A Method of Coding Television Signals Based on Edge Detection

By BELA JULESZ

(Manuscript received February 5, 1959)

A method is described for transmitting digitalized video signals to reduce channel capacity from that needed for standard PCM. This method takes advantage of the inability of the human eye to notice the exact amplitude and shape of short brightness transients. The transmitted information consists of the amplitudes and times of occurrence of the "edge" points of video signals. These selected samples are coarsely quantized if they belong to high-frequency regions, and the receiver then interpolates straight lines between the samples. The system was simulated on the IBM 704 computer. The processed pictures and obtained channel-capacity savings are presented.

I. INTRODUCTION

There is an increasing trend in the communication field to utilize the physiological and psychological properties of the ultimate receiver — the human observer. Some of these properties were applied many years ago in establishing television transmission standards — for example, visual acuity and flicker-fusion frequency thresholds. The development of information theory made this trend even more apparent, particularly in Shannon's first coding problem, where he posed the question of finding an optimum code for a continuous information source when the fidelity criterion of the receiver was given.

Unfortunately the fidelity criteria of human observers are not known. This lack of knowledge is particularly apparent in visual processes, even though in this field the challenge of possible channel-capacity saving is tempting. From a theoretical standpoint, the solution of the first coding problem must be postponed until enough psychological data are collected. But, from a practical point of view, it is possible to overcome this barrier. Instead of searching for human fidelity criteria, we can proceed in the following simpler way.

First, we take the present television pictures of toll quality as a standard. Then we process the signals of the television source by performing some reasonable operations which reduce information rate, and compare by mere inspection the resulting pictures with the standard pictures. If, for a well-chosen class of different pictures representing most of the possible cases, the results of statistical preference tests do not discriminate between the processed pictures and the standard toll quality pictures, we can regard the obtained rate of information as a practical upper bound. If we choose a processing which codes the continuous source in binary digits, and assume an error-free binary channel for transmission (e.g., PCM as a good approximation), we can ensure that no further picture quality degradation will occur. Thus, the viewer will get the same quality of pictures he was accustomed to seeing, but with less channel capacity.

As we see from the above considerations, the search for an optimum code becomes a trial-and-error procedure. The problem now is to find reasonable operations for the processing. They should be based on psychological facts or hypotheses and should not be too complicated for realization. In the last few years several ideas have been tried out along these lines, with more or less success.[1,2,3] The complexity of the required instrumentation limited or prevented a thorough investigation of these ideas. However, the rapid development of general-purpose digital computers has made it possible to test new ideas without actually building equipment. We can simulate any system on a computer by writing a program which converts the general-purpose computer to a special-purpose computer. Special input and output transducers convert the input pictures to sequences of digitalized numbers and, after processing them, reconvert the output of the computer to pictures. Such equipment was developed by and is used now in the Visual and Acoustical Research Department of Bell Telephone Laboratories as a valuable research tool.[4,5] For the processing we use an IBM 704 computer. Although at present we cannot perform the simulations of television coding schemes in real time on the existing computers, we can evaluate many aspects of a system's performance without building it. Thus, it is possible to compare systems and choose the best one before actual realizations.

II. PROPERTIES OF TELEVISION SIGNALS AND OF THE HUMAN RECEIVER RELEVANT TO EDGE DEFINITION

This paper describes a system which transmits only certain points of a television signal, depending on some given signal properties, and, after reception, interpolates between the points according to a given law. Several similar systems are described which differ only in the criteria by which the transmitted points are selected and coded, and in the function used for the interpolation.[6,7] In our case these criteria were chosen to match some properties of both the usual television pictures and vision.

Television waveforms, in contrast to acoustical signals, include fast transients followed by horizontal or slowly changing sections, and are relatively poor in damped oscillations. Because of this, a recent system[6] which transmits only the extremals of acoustical signals and interpolates the output at the receiver according to a given law is not suitable for television. We tried out this system by simulating it on the computer and in the pictures on the left in Fig. 6 results are shown — that systems which perform well for acoustics may not work for vision. Furthermore, there is experimental evidence that the human eye is not very sensitive to the exact amplitude and shape of sudden brightness changes, but is able to locate the starting and ending points of these brightness transients fairly accurately. (The meaning of this property will be made clearer by quantitative results explained in the course of this paper.) Because of these properties of the source and of the ultimate human receiver, we chose to transmit only the end points of the brightness transients. Provided the standard horizontal scanning technique is used, it is quite simple to give a mathematical criterion for selecting such "edge" points.

To locate an edge it seems natural to require that some combinations of the first and higher order derivatives of the input signal should comprise an extremum. Now, according to the sampling theorem, the least rate of discrete sampling points which determine a band-limited signal (limited to bandwidth W) must occur at the Nyquist rate ($2W$). These samples are enough to determine also derivatives of any order. If $u(t)$ is the continuous band-limited input signal and is sampled at Nyquist rate, which yields u_i ($i = \cdots -2, -1, 0, 1, 2, \cdots$) samples, the samples u_i' of the derivative signal $u'(t)$ are given by the following linear transformation:

$$\mathbf{u}' = \mathbf{A}\mathbf{u}, \tag{1}$$

where

$$\mathbf{u} = (u_1, u_2 \cdots u_n); \qquad \mathbf{u}' = (u_1', u_2' \cdots u_n')$$

and the elements of the transformation matrix $\mathbf{A}$ are

$$A_{mm} = 0; \qquad A_{mn} = 2W \frac{(-1)^{m-n}}{m - n}.$$

For the processing on digital computers we get the input data in sampled and quantized form. As we see from (1), to compute only a first deriva-

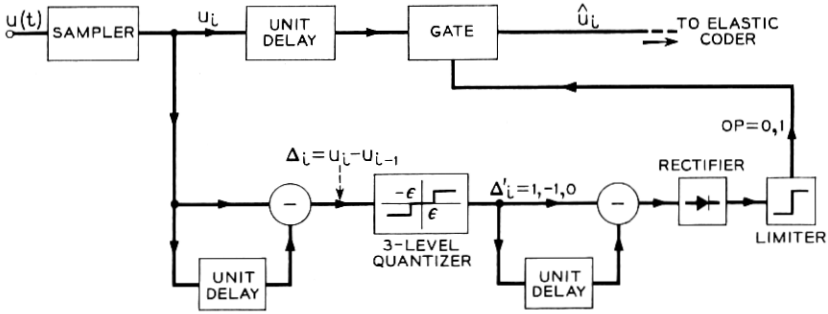

Fig. 1 — Actual system used to transmit only certain points of television signal.

tive would require a vast amount of computations, taking into account all sample values of the input signal. We can reduce the number of operations to a few subtractions if we introduce in place of the derivatives differences between sequential samples. If we now define an edge in terms of differences, we get a new system which resembles the previous one superficially, but in the microstructure (i.e., in the determination of a point within one Nyquist interval) the systems may differ considerably. We must not forget that, on account of human vernier acuity, ambiguities within one Nyquist interval $1/(2W)$ may be clearly visible. Because we cannot decide by mere speculation which of these systems will prove to be superior, we investigate the simpler one first.

III. DEFINITION OF AN EDGE POINT

The actual system is given in Fig. 1. The band-limited video signal $u(t)$ is sampled at Nyquist rate, and the difference ($\Delta_i = u_i - u_{i-1}$) between sequential samples is computed. A three-level quantizer with decision levels ϵ and $-\epsilon$, and with representative levels 1, -1 and 0 performs the following quantization:

$$\begin{aligned} \Delta_i' &= 1, && \text{if } \Delta_i \geqq \epsilon, \\ \Delta_i' &= 0, && \text{if } |\Delta_i| < \epsilon, \\ \Delta_i' &= -1, && \text{if } \Delta_i \leqq -\epsilon. \end{aligned} \tag{2}$$

Here the ϵ decision level has to be set experimentally. If it is too small, the operation will be affected by trivially fine structure; if it is too large, the fine details in the picture will be lost.

Now we define a sample point as an edge point ($\hat{u}_i$) if the quantized left- and right-hand differences (Δ_{i-1}' and Δ_i') of that point belong to one of the six cases, given by (3) and shown in Fig. 2(a):

$$\begin{aligned} &\Delta_{i-1}'\Delta_i' < 0, \\ &\Delta_{i-1}' = 0 \text{ and } \Delta_i' \neq 0, \\ &\Delta_{i-1}' \neq 0 \text{ and } \Delta_i' = 0; \end{aligned} \tag{3}$$

that is (in a more efficient notation),

$$\Delta_{i-1}' \neq \Delta_i'.$$

These cases refer to the local maxima or minima of the differences and to the end points of horizontal sections, provided the changes are above the ϵ threshold. Sample points on monotonic increasing, decreasing or horizontal sections will be omitted. These nontransmitted samples thus fall in the next three cases, shown in Fig. 2(b).

To select (from the nine possible cases) the six cases which correspond to an edge point, we have to perform the operations indicated in Fig. 1.

The output of the quantizer is again delayed one sample period and subtracted from its undelayed form to obtain the difference of the quantized left- and right-hand differences. After the second subtraction we get $OP = \Delta_i' - \Delta_{i-1}'$, which is nonzero for samples for which we want to define an edge and is zero otherwise.

The OP signal after full-wave rectification and limitation operates as a gating pulse and specifies which samples have to be transmitted.

As the result of these operations we get samples at an irregular rate, the average of which is substantially less than the Nyquist rate. This average rate depends on the picture material and on the ϵ threshold. Because of the irregular occurrences of the selected samples we also must

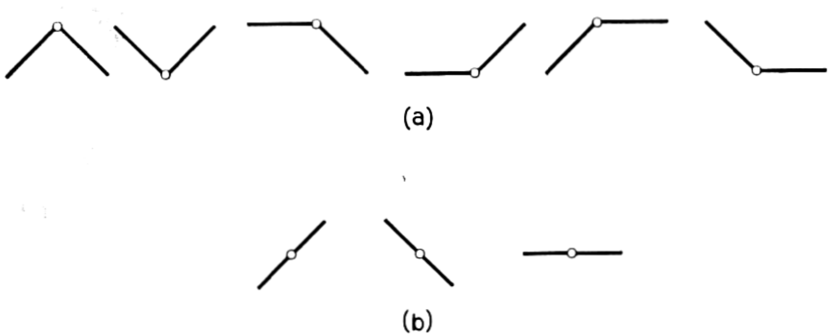

Fig. 2 — (a) Six cases in which quantized left- and right-hand differences are not equal; (b) monotonic increasing, decreasing and horizontal sections.

Fig. 3 — Picture material before and after processing (finer threshold setting).

specify their positions in time. That requires additional information to be transmitted; thus, the saving in required channel capacity is less than the ratio between the Nyquist rate and the average rate of the edge points. The net saving depends considerably on the coding schemes we apply to transmit the values and locations of the chosen samples, as will be discussed later.

IV. INTERPOLATION

After we specify the criteria for selecting the samples, we have to decide on an appropriate interpolation function. Because the selected samples occur less frequently than at Nyquist rate, they can be considered independently of each other. This means that there is no preferred curve connecting the selected samples from a mathematical point of view. From a psychological standpoint, the eye is not sensitive to the exact shape of a short transient, and thus the simplest choice in that region is a linear interpolation function. Furthermore, the longer monotonic increasing or decreasing sections between two edge points can be convex or concave and, in the average case, the best interpolation is again the linear one.

V. COMPUTER SIMULATION

According to the above considerations a program was written to determine the edge points by using the criteria given in (3) and to interpolate straight lines between them.[8] The program also provided the statistics of the distribution of the distances between adjacent edge points. The time fluctuation of the selected sample rate also was recorded.

The picture material before processing but quantized in time and amplitude is shown in Fig. 3 (middle column). The picture consists of 100 lines, each containing 120 picture elements. For synchronization and blanking we used 15 picture elements in every line and the complete first line; thus the number of picture points is $99 \times 105 = 10{,}395$. This resolution corresponds to a television picture $\frac{1}{25}$ the area of the present standards. That means that the given pictures have to be observed from five times greater distance to get the usual resolution. If we take four times picture height as the usual viewing distance for standard television, the presented pictures have to be judged from a distance of 20 times picture height. The reason for the choice of this coarser resolution was a compromise between acceptable picture quality and computer storage capacity. The amplitudes were quantized into 10 bits (1024 levels) between the white and blacker-than-black levels, and into 9 bits

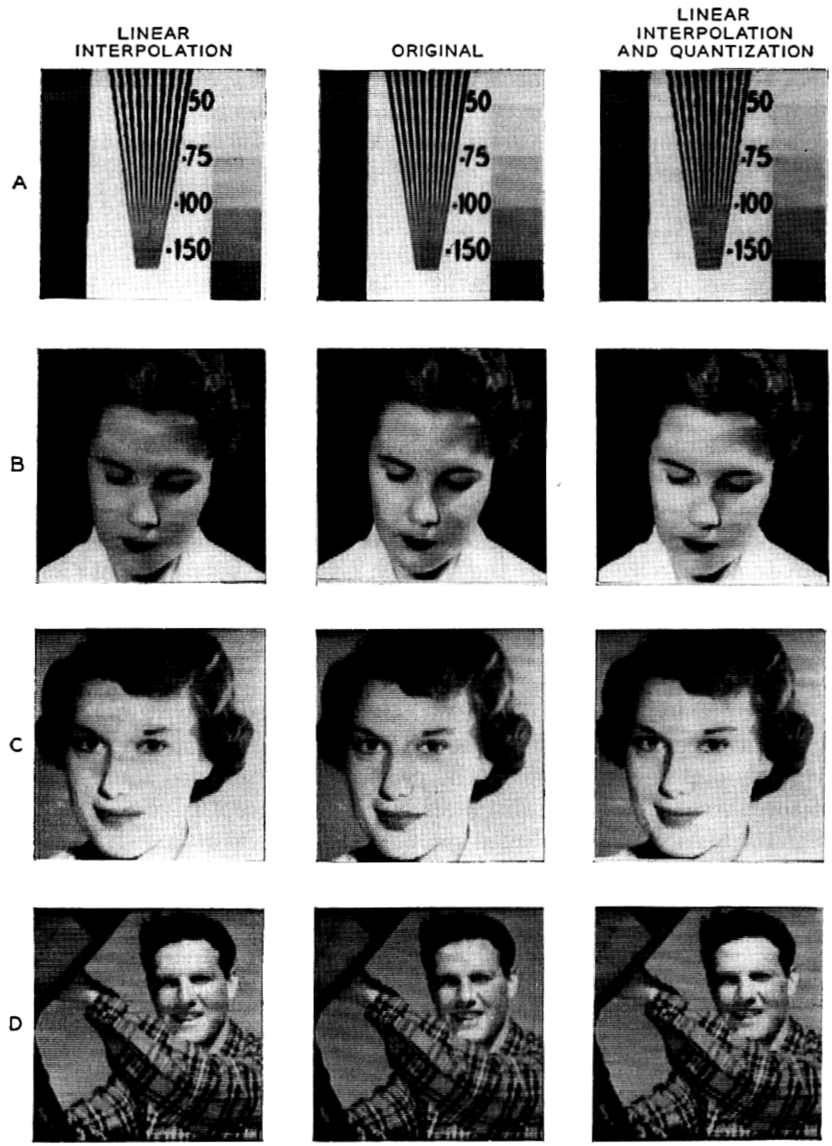

Fig. 4 — Picture material before and after processing (coarser threshold setting).

(512 levels) between the white and black levels, although 7 bits are enough for excellent quality.[1] The sampling and quantization were performed by an analog-to-digital converter, which could perform the opposite operations as well. A slow-speed scanning system converted the pictures into electrical signals and back to pictures. The sampled and quantized signals were put on a tape which served as an input to the IBM 704. The processed output of the computer was written on tape too, and the same devices in reversed operation converted it into pictures. The pictures which are designated as "original" (Figs. 3 and 4, middle columns) went through all these devices, but the program of the computer was such that it copied the input tape unchanged onto an output tape.

After we tried the processing with several ϵ threshold values we got the surprising result that, although the number of selected samples increased with decreasing ϵ values, the over-all appearance became worse. The most apparent defects were at vertical edges. The explanation of this effect is as follows: With decreasing ϵ thresholds the positioning of an edge point at the endings of horizontal sections becomes very sensitive. A little change in slope can shift the edge points several Nyquist intervals apart (see points ϵ_1 in Fig. 5). At a vertical edge each slope of the transients differs slightly from the one in the lines above (a small amount of added noise has the same effect), giving a very annoy-

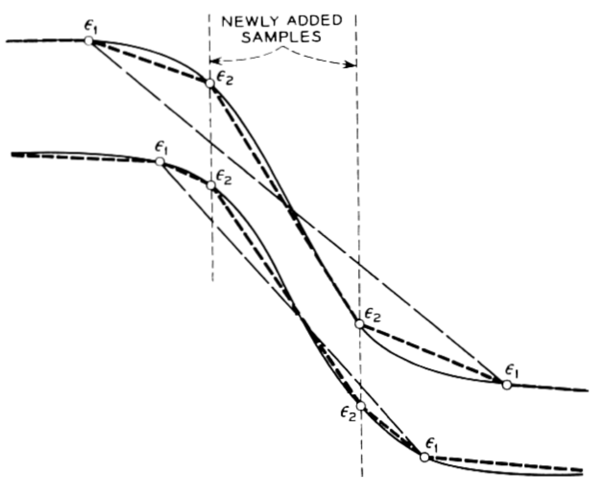

Fig. 5 — Slight change in slope (as in upper curve) moves edge points ϵ_1 several Nyquist intervals apart.

Fig 6 — (left) Distorted pictures show that systems which perform well for acoustics may not work for vision; (right) output of system with only one fine threshold.

ing fuzziness to the vertical edge. The right-hand pictures of Fig. 6 show the output of this system with a fine threshold setting ($\epsilon = 3.6$ per cent) and the above-mentioned defects are clearly visible. If we increase the ϵ threshold, this sensitiveness to edge positioning decreases, but the quality of the pictures also decreases. The reason for this is that, by taking increased threshold settings, we get fewer selected samples, and thus fine details in the pictures will be lost. With small threshold values, we get phase errors in edge positioning. Therefore, ϵ can be neither too small nor too large, and even the best compromise does not ensure adequate picture quality.

VI. OBJECTIONABLE SENSITIVITY TO EDGE POSITIONING AND ITS CORRECTION

There is a way to get rid of this annoying fuzziness and still be able to choose a value of ϵ that is small enough. If we take two threshold values (ϵ_1, ϵ_2) such that $\epsilon_1 \ll \epsilon_2$, we get two sets of edge points. We then take the union of these two sets. In most cases the set of edge points determined by the finer threshold contains the set determined by the coarser threshold, and thus does not increase the number of selected points. In the few cases when that is not the case, the additional points help to cure the sensitivity to small slope changes. In Fig. 5 we see that the edge points determined by ϵ_2 remain fixed in subsequent lines, and the phasing errors due to the edge points given by ϵ_1 have no effect on the over-all interpolation.

We determined the number of selected samples for this system using the picture material shown in the middle column of Fig. 3. We chose $\epsilon_2 = 10$ per cent (of the peak-to-peak value between black and white) and $\epsilon_1 = 3.6$ or 5 per cent. The increase of selected samples due to the additional samples determined by ϵ_2 depended on the picture material and was small (less than 11 per cent for pictures A, B, C and 17 per cent for picture D). The ratio of the selected samples to the total number of samples is given in Table I in per cent.

To simplify the design of coding devices, we limited the maximal distance to 16 Nyquist intervals. If, after determining an edge point and scanning further from left to right 16 Nyquist intervals, we did not find a next edge point, we selected a new sample 17 Nyquist intervals away from the previously selected sample. The frequency of occurrence of such a case is very small; thus, the increase due to these newly selected points is negligible.

The foregoing process gives good results in nearly every case. In exceptional cases, the pictures leave something to be desired. The reason for this and its correction are discussed next.

VII. THE "TUNNEL EFFECT" AND ITS CORRECTION

The system discussed above selects the edge points by analyzing the quantized differences according to (3). If the difference between subsequent samples is less than the ϵ_1 threshold, we do not transmit any sample. Now the pictures may contain hill- or valley-like sections with slopes so mild that the left- and right-hand differences around the maximum or minimum are less than ϵ_1, and thus we do not select these maximum or minimum points for transmission. The linear interpolation between the subsequent edge points looks like a tunnel, and if $t_j - t_i$

TABLE I — RATIO OF SELECTED SAMPLES TO TOTAL (PER CENT)

Scene	System Setting	
	$\epsilon_1 = 3.6$ per cent; $\epsilon_2 = 10$ per cent	$\epsilon_1 = 5$ per cent; $\epsilon_2 = 10$ per cent
A	32.9	29.1
B	30.1	24.0
C	34.9	28.3
D	47.0	42.0

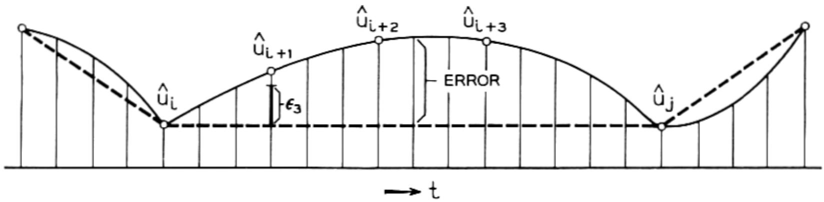

Fig. 7 — The tunnel effect between edge points $\hat{u}_j$ and $\hat{u}_i$ and its correction with ϵ_3 and $\tau = 3$.

(the distance between edge points $\hat{u}_j$ and $\hat{u}_i$) is long enough, large errors can be committed (see Fig. 7). It is possible to correct this effect in many ways. We used the following procedure: We subtracted the original picture from the processed one to give the interpolation errors. A new threshold value ϵ_3 was chosen. Whenever the error exceeded this threshold at time, t_k, the routine searched for the closest edge points left and right. If the distance, τ, on both sides was equal to or more than three Nyquist intervals, the routine selected the sample, u_k, for transmission; if the distance was less, no additional samples were selected. Thus we left the errors uncorrected for short sections (less than six Nyquist intervals long), utilizing the same psychological effect; i.e., the eye is not sensitive to the exact value of brightness changes in short times. This last manipulation improved the picture quality further. The number of selected points in this system is given in Table II. The distribution of the distance between the edge points for scene B is given in Fig. 8. Here, P_i is the frequency of the distances between subsequent edge points, and the index refers to the distances in Nyquist intervals.

By comparing Table I with Table II we see that the tunnel effect occurs very seldom, and that the increase in transmitted samples is negligible. The pictures obtained by this variant are shown in the left columns of Figs. 3 and 4. Fig. 3 corresponds to the finer threshold setting ($\epsilon_1 = 3.6$ per cent, $\epsilon_2 = 10$ per cent, $\epsilon_3 = 5$ per cent, $\tau = 3$ Nyquist

Table II — Ratio of Selected Samples to Total (Per Cent)

Scene	System Setting	
	$\epsilon_1 = 3.6$ per cent; $\epsilon_2 = 10$ per cent; $\epsilon_3 = 5$ per cent; $\tau = 3$	$\epsilon_1 = 5$ per cent; $\epsilon_2 = 10$ per cent; $\epsilon_3 = 7.2$ per cent; $\tau = 3$
A	32.9	29.2
B	31.3	25.3
C	35.8	29.3
D	47.3	42.4

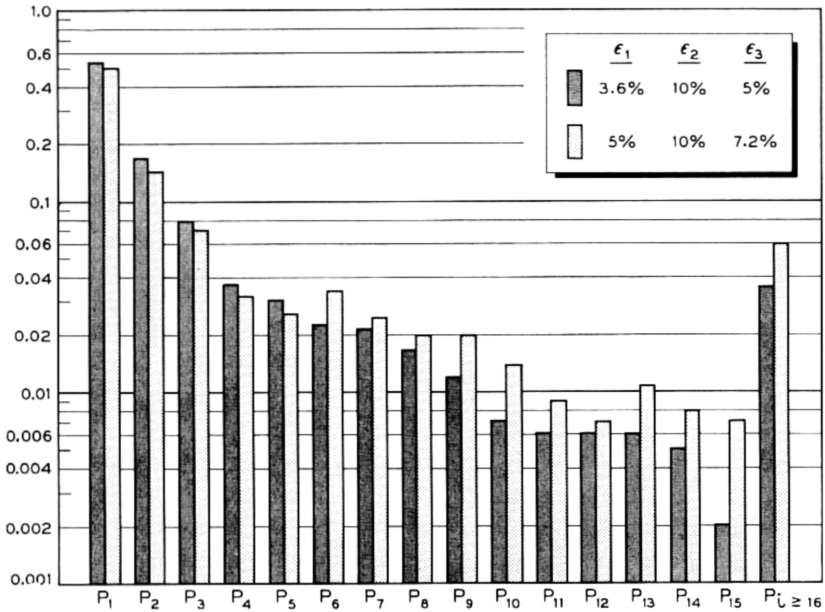

Fig. 8 — Distribution of the distance between subsequent edge points in scene B.

intervals); Fig. 4 to the coarser setting ($\epsilon_1 = 5$ per cent, $\epsilon_2 = 10$ per cent, $\epsilon_3 = 7.2$ per cent, $\tau = 3$ Nyquist intervals).

As we see, minor modifications in the program parameters improve the appearance of the pictures considerably. The statistics of the selected points for scene B are given in Fig. 8. Another advantage of choosing linear interpolation becomes apparent. As we add new points to the original edge points according to some different criterion, we need not label them separately because, in the case of linear interpolation, every received sample can be treated equally.

VIII. EVALUATION OF PROCESSED PICTURES

In Figs. 3 and 4 the left columns show the pictures processed according to the last variant. This variant, which includes edge determination by fine and coarse thresholds and tunnel-effect correction, we shall refer to simply as "linear interpolation." As we see, small changes in the threshold greatly affect the number of selected samples and the picture quality. If we decrease the thresholds, the number of selected samples reaches an asymptotic value which, depending on the picture material, is considerably higher than those obtained by $\epsilon_1 = 3.6$ per cent. (For example, in scene B the asymptotic ratio of selected samples to the total is 49.3 per cent.) Nevertheless, the improvement of the processed picture is very slight if $\epsilon_1 < 3.6$ per cent. So, the setting $\epsilon_1 = 3.6$ per cent, $\epsilon_2 = 10$ per cent, $\epsilon_3 = 5$ per cent, $\tau = 3$ Nyquist intervals presents a good compromise between picture quality and information savings.

The pictures taken with settings $\epsilon_1 = 5$ per cent, $\epsilon_2 = 10$ per cent, $\epsilon_3 = 7.2$ per cent, $\tau = 3$ Nyquist intervals could be taken as another compromise, with an emphasis on economy rather than on quality.

The reason that the picture quality does not improve much with decreasing thresholds can be explained simply: The sensitivity of the eye to phase errors in locating the edge points within one Nyquist interval is the major cause of the deteriorated appearance of the processed pictures. This ambiguity within one Nyquist interval will not improve much as we set the thresholds finer. One way to get better results would be to specify the location of the selected edge points more accurately than one Nyquist interval. Even a modest oversampling of a factor of two would be advantageous. In the next section it will be apparent that this operation will not increase the required channel capacity by more than 12 per cent in the worst case (scene D) but would be beneficial in locating the edge points within one-half Nyquist interval.

The above-mentioned phase errors are most disturbing on the vertical edges of scene A and on the outline of the face in scene B. For more detailed material this effect is much less objectionable.

IX. COARSE QUANTIZATION OF FAST-TRANSIENT REGIONS

Some recent work has exploited the same psychological phenomenon (that is, the insensitivity of the eye to the amplitude and shape of sudden brightness changes) from a different approach.[9,10,11] These authors quantized the amplitude of the samples in the region of fast transients into fewer levels than for the rest of the picture. We can add this feature advantageously to the linear interpolation between the edges. The obtained benefits are complementary: For pictures with many fast transients the number of selected samples is large, but these are just the samples which can be quantized in fewer number of bits. On the contrary, for pictures with fewer details (thus with fewer fast transients) we have to specify the selected samples more accurately, at least by 7-bit quantization.

We incorporated this feature in the linear interpolation system in the following way: The selected samples were divided in two categories. The first category contained those edge points which were no more than

TABLE III — RATIO OF COARSELY QUANTIZED SAMPLES TO SELECTED SAMPLES

Scene	System Setting	
	ϵ_1 = 3.6 per cent; ϵ_2 = 10 per cent; ϵ_3 = 5 per cent; $\tau = 3$; $t = 2$	ϵ_1 = 5 per cent; ϵ_2 = 10 per cent; ϵ_3 = 7.2 per cent; $\tau = 3$; $t = 2$
A	0.71	0.63
B	0.52	0.44
C	0.59	0.54
D	0.77	0.72

two Nyquist intervals from the left and the right neighboring edge points. These points thus belonged to high-frequency regions and were quantized coarsely. In the experiments, 3- and 4-bit (8- and 16-level) quantization was tried out. The remaining edge points which were the end points or inner points of low-frequency regions had to be quantized into finer steps. We used here 9 bits (512 levels), as in the linear interpolation system, but 7 bits would probably be very satisfactory. A 3-bit quantization for the fast-transient region turned out to be very noticeable, but 4-bit quantization gives quite satisfactory results, as the right columns of Fig. 3 and 4 show. The ratio of the coarsely quantized samples to the selected samples is given in Table III. Here t is the parameter which defines the fast-transient regions and, as mentioned, was set for two Nyquist intervals. According to this setting, edge points falling in regions which contained oscillation higher than half of the maximum frequency of the signal were coarsely quantized. We might have increased t even further from a psychological point of view, but the additional reduction in channel capacity would have been slight. In the following section we evaluate the obtained statistics and give the channel-capacity figures for possible coding schemes.

X. CODING AND AMOUNT OF CHANNEL-CAPACITY SAVINGS

After the processing of the pictures, the second step is a subjective evaluation of them. Provided we accept the obtained picture quality, the next step is to evaluate the information content and the obtained channel-capacity savings. Information theory enables us to get a theoretical lower bound of the information content of the processed pictures, but to realize it even approximately requires very involved coders an decoders. Therefore, we also make computations with simpler coding devices. Such devices do not make use of the obtained statistics of distances between edge points, but regard all possible distances as equally probable. Because the greatest distance between selected samples is restricted to 16 Nyquist intervals, 4 bits are required to specify the location of a selected sample from its previous neighbors. To describe the amplitude of the selected sample, 7 bits are adequate. Thus, 11 bits are required to specify the location and amplitude of a selected sample point. In conventional systems, 7 bits are enough to specify the amplitude of samples occurring at Nyquist rate.

Aside from the foregoing, the saving in the transmitted information obviously would be the ratio of the selected samples to the total number of samples. Because of the foregoing, the saving is diminished in the ratio of 11/7. If N is the total number of samples and N' is the number of selected samples, then the average rate of information is

$$R = 11 \frac{N'}{N} \text{ bits/sample.} \tag{4}$$

The coder contains a time-variable buffer storage to smooth out the incoming signals, which arrive at an irregular rate, and to transmit them on the channel at a constant average rate. At the receiver, the inverse elastic operation is performed in the decoder. If $N'/N < 7/11$, we get a saving in information rate over the conventional Nyquist rate sampling.

The rate of information is computed for the linear interpolation system with and without quantization. In the quantized case the required rate is

$$R_q = 11 \frac{N' - N^*}{N} + 8 \frac{N^*}{N} \text{ bits/sample,} \tag{5}$$

where N^* is the number of coarsely quantized samples.

If we take advantage of the highly peaked distribution curve of the distance between selected samples, and use a Shannon-Fano code to encode them, the rate of information for the linear interpolation system without and with quantization is as follows:

$$R_m = \frac{7 + H_X}{d}, \tag{6}$$

$$R_{mq} = \frac{4 + H_X}{d} k + \frac{7 + H_X}{d} (1 - k), \tag{7}$$

where

$$d = \frac{N}{N'}, \qquad k = \frac{N^*}{N}$$

TABLE IV — INFORMATION RATE FOR PROCESSED PICTURES (BITS PER SAMPLE)

Scene	System Setting							
	ϵ_1 = 3.6 per cent; ϵ_2 = 10 per cent; ϵ_3 = 5 per cent; τ = 3; l = 2				ϵ_1 = 5 per cent; ϵ_2 = 10 per cent; ϵ_3 = 7.2 per cent; τ = 3; l = 2			
	R	R_q	R_m	R_{mq}	R	R_q	R_m	R_{mq}
A	3.62	2.92	2.92	2.22	3.21	2.66	2.65	2.01
B	3.44	2.96	2.94	2.45	2.78	2.45	2.43	2.34
C	3.94	3.31	3.27	2.64	3.23	2.75	2.76	2.28
D	5.20	4.10	4.08	2.99	4.66	3.76	3.72	2.80

and

$$H_X = -\sum_{i=1}^{16} P_{Xi} \log_2 P_{Xi} .$$

The P_{Xi} are the frequencies of the distances between extremals of a given picture, X; R, R_q, R_m and R_{mq} are tabulated for different scenes and system settings in Table IV. The obtained information reduction is considerable, and it is an advantageous situation that the smallest entropy values, H_X, are obtained for the most involved pictures which require the most selected samples.

If we use statistical coding (e.g., Shannon-Fano or Huffman codes), we have to use the same code for all scenes. If we choose the code according to a scene Y, and we have to encode a different scene X, the expected code length in binary digits will be approximately

$$L_{XY} = -\sum_{i=1}^{16} P_{Xi} \log_2 P_{Yi} , \tag{8}$$

where L_{XY} is always greater than $L_{YY} = H_Y$. To see how these values compare with the entropies, we computed them for the 16 possible combinations of scenes using the finer threshold settings. Table V shows L_{XY}, which is not very sensitive to Y (i.e., to the particular code used).

TABLE V — EXPECTED CODE LENGTH L_{XY} IN BITS

Scene		Y: A	Y: B	Y: C	Y: D
X	A	1.88	1.99	1.95	1.99
	B	2.50	2.37	2.41	2.51
	C	2.20	2.16	2.13	2.20
	D	1.71	1.75	1.70	1.65

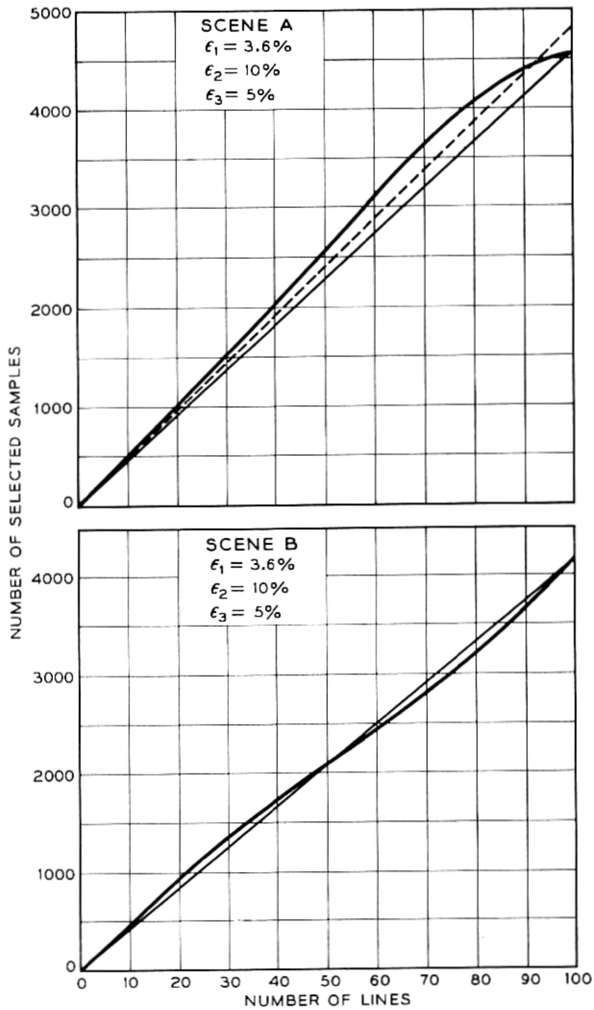

Fig. 9 — Fluctuation in time of number of selected samples at input of the buffer storage.

The channel capacity has to be enough to transmit all possible picture material. Because we do not know whether the selected few pictures are good representatives of all possible entertainment pictures, we cannot state theoretically anything definite about channel capacity, but we hope that the results are close to reality. If we regard the pictures as a whole, instead of as a 25th portion, and look at them from the usual four times picture height instead of from the distance of 20 times picture height, we can get an impression of how the quality would look for the most crowded scenes.

The picture materials used were not far from the stationary case; i.e., entropy values calculated from statistics gathered across different lines of a picture did not fluctuate much.

XI. BUFFER-STORAGE REQUIREMENTS

In Fig. 9 we show how the number of selected samples fluctuates in time at the input of the buffer storage (curved lines). If we read out the data at a constant rate, we get a straight-line representation as a function of time. If we choose this constant rate at the output as the average rate of the input, the straight line starts at the origin and hits the input curve at the end point. The maximal difference between the input and output curves gives a good estimate of the buffer-capacity requirements. The curves for scenes A and B are shown for the fine setting of the linear interpolation system without quantization or statistical coding. The coordinates are equivalent to time and are specified in terms of the number of scanned lines. The abscissae are the number of selected samples at the input and output of the quantizer, with the synchronization signals added. The requirement in storage capacity is about one scanning line (120 samples) for scene B and about four scanning lines for scene A. If an increased output rate (dashed straight line) is used, the storage-capacity requirement can be reduced.

XII. SUMMARY

The above-described experiments used the inability of the eye to notice the exact amplitude and shape of short brightness transients. By using straight-line interpolation between edge points and coarse quantization of edge points in fast-transient regions, we can transmit information at a rate of 3 bits per sample or less for the given scenes and shown picture quality. If we take the present 7 bits per sample rate as a reference, the greatest possible saving for scene D is $7/2.99 = 2.3$ times, and for scene B it is 2.9 times. Naturally, with practical buffer-storage size we cannot average out the differences in information rate for the different scenes, and we have to match the channel to the worst case.

If we use additional information to specify the location of edge points within a Nyquist interval, the quality of the pictures will greatly improve.

The obtained savings are modest and close to the figures achieved by other authors. Probably the results are interesting more because of what they reveal of visual perception than because of their immediate engineering applicability.

XIII. ACKNOWLEDGMENTS

The author is very much indebted to R. E. Graham and J. L. Kelly, Jr. for many helpful suggestions.

REFERENCES

1. Goodall, W. M., Television by Pulse Code Modulation, B.S.T.J., **30,** January 1951, p. 33.
2. Harrison, C. W., Experiments with Linear Prediction in Television, B.S.T.J., **31,** July 1952, p. 754.
3. Oliver, B. M., Efficient Coding, B.S.T.J., **31,** July 1952, p. 724.
4. David, E. E., Jr., Mathews, M. V. and McDonald, H. S., Experiments with Speech Using Digital Computer Simulation, I.R.E.-Wescon Conv. Rec. (Audio), August 1958, p. 3.
5. Graham, R. E. and Kelly, J. L., Jr., A Computer Simulation Chain for Research on Picture Coding, I.R.E.-Wescon Conv. Rec. (Computer Applications), August 1958, p. 41.
6. Mathews, M. V., unpublished manuscript.
7. Youngblood, W. A., Picture Processing, Quart. Prog. Rep., M.I.T. Res. Lab. of Elect., January 1958, p. 95.
8. Julesz, B., unpublished manuscripts.
9. Graham, R. E., Predictive Quantizing of Television Signals, I.R.E.-Wescon Conv. Rec. (Electronic Computers), August 1958, p. 147.
10. Kretzmer, E. R., Reduced Alphabet Representation of Television Signals, I.R.E. Conv. Rec., Part IV, 1956, p. 140.
11. Schreiber, W. F. and Knapp, C. F., TV Bandwidth Reduction by Digital Coding, I.R.E. Conv. Rec. (Information Theory), 1958, p. 88.