THE BELL SYSTEM TECHNICAL JOURNAL
Vol. 51, No. 1, January, 1972
Printed in U.S.A.

Buffering of Data Generated by the Coding of Moving Images

By J. O. LIMB

(Manuscript received August 20, 1971)

Data are generated at a fairly uneven rate when video telephone signals are coded by transmitting only the parts of the picture that change from frame to frame. Complete smoothing of these data is impracticable because of the size of the required buffer. Obviously, even a small buffer provides some advantage however. The object of this paper is to explore the relation between buffer requirements and channel rate under varying experimental conditions.

The study was made by recording three minutes of data (covering a range of action) on a digital computer and simulating buffer behavior for various channel rates and operating conditions.

With little or no buffering, a large rate is necessary. As the size of the buffer is increased, the required channel rate typically decreases quite rapidly until the buffer is large enough to smooth the data over an entire field. Beyond this point there is relatively little improvement until the buffer is large enough to smooth the data generated by a moving user from one movement to the next.

At times data are generated at a faster rate than can be handled by the buffer-channel combination. Reduction of the rate of data generation during these periods can be controlled either by using the amount of activity in the picture as a control or by using the state of the buffer as a control. Both methods have distinctly different effects on the buffer-size versus channel-rate curve. The two modes of control can be effectively combined in developing a successful control strategy.

The buffer size required to achieve within-field smoothing can be reduced dramatically if the data within the field are not taken in the order in which these data are generated but instead are interleaved in a systematic manner; this is because of the nonuniform rate of data generation within a field.

I. INTRODUCTION

Although most coding schemes for pictures have considered only a single frame at a time, it has long been realized that if only the changes between frames were transmitted the average bit rate could be reduced significantly.[1,2] For *Picturephone*® service this is particularly true because most of the time the person using a *Picturephone* set sits quite still or only moves his mouth. However, extreme movement can occur when a camera is panned, when a user leans forward to adjust the controls, or when he stands up and moves out of the field of the camera. In such instances the amount of movement can exceed even that which would be expected in movies or broadcast television. But even if such extremes were unimportant, it would still be necessary to buffer the data generated during normal human movement in order to take advantage of the large reduction in average bit rate which occurs when one transmits only those parts of the picture which change.

Movements exceeding one second in duration are probably not unusual during videotelephone use. To smooth the peaks in data generated during such movement would require a large buffer, capable of storing a significant fraction of the data generated during the movement. Even if buffers were cost-free items, it would probably not be feasible to smooth, completely, the flood of data generated during movement because of the delay introduced into the signal path by the buffer. (The maximum round-trip delay that can be tolerated in a conversation is between one-half second and one second.[3])

J. C. Candy, et al., describe experiments with a frame-to-frame buffered coder operating with feedback control to reduce the accuracy (both spatial and amplitude) with which the picture elements are reproduced.[4] They use a 67,000-bit buffer and assume a 2-megabit/second channel. The algorithm for selecting those points to be transmitted differs from the one used in this study and will affect the results somewhat.

The object of this study is to explore the relation between channel-rate and buffer requirements and to determine how much smoothing the buffer can achieve. The effects of two different buffer control techniques for reducing the spatial resolution in moving areas are compared and the saving in buffer capacity, using a technique of interleaving data (which in effect uses the frame memory of a coder to obtain some smoothing), is investigated.

II. DESIGN OF EXPERIMENT

There is currently no adequate statistical model to describe the movements of videotelephone users. If there were, transforming this to a data rate and then obtaining overflow statistics for the buffer would still be a difficult task. Therefore we studied buffering by simulation.

A repeatable source of data is a necessity if required buffer sizes are to be computed and compared for many channel rates and under different conditions. This precludes the use of a live subject for each experiment, since variations between runs could mask the effect being studied. We would like to have recorded several minutes of a typical videotelephone signal in binary form but this would have required much more storage than was available. Instead the signal was recorded in a reduced data form which could still yield the necessary information for the study. This is described more fully below.

The picture was divided into moving and stationary areas using a moving area detector.[5] A typical segmentation is shown in Fig. 1. For each horizontal line in the picture two measurements were made to represent quantitatively the moving (or changing) area. First, the elements falling in the moving area were counted; second, the number of moving area segments was counted. These two figures were each represented as 8-bit numbers, packed into one word, transmitted to the computer, and stored on a digital disk pack. Knowing these two numbers, we can calculate the number of bits generated per line by many different coding schemes. Buffer simuations can then be performed by incrementing a counter (each line) with the data generated during the line, and then decrementing the counter by the amount the channel transmits during the line.

In calculating the buffer requirements for a particular channel rate the only important variable is the ratio between the number of bits used to transmit information about the address of each segment and the number of bits used to transmit the amplitude of each pel (picture element) within the segment. For example, if 16 bits are used to denote each segment and 4 bits are used to denote the amplitude of each pel within a segment, the ratio is 4. If we have the buffer-size versus channel-rate curve for this allocation of bits and want to find the curve for a coding scheme which allocates 32 bits to each segment and 8 bits to each pel within the segment, we need only halve the scales of the axes. Most results have been obtained assuming 16 bits per segment and 4 bits per element. The 16 bits allocated to each segment can be distributed in many ways. For example, 8 bits can be used for a starting address, 4 bits for an error-correcting code, and 4 bits for a terminating code. Or, the 16 bits can be used just for a starting address and a stopping address. The 4-bit amplitude code can be used for an element difference code or a frame difference code. Of course, the actual allocation

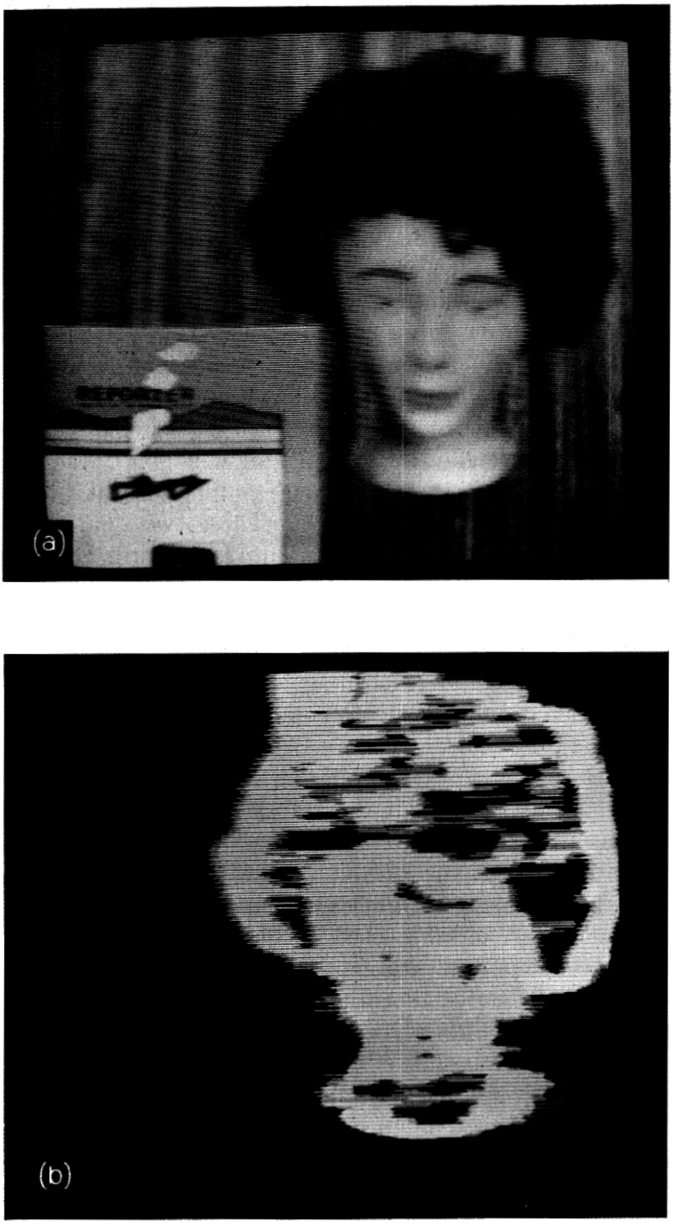

Fig. 1—(a) Picture showing moving head against stationary background; (b) Flags showing area deemed moving.

of bits is rather immaterial as far as a buffer study is concerned. Bookkeeping bits which occur every line, such as synchronization words and some forms of forced replenishment, were not explicitly incorporated in the simulations. These can be simply incorporated by modifying the channel rate to take these factors into account. Incrementing the buffer at a line rate rather than at an element rate makes little difference when we are interested in studying the smoothing that takes place over a frame and from frame-to-frame rather than over a line.

A rather basic assumption in comparing the buffer requirements of various schemes is that the operation of the movement detector does not change with the type of processing that is done (e.g., subsampling). This is almost true of the 2-dimensional movement detector that was used, and further sophistication in the design of movement detectors should make them even less sensitive to the type of coding employed. In any event the required buffer size is not strongly dependent on the segmentation criterion used because during periods of rapid movement when the buffer is most likely to be full it is quite easy to pick out the moving area.

2.1. *Recording of Data*

An RCA vidicon camera was used having a signal format roughly equal to that used in *Picturephone* service: 248 elements and 271 lines with 24 blanked elements in the line and 16 blanked lines in the frame; a 30-Hz frame rate with a 2-to-1 interlace. The signal was fed to a 7-bit quantizer via a 1-MHz low-pass filter. The frame difference signal was formed by subtracting the frame delayed signal from the undelayed signal. The difference signal was then fed to the movement detector which decided whether a point belonged to either the moving area or the stationary area. In order to make a decision the movement detector examined the frame-difference threshold signal of a block of 24 samples: 8 from the line being coded, 8 from the line above, and 8 from the line below. The frame-difference threshold signal is unity for those samples whose magnitude, as compared with the previous frame, changed by more than a threshold value; otherwise it is zero. In our experiment the threshold was set at 3/128ths of the peak signal value. The movement detector will change from the stationary to the moving mode when 8 or more of the 24 samples currently being evaluated have a frame-difference threshold signal of value 1. The movement detector will change back to the stationary mode when two or less of the 24 samples have a frame-difference threshold signal of value 1. Thus the decision is hysteretic, and, consequently, the movement detector yields a reasonably contiguous moving area (see Fig. 1). If a stored signal is updated in only the segmented area then the resulting picture is of reasonable quality.

Just over three minutes of data were recorded. During the first minute (inactive data sample) the subject was asked to talk naturally but quietly, keeping his head and body relatively still. During the second minute (medium activity data sample) the activity was increased with the subject making occasional hand movements and more pronounced head movements. The subject was most active during the third minute (active data sample) making exaggerated head movements and hand movements but no large body movements such as standing or moving out of the field of view. Some characteristics of the data are given in Table I.

The statistics shown in Table I are (*i*) the mean value, (*ii*) the value which was exceeded only 1 percent of the time, and (*iii*) the maximum value. The mean number of segments in a line increased from less than 1 to 1.5 as the activity increased from inactive to active. The number of pels in the segmented area increased proportionately more — from 18 to 75. The small increase in the average number of segments is probably caused by short segments combining to form a single longer segment as the amount of movement increases. The maximum value and the 99-percent value do not show such a marked increase with movement as does the mean. Notice, however, that the maximum value of the number of pels for the active data cannot increase further, since it is equal to the number of pels in a line.

For the coding techniques in which we were interested, the required buffer size was more dependent on that portion of the data used for describing amplitudes of pels than that portion used for positioning

Table I—Mean, 99% Value, and Maximum of the Number of Moving Area Segments and Pels in a Line.

	Inactive		Medium		Active	
	Seg.	Pel	Seg.	Pel	Seg.	Pel
Mean rate per line	0.7	17.6	1.3	44	1.5	75
99% value per line	3.0	100	4.0	122	5.0	183
Maximum rate per line	7.0	119			9.0	224

segments. This does not necessarily mean, however, that the segmenting criterion should be changed so as to decrease the number of pels at the expense of increasing the number of segments. Such a policy could lead to an increase in the overall data rate since the reduction in data rate due to the decrease in the number of pels to be transmitted could be more than offset by the increase in data rate due to the increased number of segments. The segmenting criterion which gives the minimum buffer requirement will depend on other factors such as whether or not subsampling[4] is used, what buffer control,[5] if any, is employed, and so on. Thus, at this early stage we have simply demanded that the segmenting give a reasonably contiguous coverage of the moving area, leaving the fine tuning until the basic structure of the coder is known.

III. RESULTS

Figure 2 shows a number of results relating required buffer size to channel rate. The channel rate is specified in bits-per-line. Thus a channel rate of 1 bit per pel corresponds to a channel rate of 224 bits

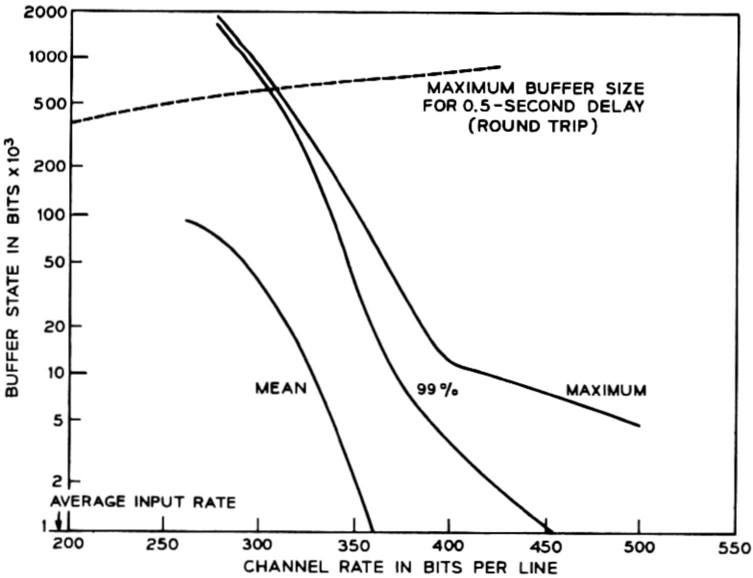

Fig. 2—Channel rate vs buffer state for medium activity data. Curves are for maximum buffer state, 99-percent point (1 percent overflow), and mean buffer state.

per line, excluding line address and line synchronizing bits. The simulations assume no data are generated or transmitted during the vertical blanking interval. This period can be used for the transmission of periodic data such as forced picture updating.[4]

The buffer size is specified in kilobits and plotted on a logarithmic scale to accommodate a large range of values. The term "state of the buffer" will be used to mean the fullness of the buffer at a particular instant. "Capacity of the buffer" will be used to denote the number of bits the buffer can store before overflow occurs. In this study the buffer was always large enough so that, for the channel rates used, the buffer never overflowed and hence may be regarded as infinite. Three statistics of buffer state were measured: (*i*) the mean, (*ii*) the buffer state which was exceeded only one percent of the time, and (*iii*) the maximum buffer state.

It is difficult to decide what measure of buffer state is appropriate in determining the capacity required for a buffer in a real system. The 99-percent point would probably result in too frequent buffer overflow although, of course, it depends critically on what actually happens to the received picture when the buffer does overflow. Note that using the 99-percent curve would result in less frequent overflow of a finite buffer than of an infinite buffer* since a finite buffer will cease to overflow as soon as the data generation rate falls below the channel rate. However, it may be a while before meaningful data can again be transmitted after overflow has occurred.

Conversely, the maximum buffer state is severe as a measure of practical buffer requirement because overflow would never occur. However, if the buffer in a working system were to overflow just once every minute (this curve was obtained for a 1-minute sample), then overflow would probably be too frequent. In the results that follow, the maximum buffer state is the measure that is most generally used as an indication of buffer capacity requirement.

The one-way signal delay through the send and receive buffer is assumed to be

$$\Delta = \text{buffer capacity/channel rate.} \tag{1}$$

Thus, if the round-trip delay, 2Δ, is to be kept less than 0.5 second then the send buffer and the receive buffer must each have a capacity

---

* Of course an infinite buffer, strictly speaking, will never overflow; however, we regard overflow as occurring whenever the simulated buffer size is exceeded. An infinite buffer continues to operate normally after the simulated buffer size is exceeded.

less than that given by the dotted limiting curve [derived from equation (1)] of Fig. 2.

The curves of Fig. 2 have two distinguishing features:

(*i*) A sharp change in slope at high channel rates.
(*ii*) A gradual turnover of the curves at lower channel rates.

This is most apparent in the maximum buffer-state curve. The change in slope, at the far right of each curve, is the point at which the buffer ceases to smooth over a whole field and we hypothesize that the upper change in slope is the point at which the buffer ceases to smooth significantly between one movement of the user and the next. The curves obtained from data interleaving (see later) provide evidence relating the "rightmost" change in slope (elbow) to the transition from intrafield to field-to-field smoothing.

Regarding the upper change in slope, we can calculate approximately the number of frames stored in the buffer during periods of peak data generation in the following manner: about 325,000 bits are generated per frame when the subject is active. If only 70,600 bits are transmitted per frame (275 bits per line) then the buffer has to hold about 255,000 bits for the frame, and if the buffer were full (1.4 million bits) then about 5-1/2 frames of active data would be stored in the buffer. The difference in data generation rate between two frames, five frames apart, is probably sufficient to explain the change in slope of the curves for large buffers.

From the elbow of the maximum buffer state curve to where the curve turns over at the top the slope is very steep. In reducing the channel rate from 400 bits per line to 300 bits per line the buffer size increases from 15,000 bits to 830,000 bits, an increase of 55 times. Thus, the elbow seems a suitable compromise for an operating point, enabling one to obtain the advantages of within-field smoothing for a reasonable buffer size. Of course, the actual operating point that is chosen will be a measure of economics, balancing the cost of the channel against the cost of the buffer.*

3.1. *Relation Between Activity and Buffer Requirements*

The maximum buffer state is plotted as a function of channel rate in Fig. 3 for the first, second, and third minutes of recorded data. It is assumed that each pel in the segmented area is specified with

* Or if the channel costs are relatively high the operating point will be determined by the maximum tolerable signal delay.

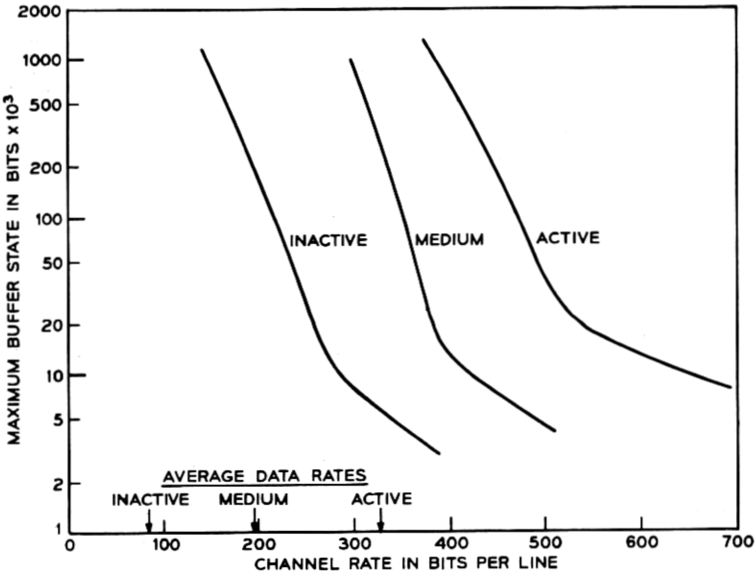

Fig. 3—Channel rate vs buffer state for three different amounts of activity in the data sample. The lower elbow moves upward and to the right as the amount of activity increases.

4 bits and each segment is specified with 16 bits. The average rates are again shown by arrows on the abscissa.

The active sequence has more than four times the average data rate of the inactive sequence, while the average data rate of the sequence having medium activity falls roughly midway between the other two. These three curves have much the same shape with the same pronounced elbow. As the amount of activity decreases, the slope of the segment to the right of the elbow becomes steeper and the position of the elbow shifts downward. This increase in the slope of the curve below the elbow is probably attributable to the fact that for inactive data human movements are shorter in duration and the segmented areas are less clumped than in active data, making smoothing within a field less profitable. Thus larger benefits are obtained with smaller buffers (below 10^3 bits) leading to smaller benefits as the channel rate is reduced to approach the elbow point in the curve.

A coder should be capable of handling the data generated during the active data sample with only occasional overflow since this type of movement could easily be encountered in practice. Thus, if we have a buffer of 50,000 bits, a channel rate of 490 bits per line would be required, according to the curve, for active data. To determine what the buffer is buying us in this instance, compare this with the data rate required by a simple in-frame coder which transmits 4 bits for each picture point. The picture quality of the two schemes would certainly be comparable but the required channel rate for the simple in-frame scheme would be 890 bits per line. Thus, simple frame-to-frame coding with a 50,000-bit buffer in this instance has enabled us to almost halve the required channel-rate.

3.2. *Change of Buffer Requirements as Segment-to-Data Ratio Varies*

The curve labeled (16, 4) in Fig. 4 indicates the maximum buffer-state versus the channel-rate for the medium activity data sample (also shown in Figs. 2 and 3). If all conditions are left the same except for the number of bits used to denote the amplitude of each data point, which is reduced from 4 bits to 2 bits, then the curve labeled (16, 2) results. To compare these curves assume that we have a 50K-bit buffer and determine what channel-rate will just fill this buffer. For the

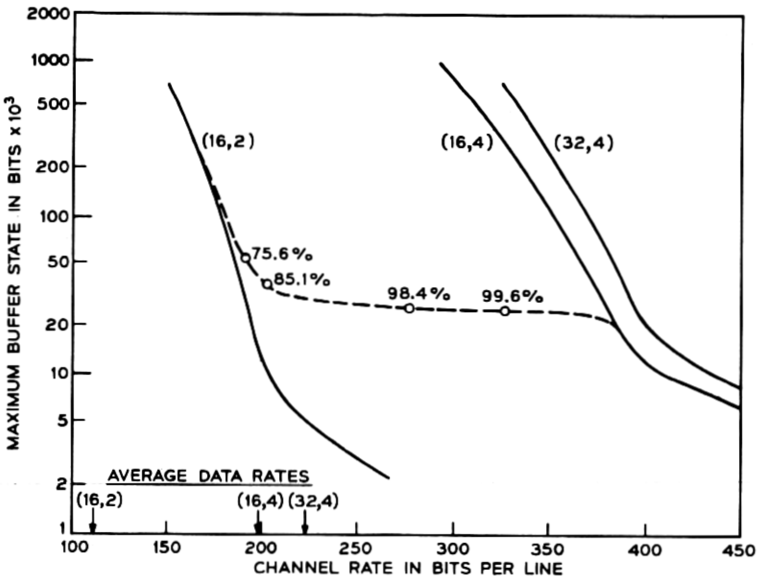

Fig. 4—Channel rate vs buffer state for the medium activity data sample as (*i*) the number of bits/pel required to code the picture is halved (16,2) and (*ii*) the number of bits to code a segment is doubled (32,4) relative to the normal coding (16,4). Also shown is the effect of subsampling when the buffer state exceeds 25,000 bits. Percentage points show time spent in full-resolution mode.

(16, 4) curve the rate is 367 bits per line and for the (16, 2) curve the rate is 185 bits per line, a reduction by almost half (0.504). This is not too surprising in that the average data generation rate is reduced by nearly half (0.556), and reflects the fact that the part of the generated data which specifies the segments is but a small fraction of the total data generated [12 percent for the (16, 4) case].

If, on the other hand, we double the bits used to specify each segment from 16 to 32 the curve labeled (32, 4) results, and as we would expect gives a negligible increase in channel rate for a given buffer capacity (4.6 percent). Consequently, it may prove well worthwhile to transmit additional error protective bits with each segment in an effort to limit the degrading effects of channel errors.

3.3 *Buffer Feedback*

One method for reducing the required buffer size is to use some form of feedback so that when the buffer exceeds a certain capacity the accuracy with which the data is coded is reduced. This may be either a reduction in the accuracy with which the amplitude of the data is specified (reduction in contrast resolution) or a reduction in the number of picture points transmitted (reduction in spatial resolution).[5] The form of the reduction in resolution is not particularly important for this simulation, but we have found that reducing the spatial resolution by a factor of two in moving areas has very little effect on the picture quality.[4,5,6]

The dotted curve in Fig. 4 indicates what happens when feedback from the buffer is used to halve the number of samples transmitted in the segmented area (or alternatively, to halve the number of bits used to specify the amplitude of each element within the moving area). In this case a barrier was put at 25,000 bits; whenever the buffer state exceeded this amount the bit rate assigned to elements within the segmented area was halved. The figure shows rather dramatically how the curve first follows the (16, 4) curve until the buffer state reaches 25,000 bits and then moves over to follow the (16, 2) curve. Thus, the buffer control has effectively limited the maximum buffer-state to approximately 25,000 bits for a large range of channel rates. As the channel rate is reduced, the amount of time spent in the lower data-rate mode increases. Percentage figures beside points on the curve indicate the percentage of time the coder was in the high data-rate mode for that particular channel rate.

The percentage of time spent in the low-resolution mode is quite low until the curve approaches the (16, 2) curve. Thus, assuming a buffer of say 40,000 bits, subsampling enables the channel rate to be reduced from about 375 bits per line to about 225 bits per line. At this level the low-resolution mode is still only used about 10 percent of the time.

The subsampling curve of Fig. 4, however, only gives the story for one level of activity. In Fig. 5 the curves show the effect of subsampling for three amounts of activity. Before one can decide on an operating point for a coding system, that is, a channel rate and a buffer size, one must decide on the maximum level of activity which will be tolerated before additional degradation of the signal takes place either in the form of buffer overload, or, for instance, further reduction in resolution. If a curve is available for this level of activity then an operating point can be selected, for instance, just below the sharp upward turn of the curve.

A corollary of the above is that given a buffer of a particular size the position of the barrier beyond which subsampling is used should be placed as close to the top of the buffer as is possible without overflow occurring (over the allowed range of data activity), or before another stage of resolution-reduction is invoked. In this way the amount of time spent in the high-resolution mode will be maximized. If two stages of subsampling are used then the barriers associated with each

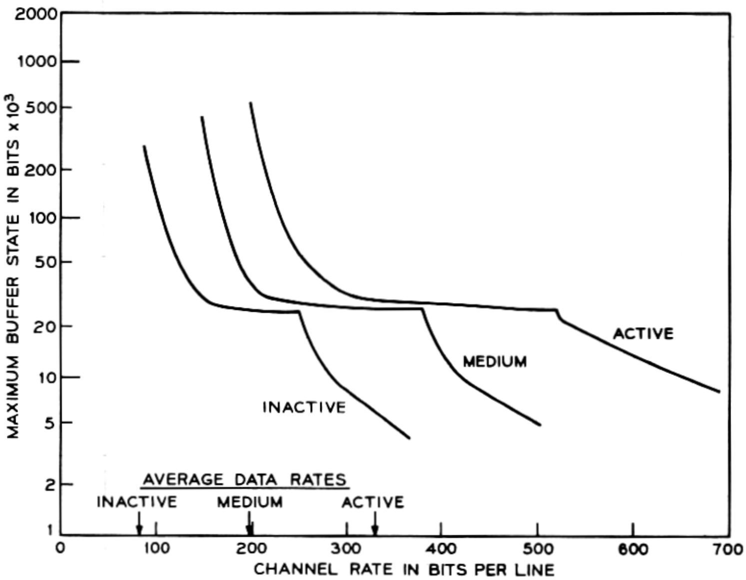

Fig. 5—The channel-rate vs buffer-state curves for buffer-controlled subsampling.

transition should be placed close together and near the top of the buffer. However, if the barriers are too close together then activity rates which would normally be handled by the first stage of subsampling will be handled by the second stage of subsampling. If on the other hand the barriers are too far apart then the first stage of subsampling will be entered prematurely.

For practical reasons one may want to change control only at the beginning of each field. Thus the whole field would be either subsampled or not depending on the state of the buffer prior to the start of the field. This condition was simulated and the results, which are shown in the dashed curve of Figure 6, indicate the importance of being able to change from one mode to another within a field. Without this provision an increase in buffer size of up to 70 percent is required, depending on the operating point.

3.4 *Activity-Controlled Subsampling*

Reduction of spatial resolution by subsampling in the moving area affects picture quality very little because the image is already blurred

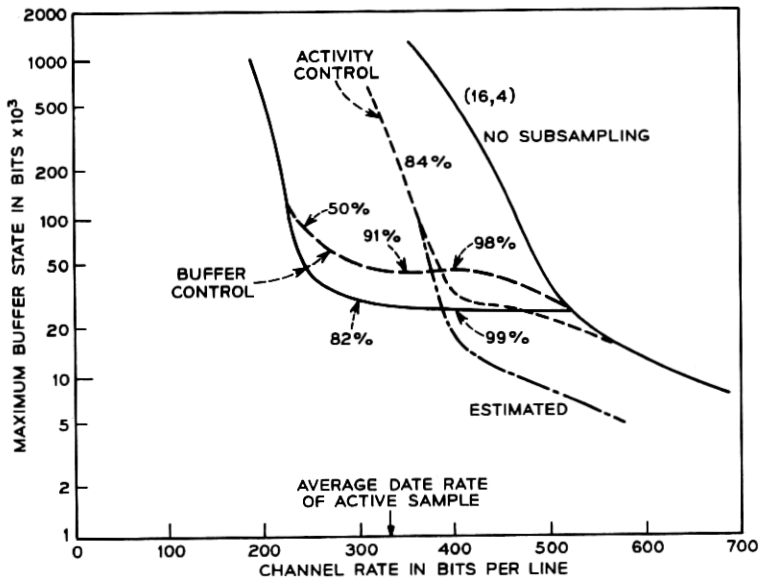

Fig. 6—The effect of activity control on subsampling (dotted curve) is compared with buffer control (dashed curve). With activity control, subsampling is invoked when more than 50,000 bits are required to code the previous field. The active data sample was used.

by the camera.* Also, the human observer is quite tolerant of blurring in a moving image.

The state of the buffer provides some measure of the amount of movement in a picture, but this measure is delayed in time because the buffer acts like an integrator and responds after the fact. A better estimate of the amount of activity is the amount of data generated in a field.

In order to simulate activity-controlled subsampling, the amount of data generated during a field was compared with a threshold. If it exceeded the threshold, subsampling was introduced in the next field. In Fig. 6 the buffer-state versus channel-rate curve for this condition with a threshold of 50,000 bits per field is shown as a dotted curve. Of course, for activity-controlled subsampling the amount of time spent in the high-resolution modes does not vary with channel rate or buffer state and in the example considered it was 84 percent.

One cannot speak of within-field activity control because a single activity figure is derived for the whole field. Subsampling active fields, in effect, alters the structure of the data being smoothed by the buffer, converting it into lower activity data but with a higher mean-to-peak ratio. Thus, the resulting channel-rate versus buffer-state curve should be similar in shape to the curve obtained without control but shifted to the left and down, as shown in Fig. 3 for decreasing amounts of activity.

The result, shown in Fig. 6, is inconsistent with the above in that the activity-controlled curve converges with the curve obtained without control as the channel rate increases. The reason for this is artifactual and stems from the way the control is derived. If one field is active enough to introduce subsampling in the next field, the subsampling will reduce the amount of activity so that the following field will not be subsampled. Thus, with active data, fields will alternate between subsampling and no subsampling. For operating points to the right of the elbow of the uncontrolled curve there will not be enough smoothing to carry the surplus generated in the field that is not subsampled into the next field, and one would expect the curves in the two cases to coincide.

One could overcome this problem by deriving an activity control that was based on the data rate prior to subsampling or (as was used by Candy, et al., in the buffer control case[4]) build hysteresis into the threshold so that the decision to stop subsampling would not be made

* The camera target integrates light falling on it over one frame period and in that time a fast moving object can move a distance of 4 or 5 pels.

until the data rate fell to about half the threshold used in the decision to switch to subsampling in the first place. The dot-dash curve is an estimate of what the channel-rate versus buffer-state curve would look like if this modified type of activity control was employed.

In comparison, we see that buffer-controlled subsampling introduces a horizontal platform between the fully sampled curve and another curve with half the number of amplitude bits assigned to the segmented pels. On the other hand, activity-controlled subsampling produces a horizontal shift of the fully sampled curve to the left.

The curves of Figs. 4 and 6 suggest the interesting possibility that perhaps buffer-controlled subsampling could be added to activity-controlled subsampling. This would give the advantage of having the subsampling phased with the data peaks when the amount of activity is moderate, but having the powerful limiting effect of buffer-controlled subsampling for active-data sources. From knowing the effect of both of the methods of subsampling one could estimate quite accurately the curve resulting from the combined strategy. The results of a simulation of the combined strategy is shown in Fig. 7 where all the conditions

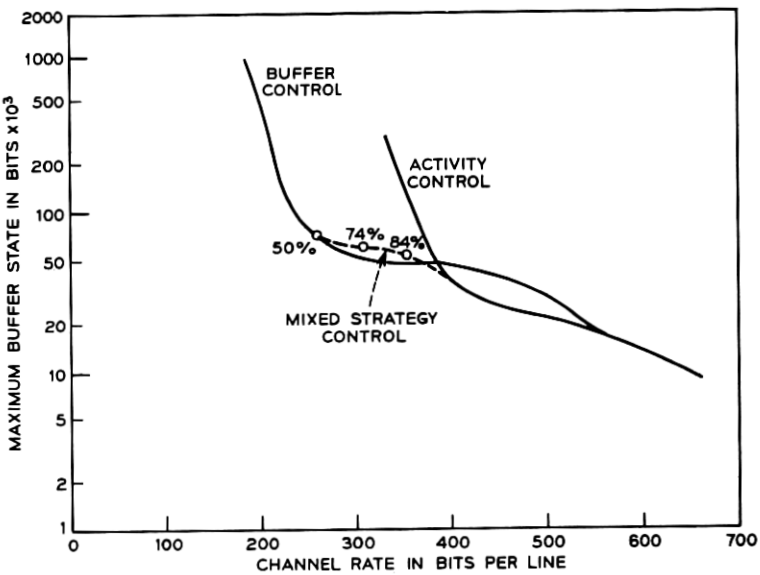

Fig. 7—A mixed strategy is employed. Activity control and buffer control are combined producing the dashed curve which makes a transition from the activity control curve to the buffer control curve as the channel rate decreases. The active data sample was used.

are the same as for the results of Figs. 4 and 6, and control is only exercised at the end of a field.

3.5. *Data Interleaving*

As mentioned previously, an efficient operating point for a coder-buffer combination is likely to be just below the elbow of the channel-rate versus buffer-state curve. At this point the buffer is smoothing within a field but providing very little smoothing from field-to field. As one would expect the average data generation rate for *Picturephone* schemes is not uniform within a frame but peaks toward the bottom.

The data generation rate as a function of vertical position within the field is shown in Fig. 8 for a single field of data. The data rate of this particular field is a local maximum, taken from the active data sample. The peak data rate is 850 bits/line occurring toward the bottom of the picture (probably associated with the model's shoulders) while the average rate is 517 bits/line giving a mean-to-peak ratio of 0.61. For less-active data samples the ratio is even more favorable, 0.47 and 0.35 for a local maximum taken from the medium-activity data sample and the inactive data sample respectively. Plots of the vertical distribution

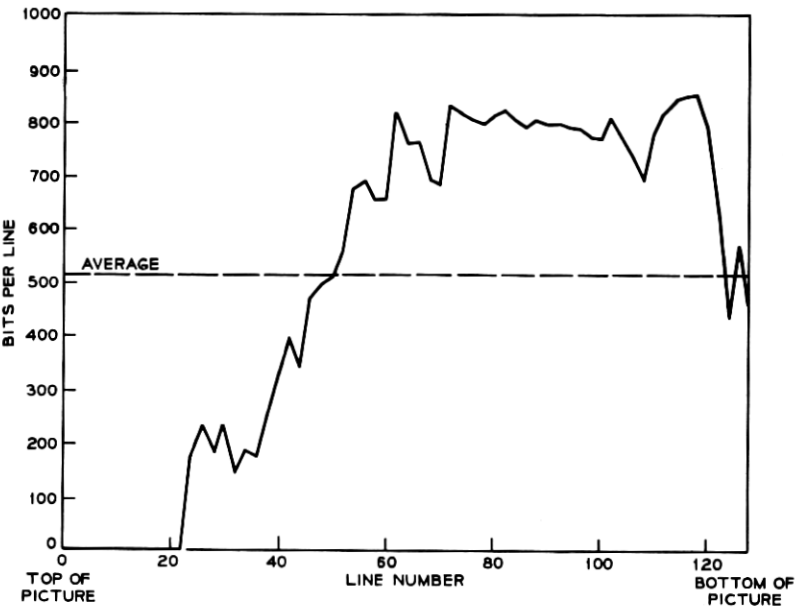

Figure 8—Vertical distribution of data-generation rate throughout field.

of data for adjacent fields, as one would expect, are very similar. Presmoothing of the data can be achieved if instead of reading out the data in the order in which they are generated the data are read so that a line coming from an area having a high average-data-generation rate will be succeeded by a line from an area having a low average-data-generation rate. This arrangement could only be achieved, in general, by having a memory capable of storing an appreciable portion of a frame of coded data at the transmitter to interleave the data and a similar arrangement at the receiver to put the data back in the correct order. In such a situation the advantage to be gained over spending the same amount of money on additional buffer capacity may be quite negligible.

In certain types of coders such as conditional PCM replenishment[7] and conditional in-frame encoding[8] the signal is stored in the transmitter frame memory in much the same form as the transmitted signal. In other coders such as the conditional element-line encoder[8] the signal stored in the frame memory can be converted to the same form as the transmitted data without a great deal of storage (a line memory in the case of the element-line encoder). Thus the frame memory can also be used to interleave the data by providing readout taps at points within a field. In such situations additional memory is required to label the information which is to be transmitted but this only amounts to an increase of about 5 percent in the size of the frame memory.

It is a simple matter to simulate data interleaving since the complete data is recorded on a digital disk in the form of two 8-bit numbers for each line and the data can be read from the disk in any order. The lines were taken in the order 1, 33, 65, 97; 2, 34, 66, 98; 3 ··· 127; 32, 64, 96, 128; 129, 161, 193, 225; 130 ··· 160, 192, 224, 256. Some small improvement may be obtained by using the order 1, 65, 33, 97; ··· over the one above.

The effect of 4:1 data interleaving is shown in Fig. 9, for the medium-active data sample. As would be expected, data interleaving has little effect with large buffers where smoothing takes place over many frames. However, at the elbow the curve for interleaving continues to fall until at a channel rate of 450 bits per line the maximum buffer states with and without data interleaving differ by more than a factor of 10. The difference in the 99-percent points (1080 bits as against 55 with data interleaving) is even larger.

IV. DISCUSSION AND CONCLUSIONS

With little or no buffering a frame-to-frame coder requires a large channel rate. As the size of the buffer is increased, the required channel

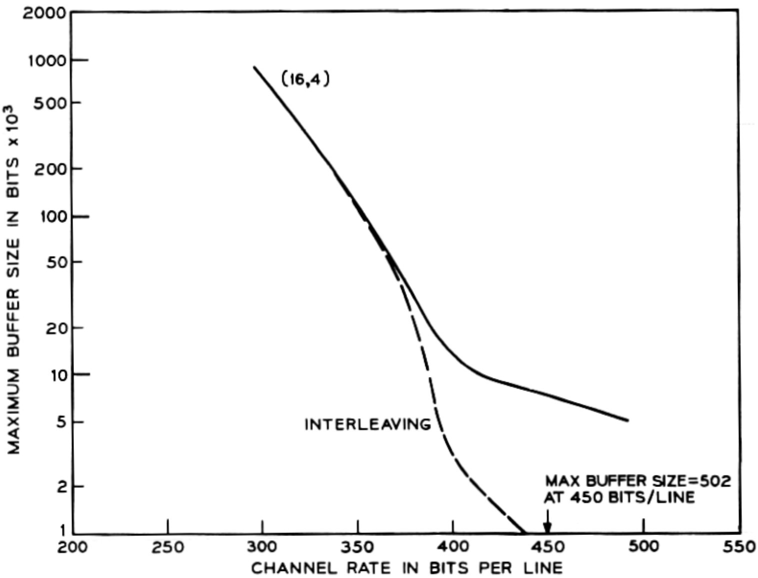

Fig. 9—Compared with normal buffering (solid-line curve) 4:1 data interleaving reduces the required buffer size by a factor of ten at higher channel rates.

rate decreases quite rapidly until the buffer is large enough to smooth the data over an entire field. Beyond this point there is very little improvement until the buffer is large enough to smooth the data from one period of active movement to the adjacent inactive period. Buffers large enough to do this (greater than 10^6 bits), however, introduce transmission delays which are intolerable in normal conversations.

The curve of maximum buffer-state versus channel-rate summarizes the buffer requirement of a coder for a particular level of activity and enables one to explore possible tradeoffs in the selection of a suitable operating point for a coder. For active movement and a 50,000-bit buffer, the use of simple frame-to-frame coding results in nearly a two-to-one saving over the channel rate required by an in-frame coder giving roughly equal picture quality. Further reduction is obtained if the rate of data generation during periods of peak activity is controlled by using either the amount of activity in the picture or the state of the buffer as the controlling variable.

It is assumed that the coder can switch between one of two modes—a normal mode and a reduced resolution mode, which has about half the data generation rate. A simple buffer threshold control effectively clamps the buffer state to the value of the threshold until the coder is operating predominantly in the reduced resolution mode. If the control signal is only permitted to change the coder mode at the end of a field then the clamping action is less effective, since the buffer, in the worst case, may have to accommodate a large amount of data generated in a busy field prior to switching. This results in an increase in required buffer size of up to 70 percent depending on the channel rate.

Activity control, on the other hand, does not change the basic shape of the buffer-state curve but instead makes the coder appear as though it has a lower overall data rate. Activity control is also an effective method of reducing the average data rate which is the most important parameter in TASI-like channel-sharing schemes.

Ideally one would like to use the speed of the moving object as the control signal rather than the data generation rate because the reduced data-rate mode would then be used only when the subjective effect on picture quality is very small.

Buffer control and activity control can be combined so that when either the buffer threshold or the activity threshold is exceeded the coder switches to the reduced data-rate mode. The activity control operates more rapidly than the buffer control, enabling the reduced data-rate mode to be restricted more to the periods of peak activity. For small buffer sizes, however, the addition of activity control would probably help picture quality very little. The combined operating mode may find application where a channel sharing and buffer sharing arrangement is applied to a small number of users. In such an arrangement both the average data rate and the peak data rate are important.

Most of the advantage in buffering is obtained by smoothing data over a field. This gain is possible only because the data are not generated at a uniform rate throughout the field but tend to be concentrated near the middle and bottom of the picture. If data from different places in the field are interleaved the resulting bit stream can be smoothed with a much smaller buffer. A four-fold interleave was simulated and the size of the buffer at the elbow point was reduced by a factor of four. For slightly larger channel rates the reduction in buffer size was greater than a factor of 10. The frame memory can only be used to obtain the temporary storage required for data interleaving when the data in the frame memory is in the same form as the transmitted data or when the data can be converted economically to be of the same form. A small, but interesting class of coders falls in this category.

V. ACKNOWLEDGMENT

I especially thank B. G. Haskell and R. F. W. Pease for their suggestions and advice. In addition B. G. Haskell was largely responsible for transferring the data to the computer, and W. G. Scholes was ever ready to help with the data acquisition.

REFERENCES

1. Jesty, L. C., "Television as a Communication Problem," Proc. IEEE *99*, pt. III A (May 1952), p. 761. See also Kell, R. D.: British Patent No. 341811, 1929.
2. Seyler, A. J., "The Coding of Visual Signals to Reduce Channel-Capacity Requirements," Proc. IEEE *109* (September 1962), pp. 676–684.
3. Brainard, R. C., Rowlinson, D. E., and Cutler, C. C., "*Picturephone*® Communication with Delay," private communication.
4. Candy, J. C., Franke, Mrs. M. A., Haskell, B. G., and Mounts, F. W., "Transmitting Television as Clusters of Frame-to-Frame Differences," B.S.T.J., *50,* No. 6 (July–August 1971), pp. 1889–1917.
5. Pease, R. F. W., and Limb, J. O., "Exchange of Spatial and Temporal Resolution in Television Coding," B.S.T.J., *50,* No. 1 (January 1971), pp. 191–200.
6. Limb, J. O., and Pease, R. F. W., "A Simple Interframe Coder for Video Telephony," B.S.T.J., *50,* No. 6 (July–August 1971), pp. 1877–1888.
7. Mounts, F. W., "A Video Encoding System Employing Conditional Picture Element Replenishment," B.S.T.J., *48,* No. 7 (September 1969), pp. 2545–2554.
8. Limb, J. O., and Pease, R. F. W.,"Frame-to-Frame Coding by Transmission of Conditional In-Frame Coded Signal", to be published.