Copyright © 1971 American Telephone and Telegraph Company
THE BELL SYSTEM TECHNICAL JOURNAL
Vol. 50, No. 6, July-August, 1971
Printed in U.S.A.

A Simple Interframe Coder For Video Telephony

By J. O. LIMB and R. F. W. PEASE

(Manuscript received February 9, 1971)

The technique of exchanging resolution according to the amount of movement in a picture has been previously described; in stationary parts of the picture the temporal resolution is reduced while in moving parts of the picture the spatial resolution is reduced. Here, we describe a method of applying resolution exchange to a differentially quantized (DPCM) signal. The resulting channel capacity required for the subjectively satisfactory transmission of the differential signal is halved. The coder is simpler than most interframe coders and should not increase the sensitivity of the system to channel errors.

I. INTRODUCTION

In a previous paper we described a way to halve the channel capacity required for the subjectively satisfactory transmission of an 8-bit PCM television signal by exchanging spatial and temporal resolution according to the amount of movement in the local part of the picture[1] Every second picture element ("pel") is sampled and in stationary areas of the picture the values of the unsampled pels are interpolated from adjacent temporal samples (reduced temporal resolution); in the moving areas the values of the unsampled elements are interpolated from neighboring sampled elements in the same line (reduced spatial resolution).

We would like to apply this technique to a signal whose bit rate has already been reduced by an element-to-element differential quantizer (EDQ), e.g., the *Picturephone*® codec. Unfortunately, halving the horizontal sampling rate, as in Ref. 1, increases the amplitude of the sample-to-sample difference signal which, in turn, requires a larger number of quantizing levels for adequate representation. There are two ways around this problem:

(*i*) Use the vertically adjacent elements as a prediction of the current pel and quantize the resulting difference. Because vertically adjacent pels are in the previous field, such a technique can be called field difference quantization and is the subject of another study.

(*ii*) Reduce the vertical resolution rather than the horizontal resolution so that the full horizontal sampling frequency is retained; every transmitted line is left completely intact. This is the method described here.

II. PRINCIPLE AND IMPLEMENTATION

In Fig. 1 we show an outline of a system which uses resolution exchange in conjunction with differential quantization of the signal. The output of the television camera is differentially quantized and fed to a movement detector which usually consists of a frame delay circuit, a difference and threshold circuit, and associated logic. When movement is not detected, that part of the picture being coded enjoys the full spatial sampling frequency but each line is only transmitted every second frame, i.e., the temporal sampling frequency is halved. When movement is detected, then the vertical sampling frequency is halved by transmitting only every second line of the picture. For a

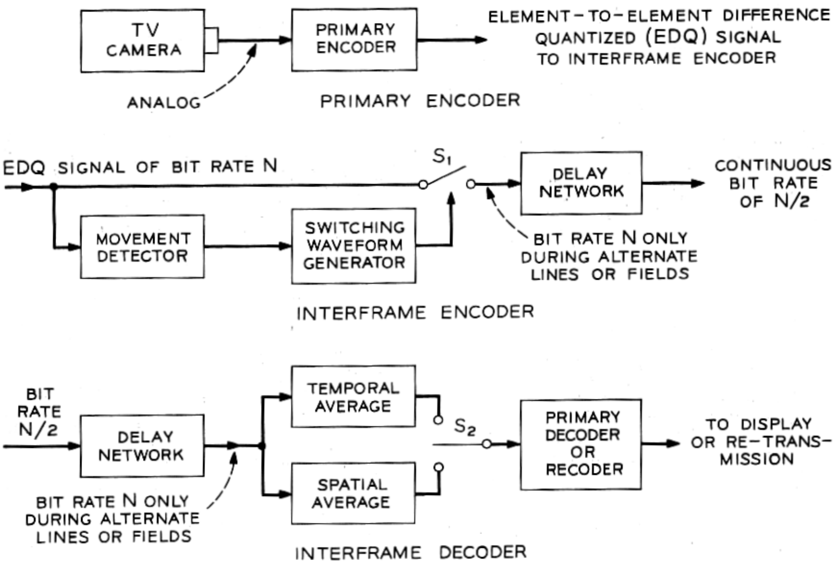


Fig. 1—Schematic of communication system using simple interframe coder.

2:1 interlaced scan this means transmitting every other field and so the temporal sampling frequency is raised to 30 Hz.

At the receiver, when movement is not detected, alternate (sampled) lines in each field are decoded and displayed. In place of the unsampled lines in the field a temporal average of corresponding lines from neighboring sampled frames is formed and displayed.*

To describe this "stationary mode" we assign to each line coordinates (y, t) which refer respectively to the vertical position in each frame (or line number) and the temporal position (or field number). Figure 2 shows diagrams of lines for a 2:1 interlaced scan format. Each sampled line is marked with a full dot and the unsampled lines are shown as circles; the averaging is denoted with arrows. In the stationary mode we receive and display alternate lines in each field, i.e., lines 2, 3, 6, 7, 10, 11, $\cdots$ in fields 2 and 3 and lines 0, 1, 4, 5, 8, 9, $\cdots$ in fields 4 and 5. The averaging is done only along the time (t) axis and so stationary pictures are displayed with the full resolution of the fully sampled picture.

The results of preliminary experiments with this mode of operation continually applied to the whole picture show better motion rendition than does frame repeating but are still unsatisfactory for moderate and fast movement of subjects of normal contrast (the degradation becomes annoying at speeds of about 2 pels per frame interval). In the case of still pictures the quality is actually improved over that of normal (3- or 4-bit) EDQ because the temporal (frame-to-frame) averaging reduces the visibility of granular noise.

When movement is detected, pels from alternate sampled fields are received, decoded, and displayed. In place of pels from the unsampled fields, an average is formed from the four nearest neighbors in the $y - t$ diagram, as shown in Fig. 2b. For example, pels in lines $(y, 1)$ are replaced with the average value of pels in lines $(y - 1, 0)$ $(y + 1, 0)$ $(y - 1, 2)$ and $(y + 1, 2)$. Thus both spatial and temporal interpolation is used to form the new value in the unsampled field.

Preliminary experiments with this mode applied continually to the whole picture showed adequate motion rendition for most head and shoulders views of a person talking; the loss in vertical resolution is barely noticeable, but in some regions (especially dark regions) of high vertical detail, some aliasing patterns can be seen. Under normal viewing conditions (see Section III), fast movement (4 pels per frame interval and above) of a contrasty edge appears slightly jerky.

* Some other work on reducing the temporal resolution is given in Refs. 2 and 3.

There are a number of ways in which the sampling mode can be controlled. The subjectively ideal, but probably most difficult, method is to change to the appropriate mode as each element is encountered. Another method is to code the whole of one line in the same mode, while a further method is to code the whole field in the same mode. The method adopted in most of our experiments is to use one mode throughout each field. The transitions from stationary mode to moving mode can be made at the beginning of any field but the reverse transition is only permitted at the beginning of any even field, counting

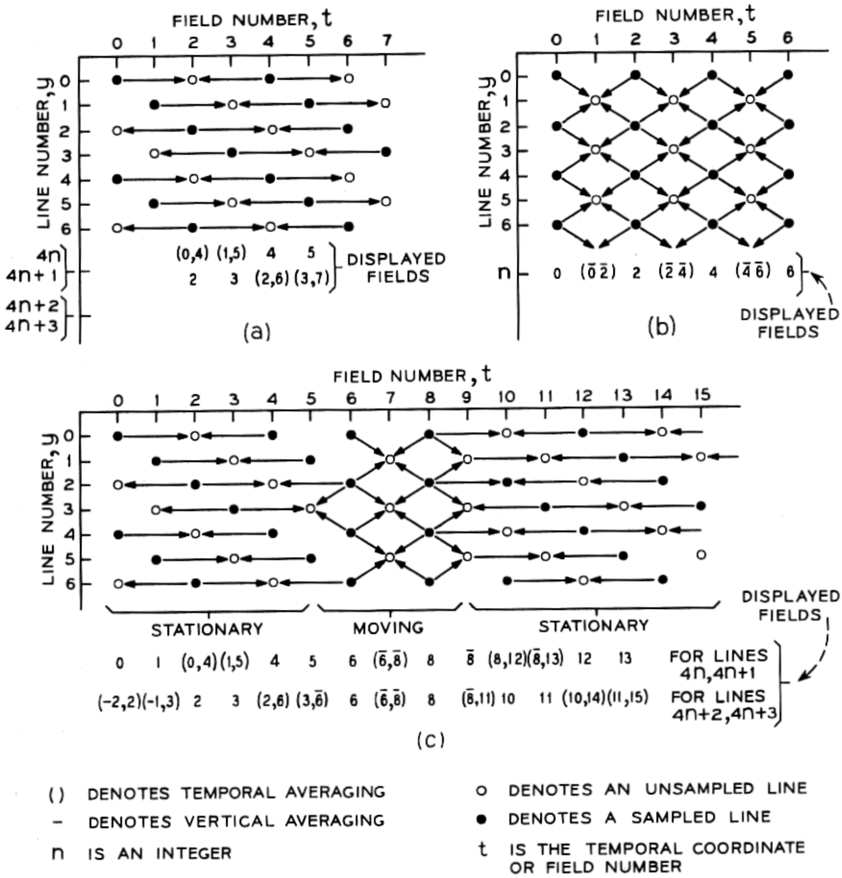


Fig. 2—Sampling patterns of lines and sequences of fields in the displayed picture for the different modes: (a) Stationary mode. (b) Moving mode. (c) A combination of modes (except near the mode changes, the fields represented in the output correspond with the temporal coordinate t).

from after the change from stationary to moving mode. This is done to prevent data being generated at a greater rate during two consecutive fields than can be handled by the channel. The mode changing information is negligible as only one extra bit of information is required after every field.

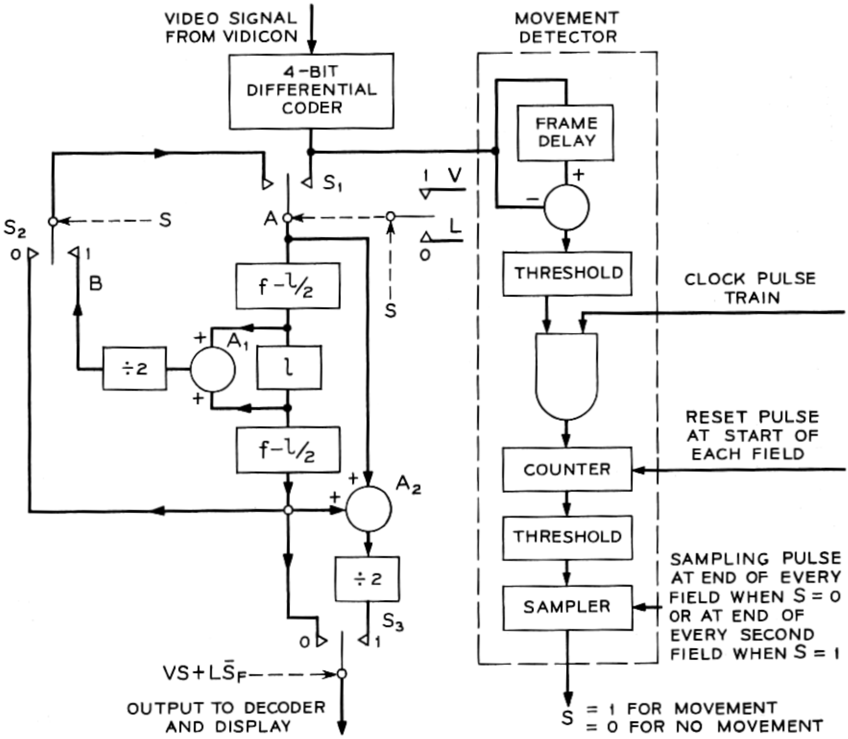
The delay network at the transmitter (Fig. 1) converts the effectively halved bit rate to a continuous bit stream of half the original rate; a similar network at the receiver reconverts the continuous bit stream back to alternate fully sampled fields or frames. The operation of these networks is described more fully in Section IV.

Fig. 2c illustrates the line sampling pattern and the interpolation used when the system starts in the stationary mode (fields 0, 1, 2, 3, 4, 5) and switches to the moving mode (fields 6, 7, 8, 9) and then switches back again. Also shown are the sequences of fields represented in the displayed picture. In both modes the lines displayed correspond either directly to the temporal coordinate or indirectly by temporally averaging from fields temporally equidistant from the current field. Near the changeovers some lines are misplaced temporally; for instance the unsampled lines of field 5 are replaced with lines from field 3 and 6 whose mean temporal position is $4\frac{1}{2}$. Different weights could be assigned to the signals of fields 3 and 6, but for simplicity we decided to subjectively test the coder using equal weight for all averaging.

To experimentally test the coder we simulated Fig. 1 with the arrangement of Fig. 3. The picture format, scenes, and viewing conditions were the same as those used in the previous resolution exchange experiments,[1] i.e., there were 271 lines per frame with a 2:1 line interlace and the frame frequency was 30 Hz; the sampling frequency was 2 MHz. The scenes were head and shoulders views of a variety of subjects engaged in conversation varying from quiet to violent. The display raster was approximately $5\frac{1}{2}$ inches horizontally by 5 inches vertically and was viewed at 3 feet under ambient illumination typical of a well-lit office (about 70 footcandles).

The primary coder is a 4-bit digital differential quantizer* whose quantizing levels are shown in Table I. The accumulated value, with a peak value of 127 levels, is fed to the frame memory via switch S_1 (Fig. 3a). In the stationary mode S_1 and S_3 are switched at the line rate and S_2 is held at 0. The waveform L, used for switching at line

* Although the quantizer has 17 levels and, strictly speaking, could not be transmitted as a 4-bit signal, certain combinations of levels were deleted so as to permit 4-bit transmission.[4]



f−l/2	DELAY OF 1 FIELD TIME LESS ONE-HALF LINE TIME
l	DELAY OF 1 LINE TIME
L,V	WAVEFORMS SWITCHING AT LINE RATE AND FIELD RATE RESPECTIVELY (SEE FIG.3B)
$\bar{S}_F$	INVERSE OF SIGNAL S DELAYED BY ONE FRAME TIME

Fig. 3a—Experimental arrangement for testing simple interframe coder. Alternate fields or lines are fed to the frame memory via switch S_1. Switch S_2 selects either the previous field or lines from the previous frame for recirculation in the (delay line) frame memory. Adder A_1 performs the vertical averaging and adder A_2 performs the temporal averaging. Switch S_3 selects the required output.

rate, is shown in Fig. 3b as two waveforms of opposite phase. One phase applies to line numbers 1, 2, 5, 6, 9, 10, $\cdots$, or $4n$ and $4n + 1$, where n is an integer, and the opposing phase applies to line numbers 3, 4, 7, 8, $\cdots$, or $4n + 2$ and $4n + 3$. Thus, during one field the value of L changes each successive line, and for a given line the waveform, L, changes polarity for each frame. Figure 3b also shows which fields are present at points A, B, C (Fig. 3a) for each of the two sets of

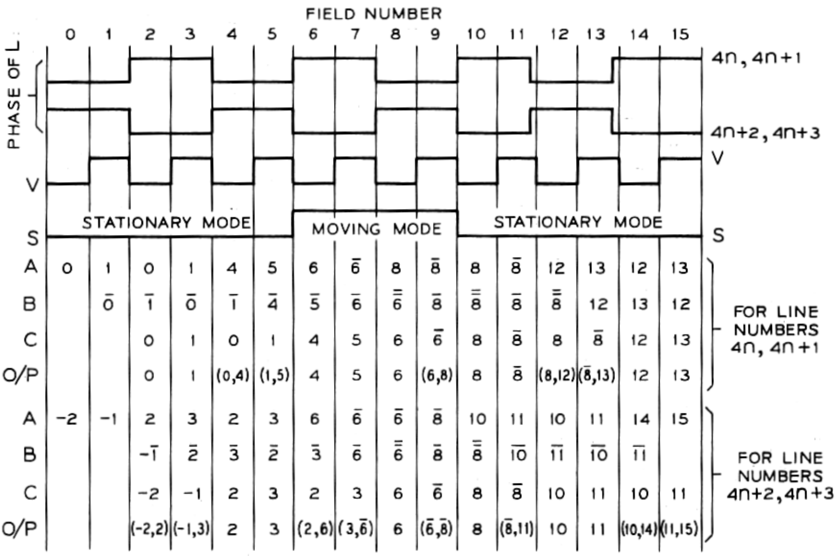


Fig. 3b—Waveforms L, V, and S and fields present at points A, B, C (Fig. 3a) and at the output for both sets of lines. In real time, waveform L switches at the line rate but is shown here as two waveforms, one for each set of lines switching at the frame rate.

lines, for the same sequence of fields and modes as described in the previous section and shown in Fig. 2. In the stationary mode ($S = 0$) the required output can be obtained either by continuously averaging the signals present at A and C or by switching S_3 at the line rate; the latter method is used because it allows greater simplicity when changing modes.

In the moving mode S_1 and S_3 are switched at the field rate (waveform V) and Fig. 3b again shows which fields are present at points

Table I—Quantizing Levels

Level	Decision Level	Representative Level
0		0
±1	1/2	1
±2	1-1/2	2
±3	3	4
±4	5	6
±5	7-1/2	9
±6	11-1/2	14
±7	16-1/2	19
±8	22	25

A, B, C of Fig 3a; the output is formed by switching between C and (A + C)/2 at the field rate. To bring about the same sequence of outputs near the mode changes, as shown in Fig. 2, the waveform-controlling switches S_1 and S_2 are changed as soon as S changes, but the change of waveform-controlling switch S_3 is delayed by 1 frame. The sequences of fields present in the output in the experimental system (Fig. 3a) are also shown in Fig. 3b and correspond to the sequences shown in Fig. 2c.

In the experimental system, movement is deemed to be present if, during any field, 512 or more pels exhibit a frame difference amplitude greater than 15 levels (out of 255). To return to the stationary mode, less than 512 picture elements must have a frame difference amplitude greater than 15 levels during the even field where the field number is counted from the field in which the transition is made to the moving mode.

III. RESULT

The picture quality was consistent with the results of the preliminary experiments conducted on the separate modes as described above. The two faults of the moving mode, the aliasing patterns in areas of strong vertical detail and the slight jerkiness of fast-moving contrasty subjects, were still visible. Degradation due to the switching of modes was seldom visible and the effect of switching the whole picture instead of just the moving areas was not troublesome even when the plane of sharpest focus was midway between the subject's head and the curtains in the background.

In some related experiments the switching of modes was confined to the moving area. The moving area detector examined a sequence of eight frame differences to decide whether the current picture element belonged in a moving area.[1] The resulting pictures of similar scenes viewed under the same conditions were no more pleasing and the movement detector setting was more critical; i.e., with a poor setting slowly moving sharp edges tended to break up due to intermittent mode switching.

IV. DISCUSSION

4.1 *Delay Requirements*

At the transmitter every second pel is delayed by one line period, or one field period (according to the mode), to bring about a constant bit rate. This delay can also be used by the movement detector for generating the required frame differences. Of course, the movement detector must now work on one-quarter as many points as before. But because of the global nature of the decision and the large number of supra-threshold points required to change to the moving mode (512), reliable detection of movement could probably be performed with even less than this number of points. Thus, if the primary (EDQ) coder has an output bit rate of 6 Mb/s (or 200,000 bits per frame) a delay store of 50,000 bits (half a field) is required at the transmitter.

At the receiver a similar delay is needed to convert the incoming constant bit rate to the required 6 Mb/s rate during alternate lines or fields and a frame memory of 200,000 bits is needed to store the pels required for display.

It is tempting to devise schemes in which the alternate fields or lines of data from different sources are multiplexed so that the 50 k-bit delay stores at the transmitter and receiver are unnecessary. However, such schemes have so far been less practicable than using the extra storage; the main difficulty lies in synchronizing any two cameras in an unsynchronized system.

4.2 *Recoding of Output*

When the primary decoder is located remotely from the interframe decoder, the unsampled fields or frames (depending on whether there is movement or not) must be recoded since, when different quantized differences (representative values) are averaged, the resulting difference will not necessarily belong to the set of representative values allowed by the in-frame decoder. A circuit to achieve this is shown in Fig. 4.* The averaged difference signal is added to the error term from the previously quantized level: the output of the quantizer is subtracted from the input to form the new error term. The advantage, previously mentioned, of reducing granular noise in the stationary mode is lost in the process of recoding. The effect of recoding was tested experimentally by differentially quantizing the output of switch S_3 (Fig. 3a) before decoding and displaying the signal. There was no appreciable increase in noise (over the primary coded signal) due to recoding either in the stationary mode or the moving mode.

In a switching system it may be desirable to convert and reconvert the signal between the full rate and the half rate many times. This

* This is Cutler's error feedback coder.[5]

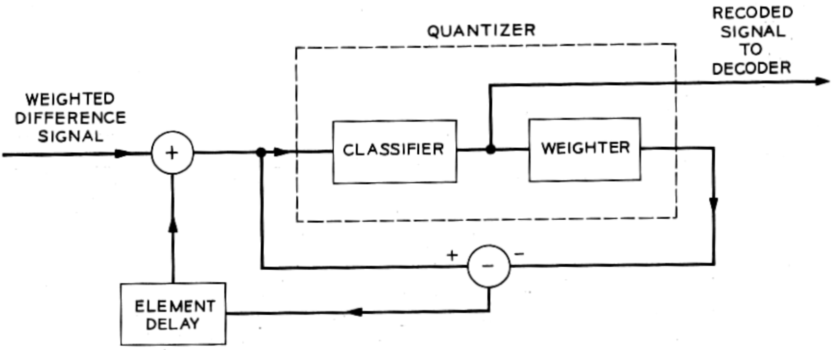


Fig. 4—Recoder to convert averaged weights to standard weights.

can be achieved with no loss of quality other than that introduced in the initial conversions by having the full-rate signal contain one code word in each field to indicate the sampling mode; thus no further movement detection need be employed and subsequent interframe encodings are simpler than the initial one. Decoding will remain unchanged.

4.3 *Comparison*

Compared with other interframe coders[6,7] the simple interframe coder has a higher bit rate and, under certain conditions, a poorer picture quality. However, there are certain offsetting advantages which are described below.

One advantage is the relatively low memory requirement (300,000 bits for the transmitter and receiver combined). Other existing interframe coders require 530,000 bits of storage for the frame memory at the transmitter and the same again at the receiver, and sophisticated buffers are also required because of the randomness in the generated bit rate of these coders. It is possible to use smaller stores in such coders by requantizing the input signal. This operation produces a small loss in picture quality and also it is not yet clear what effect it will have on recoding the signal.[8]

The simple interframe coder can be used with any primary coder which uses only previous pels along the line for prediction (see, for example, Refs. 4, 9, 10). The primary encoding stage may well be located at the first level of switching while the frame-to-frame coding section may be located at a higher level in the switching hierarchy. With such an arrangement, the secondary encoding stage can be bypassed altogether when less than half the *Picturephone* trunk connections in a group of trunks is not in use.

The effect of channel errors on the received picture is probably about the same as in a standard DPCM system. If element-to-element differential encoding were used in the primary encoder, errors would be confined to a single line (in the received signal) as the accumulators at the transmitter and receiver are reset at the end of each line.[4] If the movement detector is in the stationary mode, the effect of an error in one frame will also be displayed at half-amplitude in neighboring frames; but on the other hand, no errors will originate in the neighboring frames. Similarly, if an error occurs in the moving mode, the effect will appear at full-amplitude in the sampled field and at quarter-amplitude in the two adjacent lines in each neighboring field, but the neighboring fields will contain no indigenous errors. Most error detection techniques that can be applied to the primary encoding section can be used without modification when the frame-to-frame coding section is added (e.g., the check-summing method of error detection suggested for the differential quantizer).[4]

Apart from one mode-bit transmitted every second field the coder has no change-of-mode words, addresses, start-of-line words, or start-of-frame words. In a coder using these special words, an error in the received data, if it occurs in a special word, can be especially troublesome.

4.4 *Encoding Using the Previous Line*

There may be occasions when the primary coder uses pels from the previous line (in the same field) as well as from the present line for predicting the value of the current pel (see for example, Ref. 11). The stationary mode described above is now unsatisfactory because only every other line in each field is available at the receiver. We therefore modified the apparatus of Fig. 3 to evaluate a coder in which alternate frames are transmitted in the stationary mode and temporal interpolation is used to replace the unsampled frames. The moving mode is unchanged and whole fields are now left intact in both modes. The change from a stationary to a moving mode can now be made only at the end of a frame rather than at the end of a field, as before. In addition, the delay stores for converting to and from a constant 3Mb/s rate are twice as large as required before, but the total required memory is still less than half that of other existing interframe coders. Experimental results with this coder gave pictures which were as satisfactory as those already described.

V. ACKNOWLEDGMENTS

We would like to acknowledge the many discussions with A. J. Seyler of the Research Laboratories of the Australian Post Office and with our colleagues, J. C. Candy, D. J. Connor, B. G. Haskell, and F. W. Mounts. We are also indebted to W. G. Scholes for his technical assistance.

REFERENCES

1. Limb, J. O., and Pease, R. F. W., "Exchange of Spatial and Temporal Resolution in Television Coding," paper presented at S.M.P.T.E. Meeting, Los Angeles, October 1969; also B.S.T.J., *50,* No. 1 (January 1971), pp. 191–200.
2. Cunningham, J. E., "Frame Correction Coding," Symposium on Picture Bandwidth Compression, MIT, Cambridge, Mass., April 1969.
3. Brainard, R. C., Mounts, F. W., and Prasada, B., "Low-Resolution TV: Subjective Effects of Frame Repetition and Picture Replenishment," B.S.T.J., *46,* No. 1 (January 1967), pp. 261–271.
4. Limb, J. O., and Mounts, F. W., "Digital Differential Quantizer for Television," B.S.T.J., *48,* No. 7 (September 1969), pp. 2583–2599.
5. Cutler, C. C., "Transmission Systems Employing Quantization," U. S. Patent No. 2,927,962, March 8, 1960.
6. Mounts, F. W., "A Video Encoding System Using Conditional Picture-Element Replenishment," B.S.T.J., *48,* No. 7 (September 1969), pp. 2545–2554.
7. 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., this issue, pp. 1889–1917.
8. Candy, J. C., and Mounts, F. W., private discussion.
9. Brown, E. F., "A Sliding-Scale Direct-Feedback PCM Coder for Television," B.S.T.J., *48,* No. 5 (May–June 1969), pp. 1537–1553.
10. Bosworth, R. H., and Candy, J. C., "A Companded One-Bit Coder for Television Transmission," B.S.T.J., *48,* No. 5 (May-June 1969), pp. 1459–1479.
11. Connor, D. J., Pease, R. F. W., and Scholes, W. G., "Television Coding Using Two-Dimensional Spatial Prediction," B.S.T.J, *50,* No. 3 (March, 1971), pp. 1049–1061.