, \dots, m$ , sending a test sequence which is received by the receiver and fed back. After  $m$  transmissions, both sender and receiver know which symbol  $i$  results in  $y = 1$ . Thereafter, the transmitter simply uses the channel as a binary typewriter sending  $i$  or not  $i$  accordingly as he wishes to send 1 or 0. Thus the feedback capacity is clearly one bit per transmission.

Now, what about the capacity without feedback? Here it turns out that the receiver actually does know which symbol maps into  $y = 1$ . The transmitter sends a test sequence  $1, 2, \dots, m$  and the receiver determines the symbol  $i$  that maps to  $y = 1$ . He is then prepared to do very clever decoding in the future. Unfortunately, the transmitter is still in the dark. The best the transmitter can do is to use a uniform distribution over the  $m$  letters, in which case he is only sending  $Y = 1$   $\frac{1}{m}$ -th of the time. The resulting mutual information is  $I = \frac{1}{m} \log m + ((m-1)/m) \log(m/m-1)$ , which is approximately  $\frac{1}{m} \log m$ . Thus,  $C_{NFB} \approx \log m/m$ .

So here we have it. For large  $m$ , the capacity without feedback  $C_{NFB}$  is very near zero, while the capacity with feedback  $C_{FB}$  equals 1, and the ratio is arbitrarily large. Feedback helps by an arbitrarily large factor.

By changing the example slightly, one can get the additive difference between the two capacities to be as large as possible. So thanks to this example, we see that all is not well in showing that feedback has a bounded effect on the improvement of capacity. Nonetheless, we feel that the factor of two limit and the half bit additive limit from the previous section are typical of practical channels.

## 5 Multiple access channels.

Until now, we have talked about one sender and one receiver. The multiple access channel, modeling satellite communication, has many senders  $X_1, X_2, \dots, X_m$  and one receiver  $Y$ , where  $Y$  has the conditional distribution  $p(y|x_1, x_2, \dots, x_m)$ . Here one has interference among the senders as well as channel noise.

**Theorem 4** (Ahlsvede [3] and Liao [2]). *The capacity region for the 2-user multiple access channel is the convex hull of the union of the set of all  $(R_1, R_2)$  satisfying*

$$\begin{aligned} R_1 &\leq I(X_1; Y|X_2) \\ R_2 &\leq I(X_2; Y|X_1) \\ R_1 + R_2 &\leq I(X_1, X_2; Y) \end{aligned}$$

*for some product distribution  $p(x_1)p(x_2)$ .*

Gaarder and Wolf [7] showed that feedback increases the capacity of multiple access channels. Cover and Leung [1] went on to show that the following region is achievable.

**Theorem 5** *An achievable rate region for the multiple access channel with feedback is given by the convex hull of the union of the set of  $(R_1, R_2)$  satisfying*

$$\begin{aligned} R_1 &\leq I(X_1; Y|X_2, U) \\ R_2 &\leq I(X_2; Y|X_1, U) \\ R_1 + R_2 &\leq I(X_1, X_2; Y) \end{aligned}$$

*for some distribution  $p(u)p(x_1|u)p(x_2|u)$ .*

However, Theorem 5 is in general only an inner bound to the capacity region. Willems [9] has shown that if a certain condition is satisfied, then the above region is indeed the capacity region. The region in Theorem 5 is larger than the no-feedback region, and we conclude that feedback improves the capacity region of a multiple access channel, even when it is memoryless. See [12, 13, 14] for further results in this direction.

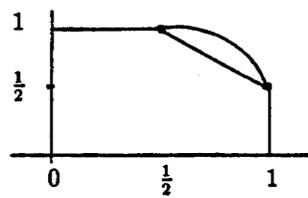
Why does feedback help? The channel is memoryless, so feedback does not help predict the noise. Here capacity is increased because feedback diminishes the interference through cooperation. Each transmitter has a better idea of what the other transmitter is sending than does the receiver  $Y$ . To this extent they can cooperate. Whether they help each other or not, they can at least get out of each other's way in some statistical sense. This is the real reason why feedback helps multiple access communications. It allows the senders to cooperate. However, this improvement in capacity is limited.

J.A. Thomas has shown that an arbitrary multiple access channel with two users has its capacity region at most doubled with feedback. He has gone on to show that for  $m$ -user memoryless Gaussian multiple access channels that feedback at most doubles capacity. He conjectures that capacity is at most doubled with feedback for arbitrary non-Gaussian  $m$ -user multiple access channels.

For multiple access channels, feedback helps, but not by very much.

## 6 Computational complexity for multiple access channels with feedback

Consider the binary erasure multiple access channel in which  $X_1, X_2 \in \{0, 1\}$ ,  $Y \in \{0, 1, 2\}$ , and  $Y = X_1 + X_2$ . The capacity region for this channel is shown in Figure 2.



Capacity Region

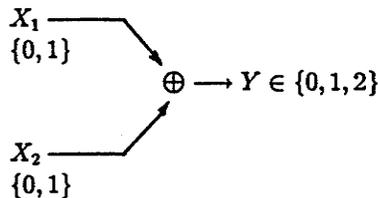


Figure 2: Binary Erasure Multiple Access Channel.

Here it is known that the region in Theorem 5 is indeed the capacity region. This is strictly greater than the no-feedback capacity region. So feedback helps capacity. The question is, how does one send information over this channel using feedback? Suppose  $X_1$  is sending at rate  $R_1 = 1$ . Thus  $X_1$  sends an arbitrary sequence of zeros and ones, each occurring about half the time. How then can  $X_2$  achieve the rate  $\frac{1}{2}$  corresponding to the point  $(R_1, R_2) = (1, \frac{1}{2})$  on the capacity boundary? Here is the strategy, relayed to me by J. Massey. Let  $X_1$  and  $X_2$  arbitrarily send their zeros and ones. For example:

$$\begin{array}{r}
 X_1 = 0 \ 0 \ 1 \ 0 \ 0 \ 1 \ 1 \\
 X_2 = 0 \ 0 \ 1 \ \boxed{1 \ 1 \ 1} \\
 \hline
 Y = 0 \ 0 \ 2 \ \boxed{1 \ 1 \ 2}
 \end{array}$$

Notice that when  $Y$  equals 0 or 2, the receiver knows instantly the values of  $X_1$  and  $X_2$ . However,  $Y = 1$  acts as an erasure. We don't know whether  $(X_1, X_2) = (0, 1)$  or  $(1, 0)$ . At this point,  $X_1$  continues sending whatever he wishes to send. After all, he is sending at rate  $R_1 = 1$ . But  $X_2$  now continues to retransmit whatever it was

that he sent when the ambiguous  $Y = 1$  was received. He continues to do this until  $Y$  equals either 0 or 2, thus correcting the erasure. It is then a simple matter for  $Y$  to go back and correctly determine previous values of  $X_1$  and  $X_2$ . So  $X_2$  has to send on the average two symbols for every one that gets through, achieving rate  $R_2 = \frac{1}{2}$ .

This point on the boundary of the capacity region could not have been achieved as simply without feedback.

## 7 Conclusion.

We have now examined many cases where feedback does and does not help the communication of information. We now go back over the previous questions and answer them with respect to these examples.

### Possible Roles of feedback:

**Correct receiver's misunderstanding?** Feedback does not increase capacity for memoryless channels, so it does not aid in correcting  $Y$ 's misunderstanding. On the other hand, feedback improves the error exponent and it helps reduce the complexity of the communication. Indeed, for additive Gaussian noise channels, the Kailath-Schalkwijk [6] scheme sends correction information to  $Y$  and achieves capacity.

**Predict and correct the noise?** Here feedback helps the capacity if the noise is dependent. On the other hand, the improvement in capacity is less than or equal to a factor of two for Gaussian additive noise channels regardless of the dependence. Also, one does not really correct the noise, but joins it in some sense.

**Improve error exponent?** Feedback helps.

**Cooperation of the senders?** Feedback allows cooperation, increases the capacity and lowers the computational complexity.

**Determination of the channel?** A simple test sequence can be used to determine the entire channel distribution  $p(y|x)$ . One can then use the code appropriate for that channel and achieve channel capacity as well as if one had known ahead of time what the channel was. Feedback definitely helps - sometimes by arbitrary factors.

**Reduction of delay?** Feedback can greatly reduce delay. The examples show small delays for many of the channels in which feedback is used. Feedback allows multiple users of satellite and computer networks to share a common channel with minimal delay.

**Reduction in computational complexity?** Feedback helps.

In summary, feedback helps communication, but not as much as one might think. It simplifies communication without greatly increasing the rate of communication.

**Acknowledgment:** I would like to thank S. Pombra and J.A. Thomas for contributing their ideas to this discussion. J.A. Thomas also brought Dobrushin's example to my attention.