Experiments with Linear Prediction in Television

By C. W. HARRISON

(Manuscript received February 28, 1952)

The correlation present in a signal makes possible the prediction of the future of the signal in terms of the past and present. If the method used for prediction makes full use of the entire pertinent past, then the error signal—the difference between the actual and the predicted signal—will be a completely random wave of lower power than the original signal but containing all the information of the original.

One method of prediction, which does not make full use of the past, but which is nevertheless remarkably effective with certain signals and also appealing because of its relative simplicity, is linear prediction. Here the prediction for the next signal sample is simply the sum of previous signal samples each multiplied by an appropriate weighting factor. The best values for the weighting coefficients depend upon the statistics of the signal, but once they have been determined the prediction may be done with relatively simple apparatus.

This paper describes the apparatus used for some experiments on linear prediction of television signals, and describes the results obtained to date.

INTRODUCTION

Linear prediction is perhaps the most expedient elementary means of removing first order correlation in a television message. Before discussing the advantages and disadvantages of linear prediction, it might be well to consider what is generally meant by correlation in a television picture and why it should be removed.

Almost every picture that has recognizable features contains both linear and non-linear correlation. Each type of correlation helps in identifying one picture from another; however, linear prediction is only effective in removing linear correlation, and for this reason, future references to correlation will refer only to its linear properties. With television, a signal is obtained as the result of scanning; hence, the correlation is evident in both space and time. Briefly, correlation is that relation which the "next" elemental part of the signal has with its past.

To leave correlation in a message is to be redundant, and this effectively loads the transmission medium with a lot of excess "words" not necessary to the description of the picture at the receiving end. It is then more "efficient" to send only the information necessary to identify the picture, and to restore the redundancy at the receiver.

EFFICIENT TRANSMISSION

The more *efficient* we are in sending pictures over a given transmission line, the more alarmed we become at the increasing amount of equipment that is required at the transmitting and receiving terminals. Certainly the design will be a compromise between the complexity of apparatus and the efficiency achieved. The ingenuity of engineers will be taxed along these lines for years to come; however basically, the general form of these systems will be similar to that shown in Fig. 1. Although not always separable, four essential operations are required-namely, decorrelating, encoding, decoding and correlating. The transmitting decorrelator and the encoder encompass the principal design problems, since the decoder and correlator at the receiving end perform the reverse operations which interpret the code and add in the redundancy that was removed.

Decorrelation involves prediction, and as the predictors are more nearly made to predict the future of the signal, the more the output signal from the decorrelator resembles random noise. The essential picture information is still present, which means that our original picture signal can be obtained at the receiving end without theoretical degradation. The basic job of the encoder is to match the picture information out of the decorrelator to the channel over which it is to be transmitted. There are several encoding operations. The first concerns the rate of information into the encoder, and that required out of it. In the case of television, there are flat, highly correlated areas as well as areas containing more concentrated detail. This means that the information

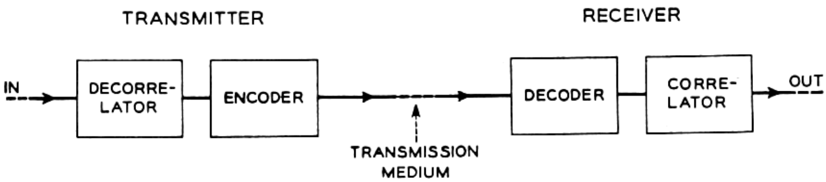

Fig. 1—Block diagram of an efficient transmission system employing reversible decorrelating and encoding means.

rate varies when the picture is scanned at the conventional uniform scanning rate. The output of the encoder feeds a transmission line that has a definite channel capacity, and if maximum efficiency is to be obtained from this transmission medium, then the rate of information into it must be held relatively uniform at a value near the channel capacity. It is the job of the encoder to take the varying rate of information from the decorrelator and feed it to the channel at a constant rate. At the receiving end, the decoder must take the constant rate of information and deliver it to the correlator at the variable rate as originally fed into the transmitter's encoder. Thus, to perform this task, a variable or elastic delay to run ahead or behind, depending on the information content of the picture being scanned, is an important part of the encoder. Over a long period of time, the variable delay would average out to some fixed value. This variable delay must never run out, even when the detail is concentrated. There are instances when this condition could not be met, such as an extended reproduction of a snow storm; however, with good design the system should fail "safe"—a slight degradation of picture quality. This condition can be made infrequent enough to cause little concern.

The encoder design must also account for noise as well as bandwidth of the channel and must consider the ultimate effect of an error that may be introduced by noise along the transmission line. As more redundancy is removed to get at the "essence" of the picture signal, the more important it is to guard this "essence," as mistakes presented to the receiver will propagate themselves longer in the absence of correlation. Errors can be minimized by rugged systems of modulation such as PCM, where the signal-to-noise ratio of the transmission line determines the base of the PCM system selected. In any event, the encoder must send the information so that the effect of errors will not appreciably disturb the picture.

DECORRELATION AND LINEAR PREDICTION

Fig. 2 illustrates, in a general way, a means of decorrelating the signal, $S_1(t)$. For purposes of explanation, the encoder and decoder have been omitted, and the transmission between the receiving and sending terminals, idealized. The predictors, P, are identical, and base their prediction, $S_p(t)$, on the signal's past history. In this way, the output of the computer represents the discrepancy between the actual value of the signal sample and the predictor's prediction. By this means we are sending only our mistakes—the amount by which the next picture element surprises us. For example, if the computer is so designed that it

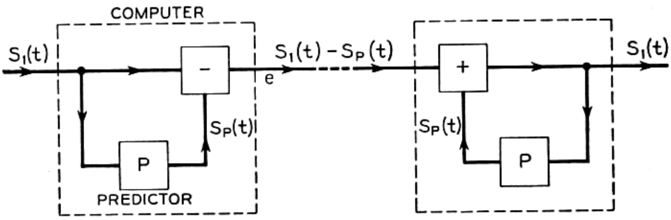

Fig. 2—Decorrelator and correlator showing reversible nature of this method of removing redundancy.

bases its prediction on the "previous frame," and we are transmitting a "still," there will be no surprises after the first frame and consequently no output signal. Certainly it is redundant to send the same picture more than once.

Linear prediction provides an easily instrumented means of removing redundancy. With linear prediction the next signal sample is simply the sum of the previous signal samples, each multiplied by an appropriate weighting factor. The best values for these weighting coefficients depend on the statistics of the signal.

Fig. 3 is a block diagram of a decorrelator employing linear prediction. The delayed versions of the input signal can be obtained from taps along the delay line. The weighting coefficients for each of the delayed signals are selected by loss in their respective paths as shown by the amplitude controls. The polarity of each signal can be determined by the switches. The output is simply the sum of these weighted signals.

If we consider the signal on a continuous basis (not quantized or sampled), linear predictors can be characterized as ordinary linear filters used to predistort the frequency spectrum of the signal. As such, they can be designed in the frequency or time domain. However as will

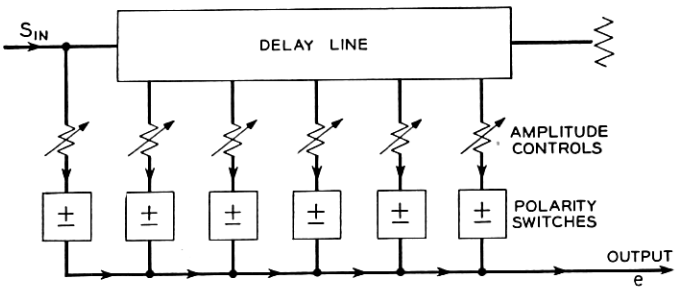

Fig. 3—General block diagram of decorrelator employing linear prediction. Linear prediction bases its prediction on the weighted sum of previous signal samples.

be shown, it is much easier to recognize circuit configurations that reduce redundancy in the time domain. To this end, and for purposes of encoding, the signal is thought of as signal samples uniformly spaced at Nyquist intervals. Thus, amplitude values obtained by sampling a 4.0 mc picture signal at $\frac{1}{8}$ microsecond intervals serve to specify the signal completely. Fig. 4 shows a small portion of a television raster where the signal is represented by signal values spaced at Nyquist intervals, τ. The coordinates shown are designated with respect to the "present value" of the signal, $S_{0,0}$. The positive coordinate directions are shown by the arrows. The past is represented by positive coordinates—the future by negative coordinates. In this way, the previous value of the signal

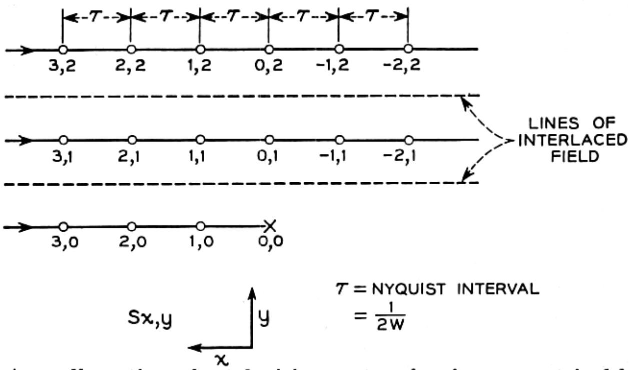

Fig. 4—A small portion of a television raster showing geometrical location of signal samples with relation to the "present value" of the signal, $S_{0,0}$.

taken one Nuquist interval before $S_{0,0}$ is designated by $S_{1,0}$ – the previous line samples by $S_{0,1}$, etc.

METHODS OF LINEAR PREDICTION

As previously stated, with linear prediction the next signal sample, $S_p(t)$, is simply the sum of the previous signal samples each multiplied by an appropriate weighting factor. Thus,

$$S_p(t) = a_{1,0}S_{1,0} + a_{2,0}S_{2,0} + a_{3,0}S_{3,0} + \cdots a_{m,n}S_{m,n}$$

represents the weighted sum of all the previous signal values. The error signal, e, as shown in Fig. 2, is represented by the difference between the present vaue of the signal, $S_{0,0}$ and the predictor's prediction.

$$e = S_{0,0} - S_p(t)$$

There are several specific types of linear prediction that deserve further explanation—namely, "previous value," "slope," "previous line," "planar" and "circular."

"*Previous value*" prediction is illustrated in Fig. 5. Here the prediction is taken to be the signal amplitude of the preceding picture element. The previous amplitude of the signal, $S_{1,0}$ is subtracted from the present value of the signal, $S_{0,0}$. The error signal is given as

$$e = S_{0,0} - S_{1,0}$$

The filter characteristic can be expressed as

$$F(\omega) = \left(2 \sin \frac{\omega\tau}{2}\right) \epsilon^{i\left(\frac{\pi}{2} - \frac{\omega\tau}{2}\right)}$$

This method of prediction proves to be rather effective in reducing the average power for most television pictures. The expression for the filter characteristic given above shows that the peak amplitude can be twice that of the original signal.

"*Slope*" prediction is illustrated by Fig. 6. "Slope" prediction is so called because it is equivalent to passing a straight line through the two previous signal values, with the assumption that this line will pass through the next signal value. The predicted signal is given by

$$S_p = 2S_{1,0} - S_{2,0}$$

The frequency and phase characteristic is expressed as

$$F(\omega) = [4 \sin^2 \omega\tau]\epsilon^{i(\pi - \omega\tau)}$$

For this method of prediction, the peak amplitude of the error signal can be as much as four times the peak amplitude of the original signal.

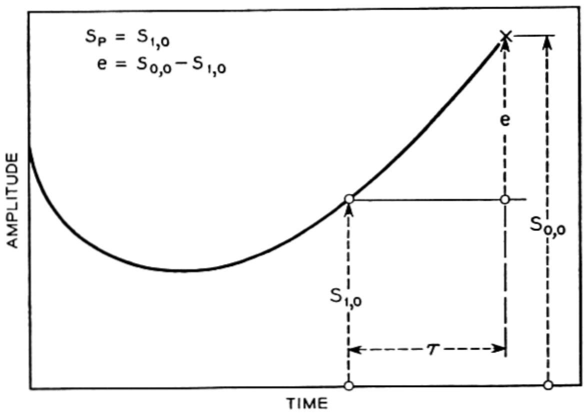

Fig. 5—Example of "previous value" prediction, where the error signal is the difference between the actual value of the signal and the previous value.

It is of interest to mention that "slope" prediction is equivalent to two "previous value" predictors in tandem. Three or more "previous value" predictors in tandem are equivalent to a binomial weighting of the previous values of the signal to form a predicted signal.

For example, the prediction for three "previous value" predictors in tandem is given by

$$S_p = 3S_{1,0} - 3S_{2,0} + S_{3,0}$$

For four "previous value" predictors in tandem

$$S_p = 4S_{1,0} - 6S_{2,0} + 4S_{3,0} - S_{4,0}$$

As can be seen from the above equations, further extension of "previous value" tandem operation results in a heavier weighting of picture elements further and further from the point to be predicted. For most pictures this leads to greater errors.

"*Previous line*" prediction, shown in Fig. 7, would be expected to be similar to previous value prediction, since a picture would presumably have approximately the same correlation vertically as it does horizontally. This would be the case except that our interlaced scanning system makes the previous line signal some 28 per cent further away from $S_{0,0}$ than the closest horizontal sample, $S_{1,0}$. The error signal, e, is given by where T is a line time. The error output has a maximum peak amplitude of twice the input.

$$e = S_{0,0} - S_{0,1}$$

The filter characteristic can be expressed as

$$F(\omega) = \left[2 \sin \frac{\omega T}{2}\right] \epsilon^{i\left(\frac{\pi}{2} - \frac{\omega T}{2}\right)}$$

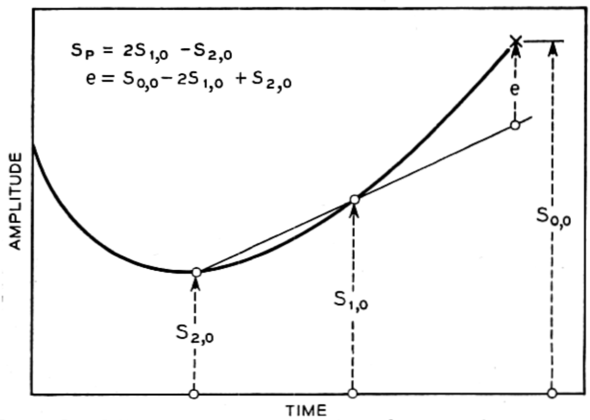

Fig. 6—Example of "slope" prediction. Here the next signal value is assumed to lie on a straight line that intersects the two previous signal values.

"*Planar*" prediction, shown in Fig. 8, is effectively tandem operation of "previous value" and "previous line" prediction. Planar prediction may also be thought of as the value represented by a plane above the present value of the signal when passed through three adjacent signal

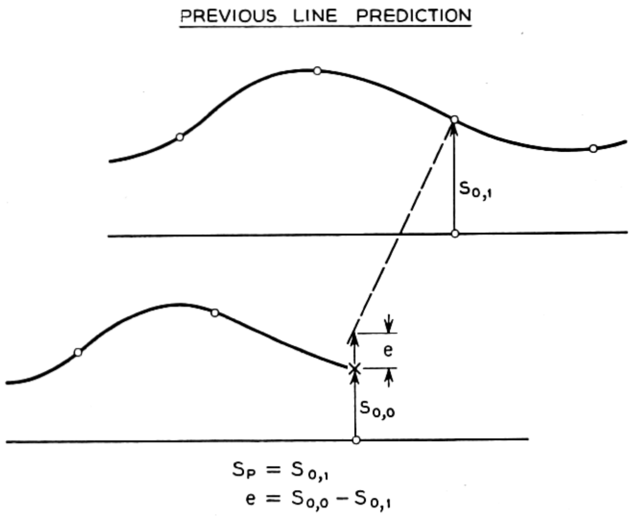

Fig. 7—Example of "previous line" prediction. Here the error signal is the difference between the actual value of the signal and the value of the signal on the line directly above.

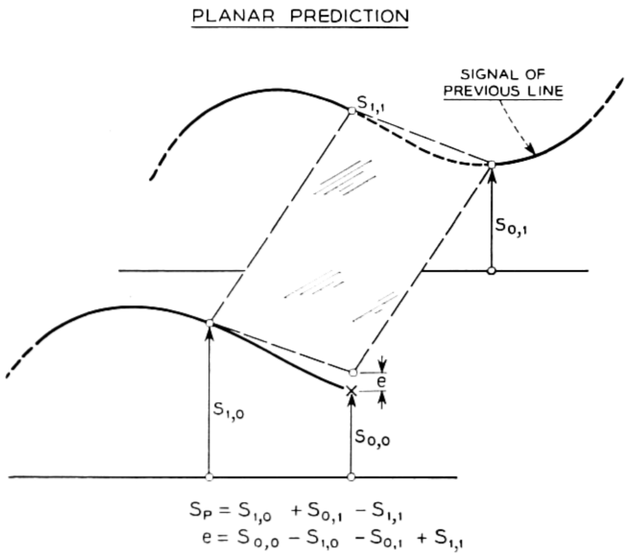

Fig. 8—An example of "planar" prediction. Here the prediction is represented by a plane that has been passed through three adjacent signal values.

values, namely $S_{1,0}$, $S_{0,1}$ and $S_{1,1}$. The predicted signal is given by

$$S_p = S_{1,0} + S_{0,1} - S_{1,1} .$$

The filter characteristic is given by

$$F(\omega) = 4 \sin \frac{\omega\tau}{2} \sin \frac{\omega T}{2} \epsilon^{i\left(\pi - \frac{\omega}{2}(T+\tau)\right)}$$

The peak error amplitude for "Planar" prediction can be four times that of the input signal.

"Planar" prediction has several good characteristics. For example, if $S_{1,1}$ and $S_{1,0}$ were white and $S_{0,1}$ black, then $S_{0,0}$ would be predicted to be black. Thus a change horizontally from white to black would produce no errors. Similary, if $S_{1,1}$ and $S_{0,1}$ were white and $S_{1,0}$ black, then $S_{0,0}$ would be predicted to be black. This indicates that a change vertically from white to black would be predicted correctly. In this manner, all vertical and horizontal contours in a picture are deleted. This philosophy can be extended to include other directions as well.

"*Circular*" prediction, illustrated in Fig. 9, is an extension of planar, since it deletes horizontal, vertical and 52° contours as well. A total of 190.5 microseconds of delay is required, making the required equipment more elaborate. Also, as more delay is required, more noise is added.

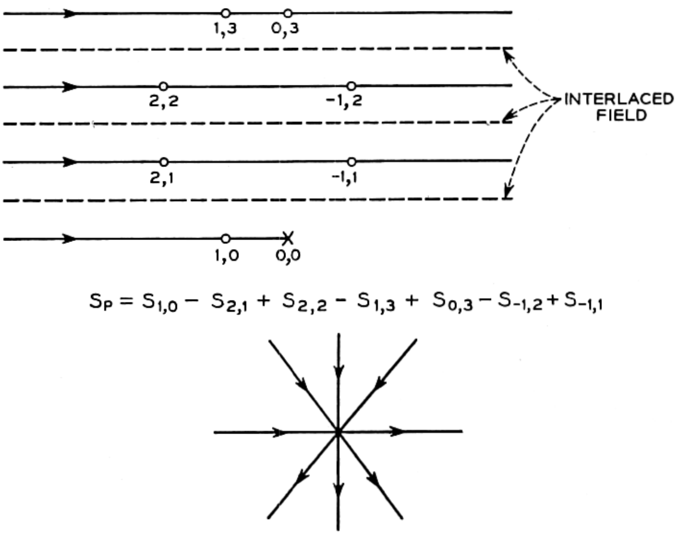

Fig. 9—Past signal samples required for "circular prediction"—a type of prediction which removes horizontal, vertical, and ±52° straight line picture contours.

Therefore, indefinite extension of this straight line contour deleting philosophy is not a paying means of prediction, at least not at the present state of the art of wide band delay lines. Furthermore, the increasing diameter of the circle for extension of circular prediction would decrease its accuracy for finely concentrated detail.

Fig. 10 shows the relative position of picture elements nearest $S_{0,0}$ if a wide band field delay were available. The methods of prediction dis-

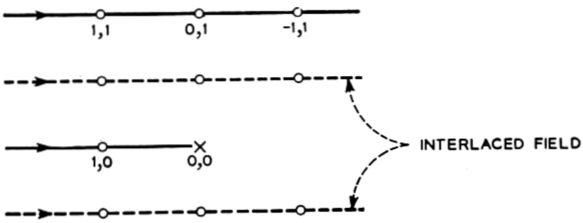

Fig. 10—Small portion of television raster showing signal samples, including those of previous field, which would enable time extrapolation-space interpolation as a method of prediction.

cussed have been essentially an extrapolation in space; however, with a field delay, interpolation in space, and extrapolation in time would also be possible.

EXPERIMENTAL CIRCUITRY

Experimentally, those types of predictors that involve only a few Nyquist intervals of delay are easiest to mechanize. Fig. 11 shows a simplified schematic of a decorrelator that enables an evaluation of linear prediction schemes having error signals given by $e = a_{0,0}S_{0,0} \pm a_{1,0}S_{1,0} \pm a_{2,0}S_{2,0}$. This enables an evaluation of "previous value" and "slope" prediction. The signal is fed into a terminated delay line having taps at Nyquist intervals. Each of these signals is individually attenuated by the potentiometers in the cathode circuit of the cathode followers. Each output is then fed to its respective polarity switch. The D.P.D.T. switch determines to which side of the differential amplifier, V_4, the particular signal is sent. Since more than one signal may require the same polarity, the signals are combined through "L" type resistance attenuators to prevent interaction between signals. The D.P.D.T. switches are so arranged that the other signals are unaffected when a polarity switch is reversed. The differential amplifier, V_4, is a cathode coupled circuit having the advantage of two identical grids which produce opposing effects in the output. The output is then matched to the line by the cathode follower, V_5. In this way we can transmit (1) the original picture signal with either polarity and any amplitude, (2) the picture signal delayed by one or two Nyquist intervals with either polarity and any amplitude or (3) any linear combination of (1) or (2).

Fig. 12 is a block diagram of the experimental set-up used to investigate prediction methods that involve previous value and previous line samples such as "planar," etc. The input is fed into a manually variable delay having 0.1 Nyquist interval steps. This delay line acts like a vernier for the 63.5 microsecond line delay. Effectively, it enables the previous line samples to be positioned directly above the previous value samples.

The 63.5 microsecond delay is a so-called "acoustic" or "ultrasonic" delay line and was developed by Mr. H. J. McSkimin. The associated circuitry was developed by Mr. A. L. Hopper.* Storage is accomplished by a fused silica bar with quartz transducers operating at a carrier

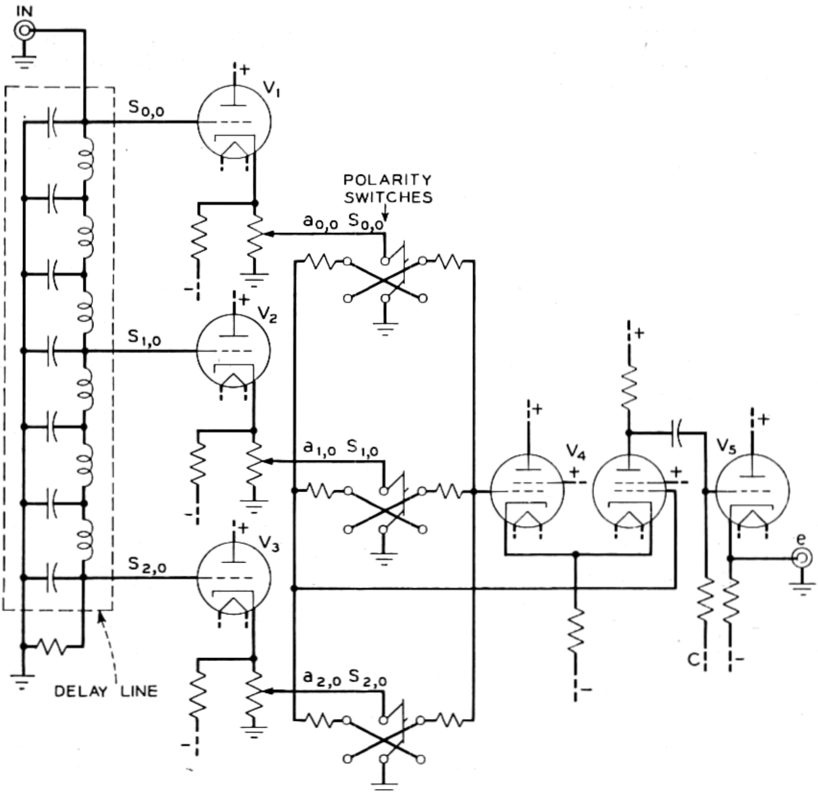

Fig. 11—Simplified schematic for "previous value" and "slope" prediction.

* A. L. Hopper, "Storing Video Information," *Electronics*, **24**, pp. 122–125, June, 1951.

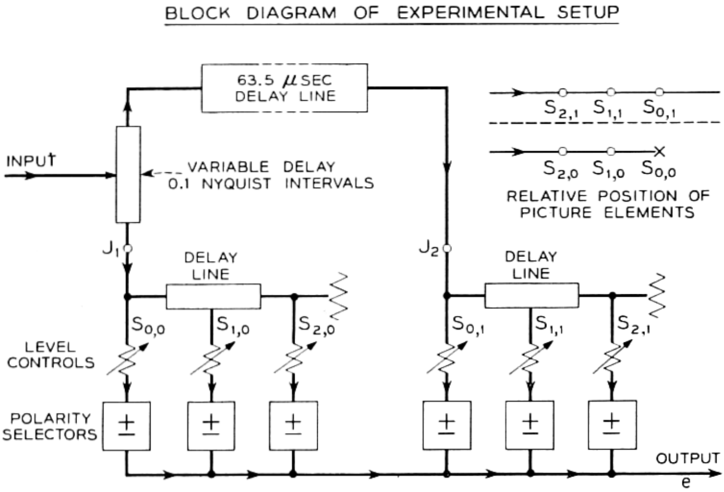

Fig. 12—Block diagram of experimental apparatus used to investigate methods of prediction involving combinations of previous signal values along a line with those on the line directly above.

frequency of 54.0 mc. The over-all video bandwidth is flat (± 0.1 db) to 5.0 mc. Nonlinear distortion is approximately one percent when the peak-to-peak signal to r.m.s. noise is 58 db. To give an idea of its complexity, two such units with their associated power supplies require a seven-foot relay rack for housing. The signal at J_1 represents the input picture signal, even though it may be delayed by a small fraction of a Nyquist interval from the actual input signal. The signal at J_2 is the same signal as found at J_1 but delayed by one line time. Each of these signals is fed into terminated delay lines to enable additional signal samples to be obtained. The geometrical location of these signal samples is illustrated in the upper right section of Fig. 12. Here, six signals are obtained instead of three as were required for "previous value" and "slope" prediction. These signals are weighted and polarized in the same manner as the three signals shown in Fig. 11. The output is the sum of these weighted signals and is given by

$$e = a_{0,0}S_{0,0} + a_{1,0}S_{1,0} + a_{2,0}S_{2,0} + a_{0,1}S_{0,1} + a_{1,1}S_{1,1} + a_{2,1}S_{2,1}\,.$$

The coefficients may assume positive or negative values.

MEASUREMENTS

It is obvious that if we are able to predict the value of most signal samples closely (which we will be able to do if there is a large amount of correlation in the picture), then the average amplitude of our mistakes will be much less than the average amplitude of the original signal. Thus, by using the decorrelator alone, we can send a message over a channel with the same bandwidth as before but with less average power. At first, this might sound like a worthwhile saving; however this lower average power is accompanied by an even higher peak amplitude which makes any direct saving less attractive. Furthermore, the low frequency attentuation of the decorrelator makes the signal vulnerable to low frequency disturbances, since the correlator must restore (emphasize) these low frequency components.

A proper but *not entirely adequate* method of evaluating the effectiveness of a predictor is by measuring the ratio of signal power to error power. This is called "Power Reduction" and is generally expressed in db. Power reduction simply provides a scale by which we can weigh a linear predictor's capabilities. The "not entirely adequate" refers to the fact that minimum error power may not provide simultaneously the lowest amount of redundancy for that given type of prediction.

As an example, Fig. 13 shows the power reduction for the relative weighting of the previous horizontal signal sample as compared to the present value of the signal, for three pictures-later to be described as Scene A, B and C. The top-most curve is for Scene B, which is a simple, soft picture that contains very little detail. For this picture, the minimum error power coincides (within measurable limits) with the minimum redundancy. For Scene A and particularly Scene C, minimum error power is considerably different than that for minimum redundancy. This difference between minimum error power and minimum redundancy also applies to decorrelators using other types of predictors as well. Minimum redundancy may also be a misleading criterion of a predictor's performance, since the value of the prediction must depend on the particular type of encoder used, and some types of encoding will require certain types of redundant information to be retained.

The following pictures are representations of the error signal as photographed from a 10-inch laboratory monitor. The signals were band limited to 4.3 mc. Fig. 14 represents the "original" for three scenes called A, B and C. These pictures represent, to a first approximation, the gamut of pictures normally expected to be transmitted. They are by no means the best or the worst pictures than can be imagined; however any system should be able to reproduce these pictures without appreciable distortion. For example, Scene C should be capable of being sent continuously without the elastic delay running out, etc.

Fig. 15 shows how the error signal appears for "previous value" prediction. "Previous value" prediction is excellent for flat white or dark

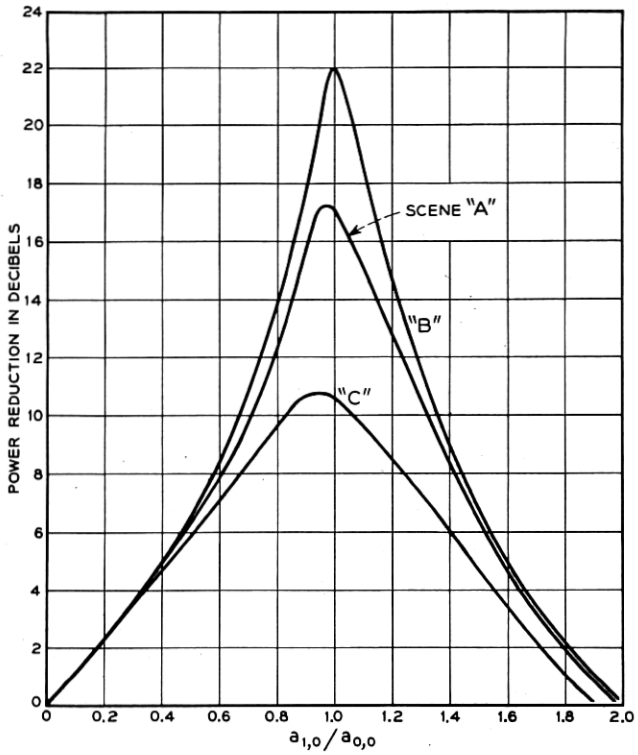

Fig. 13—Power reduction for various weighting coefficients for the previous horizontal sample.

areas as can be observed in the background of Scene A. Where the separation of a white to black area is made, the error signal is large. It is this large error signal that informs the receiver of this change in brightness, and until another change occurs, the error output is again zero. This type of performance produced the flat grey appearance of the background. In this way, the picture represents only changes in brightness—a first difference type of picture.

It may be noted that horizontal contour lines have vanished leaving only vertical contours which pertain to the brightness changes that have occurred. This effect is especially evident in Scene C. The power reductions given in the lower left hand corner of these pictures are consistent with their complexity.

Fig. 16 shows the error signal appearance for "slope" prediction. When compared to the error signal for "previous value" prediction a finer vertical granularity is observed, and this is attributed to sudden

SCENE A

SCENE B

SCENE C

Fig. 14—Three pictures as photographed from the face of a kinescope. Scene "A" is a picture of average complexity. Scene "B" is a simple, rather soft picture. Scene "C" is a complex, highly detailed picture. Roughly, these pictures represent the gamut of pictures normally expected to be transmitted.

SCENE A

SCENE B

SCENE C

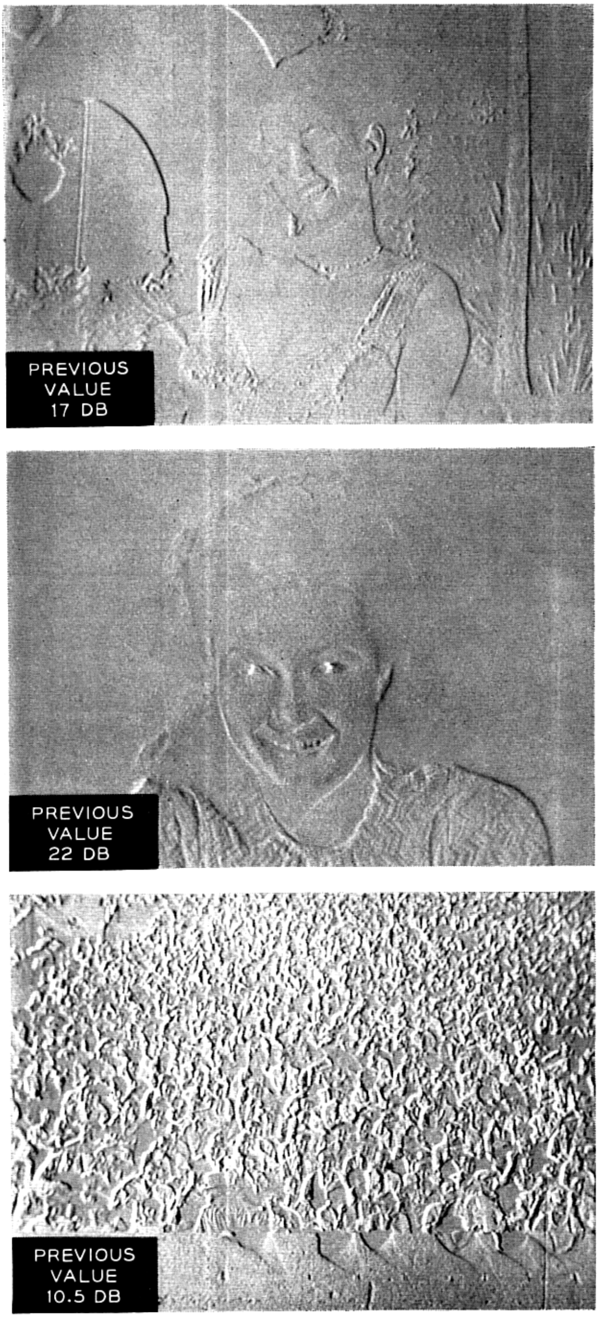

Fig. 15—Three pictures showing the appearance of the error signal when using "previous value" prediction. Note the absence of horizontal contours.

SCENE A

SCENE B

SCENE C

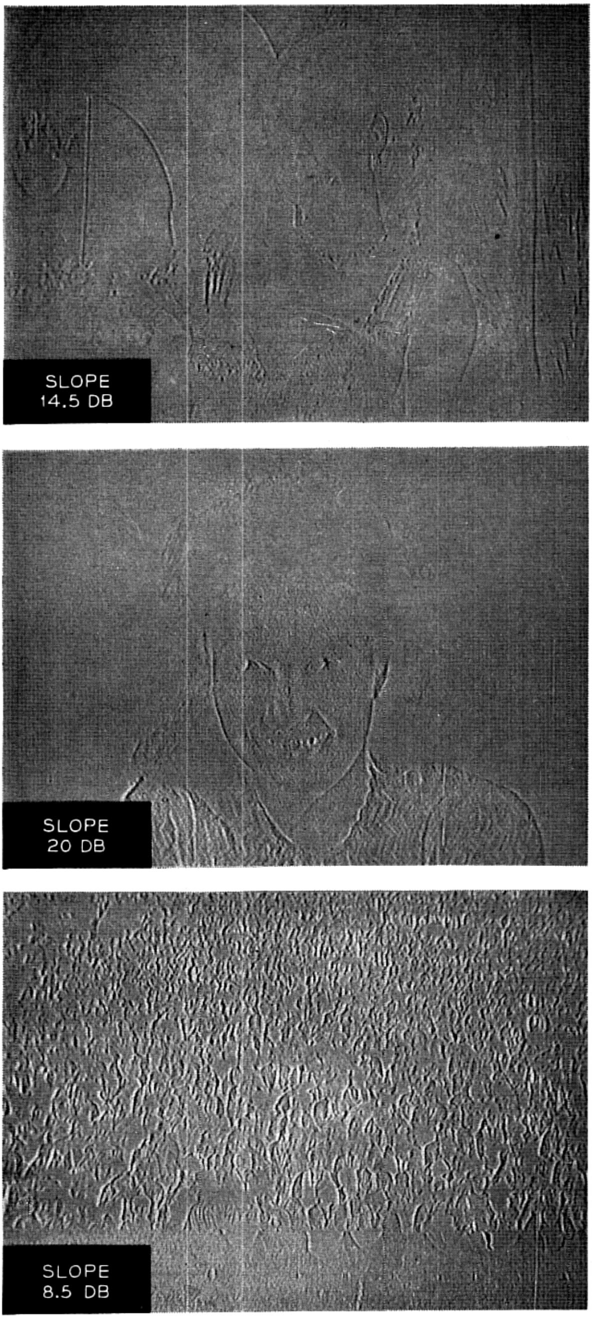

Fig. 16—Three pictures showing the appearance of the error signal when using ‘slope’’ prediction.

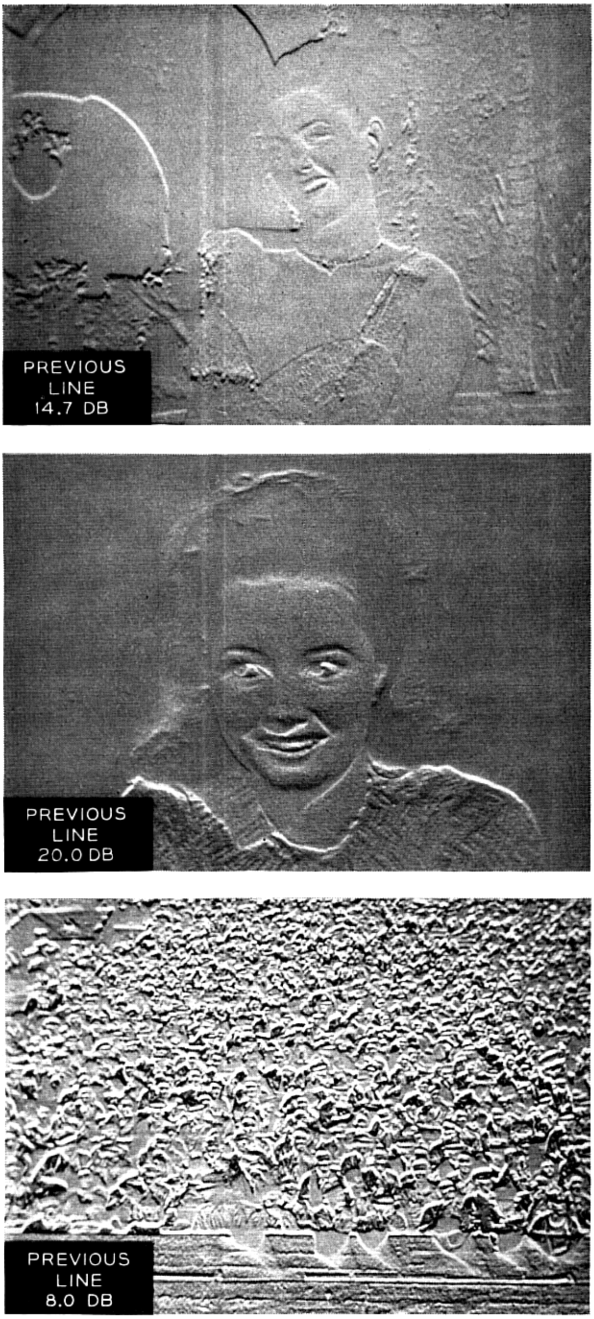

Fig. 17—Three pictures showing the appearance of the error signal when using "previous line" prediction. Note the absence of vertical contours.

SCENE A

SCENE B

SCENE C

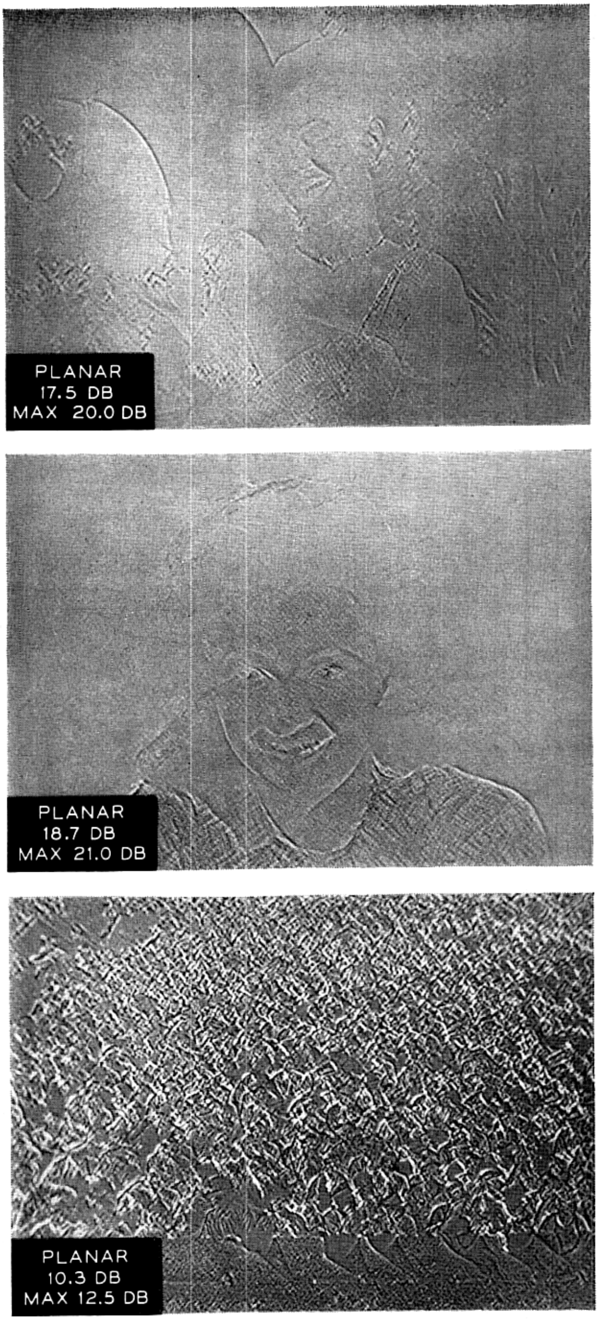

Fig. 18—Three pictures showing the appearance of the error signal when using "planar" prediction. Note the absence of horizontal and vertical contours.

changes in brightness. In the case of "previous value" prediction a sudden change in brightness produces only one error, where for "slope," two errors result. This, to some extent, accounts for the lesser amount of power reduction measured for these scenes.

Fig. 17 shows the appearance of the error signal for "previous line" prediction in the three scenes. Where vertical contour lines are predominately left after "previous value" prediction shown in Fig. 15, horizontal contours, are more prevalent now. It can be noted that the power reduction for "previous line" prediction is less than that for "previous value" prediction. This is due principally to the increased distance of the previous line sample from $S_{0,0}$. If the closest horizontal sample was taken at the same distance from the present value of the signal as the previous line sample, then the power reduction using these signal values individually for prediction would be essentially the same for most pictures.

Fig. 18 shows the error signal appearance for "planar" prediction. Here, vertical as well as horizontal contours are deleted. In Scene A the tree trunk has almost completely vanished. In Scene B the picture has an extremely flat appearance. Scene C exhibits the lack of horizontal and vertical contours best, since only sloping contours are left. The power reduction figures at the lower left hand corner also show values for minimum error power. For most pictures, the error power can be reduced by a factor of one-half again over the planar coefficients by modifying the weighting coefficients. The coefficients for this modified planar case are given by

$$S_p = \tfrac{2}{3}S_{1,0} + \tfrac{2}{3}S_{0,1} - \tfrac{1}{3}S_{1,1}$$

These coefficients generally produce an error signal with less power than the coefficients used for "planar" prediction.

While all pictures contain redundancy, the error signals from these simple linear predictors shown in Figs. 15, 16, 17 and 18 can visually be noted still to contain large amounts of redundancy. The contours of the models and of the various objects are readily identifiable. Were all redundancy removed, the picture would be completely chaotic and would appear very much like random noise, although greater efficiency in transmission would be achieved. For richer rewards, more sophisticated methods of prediction will be required.

ACKNOWLEDGMENT

The author wishes to acknowledge with grateful appreciation the invaluable guidance of Dr. B. M. Oliver. It was principally through his efforts that this study was made possible.