

Document Room, DOCUMENT ROOM 36-412
Research Laboratory of Electronics
Massachusetts Institute of Technology

#2

302

MEASUREMENT OF SECOND-ORDER PROBABILITY DISTRIBUTIONS OF PICTURES BY DIGITAL MEANS

JAMES CHURCHILL STODDARD

LOAN COPY
only

TECHNICAL REPORT 302

JULY 13, 1955

RESEARCH LABORATORY OF ELECTRONICS
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
CAMBRIDGE, MASSACHUSETTS

The Research Laboratory of Electronics is an interdepartmental laboratory of the Department of Electrical Engineering and the Department of Physics.

The research reported in this document was made possible in part by support extended the Massachusetts Institute of Technology, Research Laboratory of Electronics, jointly by the Army (Signal Corps), the Navy (Office of Naval Research), and the Air Force (Office of Scientific Research, Air Research and Development Command), under Signal Corps Contract DA36-039 SC-64637, Project 102B; Department of the Army Project 3-99-10-022.

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
RESEARCH LABORATORY OF ELECTRONICS

Technical Report 302

July 13, 1955

MEASUREMENT OF SECOND-ORDER PROBABILITY
DISTRIBUTIONS OF PICTURES BY DIGITAL MEANS

James Churchill Stoddard

This report is based on a thesis submitted to the Department of Electrical Engineering, M. I. T., May 23, 1955, in partial fulfillment of the requirements for the degree of Master of Science.

Abstract

The transmission of a picture was approached from the point of view of statistics and information theory. A picture was approximated by the processes of sampling and quantizing the video waveform that represents the intensity of the picture as it is scanned. Transmission of each sample of the picture is reduced to the problem of sending a number that indicates which of the 32 possible intensity levels occurred in that sample.

Equipment that measures the second-order probability distribution of a video waveform from a facsimile transmitter by digital means was designed and built. System tests were made to permit evaluation of the potential accuracy of measurement. Although perfectly repeatable results from a known waveform could not be obtained, measurement of the complete joint probability distribution of one picture was made by running the data until consistent results could be obtained for each set of levels. A value of entropy representing an upper bound to the information content of the picture was computed from the measured second-order probability distribution of intensity in adjacent samples.

I. INTRODUCTION

1.1 BACKGROUND OF INFORMATION THEORY

Information theory is basically a statistical study of communication systems. It can be applied, at least in principle, to almost any communication system. For examples we might take a simple buzzer signaling system or a complex color television system. Information theory does not show directly how to improve a system, except in the simplest cases, but rather provides a measure of the efficiency of a particular system.

The communication system usually encountered involves a continuously varying quantity that can assume any value within certain limits. Before the advent of information theory, the usual criterion for quality in a communication system was waveform reproduction, which was evaluated in terms of sinusoidal frequency response, distortion, and signal-to-noise ratio. A considerable amount of work has been done on the requirements of a speech channel. For a channel of certain specifications of waveform reproduction, a certain per cent of a standard set of words is intelligible to the average person. This does not mean that if some different modulation scheme were used before transmission over the channel and a suitable demodulation system were used after the reception of the signal the same criteria would hold. Information theory goes deeper than that and inquires into what has to be sent in order to convey the message. Most practical situations involve the human being as the transmitter and/or receiver and this presents a difficulty, since part of the "system" in the general sense is the sensory system of the human being.

In theoretical considerations, it is often convenient to reduce a continuous system to a discrete system. The approximate representation of a continuous quantity by a set of discrete quantities can be approached from several points of view, but the one most useful for this problem is that of sampling and quantizing. The sampling theorem states that if a time-varying function is sampled at a rate equal to twice the highest frequency present, no information is lost (1). In nature there are always factors, such as stray capacity in an electrical system, that limit the high-frequency response so that a finite sampling rate leads to a negligible error. Quantizing is the means of restricting the number of values that the time function can assume. For example, if the time function has a value between a and b , we represent it as level x , and if it has a value between b and c , we represent it as level y , and so on. There is always some kind of "noise" in a communication system; hence if the size of a quantum step is small compared with the noise, the quantized signal is a good representation of the original signal. The term noise, as used here, may be actual electric noise or it may be uncertainty in the receiving or transmitting device. In this sense, intensity-level changes that the eye cannot perceive are below the noise level in the receiver, since the eye is part of the receiving apparatus.

Another approach to the application of information theory to continuous channels is discussed by Fano (2). The signal is not quantized and the expressions that are to be

given for the discrete channel have analogs in the continuous case. However, since the measurement technique used here is digital and performed on the signal after sampling and quantization, only the discrete case will be discussed.

Two of the basic concepts of information theory are the measure of information and statistical coding. Consider the communication system in which a counter man in a lunchroom signals the short-order chef. Maybe the means of signaling is a code rung on a buzzer with dots and dashes. Certain dishes are requested more frequently than others and it saves time if a short series of dots and dashes, that is, a short code word, is used for these items. In other words, the length of the code words should be assigned according to the probability of their use so that on the average the length of code word per message would be minimized. From this discussion, it seems obvious that when the chef receives a request for a very unusual dish, he is receiving more information than usually. That is, if most of the customers at this lunch counter have scrambled eggs for breakfast, the chef should not feel that he has received unusual information whenever the order for that dish comes over the communication system.

Suppose that the messages to be transmitted are represented by a finite set X_1, X_2, \dots, X_n ; then the amount of information supplied by X_k alone is given by $I(X_k) = -\log P(X_k)$, where $P(X_k)$ is the probability of the X_k -th message. This is based on the assumption that all successive messages are statistically independent, that is, that the probability of a message is independent of all precedent messages.

The logarithmic measure of information is the only measure that satisfies the desired properties of a measure of information. The base of the logarithm determines the size of the units of the measure, and it is usually taken as two. (In this report all logarithms are to the base two unless otherwise specified.)

The average of the information supplied by x_k over all X is the comentropy of the set X and is given by

$$H(X) = - \sum_{k=1}^M P(x_k) \log P(x_k)$$

where n is the number of messages. This comentropy represents the number of binary digits that are necessary to specify a particular message of the message set. In this discussion lower case letters denote individual messages and capital letters denote sets of messages.

In many communications systems, the assumption that successive messages are statistically independent is far from valid. Suppose that the probability of a particular message depends only upon the preceding message, then the conditional information supplied by the message x_k is given by $I(x_k | y_i^{(1)}) = -\log P(x_k | y_i^{(1)})$ where $y_i^{(1)}$ is the message preceding the message x_k . A conditional comentropy is defined as the average of the conditional information over the set and is given by

$$H(X|Y^{(1)}) = - \sum_X \sum_Y P(x_m; y_i^{(1)}) \log P(x_k|y_i^{(1)})$$

This represents the average amount of information needed to specify a message, given the preceding messages. This procedure can be extended to include as many of the preceding messages as are necessary until the statistical dependence on the remaining preceding messages is zero. The limiting value of $H(X|Y^{(1)}, Y^{(2)}, \dots, Y^{(n)})$ as n goes to infinity, where x is the message and $y_i^{(1)}, y_i^{(2)}, \dots, y_i^{(n)}$ are n preceding messages, is the information content of the complete message set. The function $H(X|Y^{(1)}, Y^{(2)}, \dots, Y^{(n)})$ is a monotonically decreasing function of n (see Appendix I). Therefore, $H(X|Y^{(1)} \dots Y^{(n)})$ for any n represents an upper bound to the information content of the message set that becomes asymptotically a better and better upper bound to the information content as n increases.

1.2 INFORMATION THEORY APPLIED TO THE TRANSMISSION OF PICTURES

The problem of the transmission of pictures has several approaches. A picture can be satisfactorily represented by a process of sampling and quantizing. The picture is first sampled in one direction by scanning successive lines. The number of lines determines the resolving power of the system in one direction, whereas the sampling rate of the resultant video waveform together with scanning speed determines the resolving power in the other direction. The question of the number of levels that are necessary to reproduce a satisfactory picture, when the intensity is quantized, has been investigated by Goodall and Gibbons (3, 4). Some number of levels between 32 and 64 is adequate, depending upon the amount of noise in the system. The trouble is that the eye can detect rather small changes in intensity in adjacent areas; hence the step effect on the original slow change in intensity is rather bothersome. This fact was suggested to the author by W. A. Youngblood, of this laboratory, who is working on the quantization of rates of change or derivatives in pictures.

In summary, we can represent a picture as a set of numbers, one for each cell. These numbers will be transmitted somehow to the receiver. Tying the problem down to information theory is simply deciding how the information represented by these numbers can be transmitted by a suitable coding scheme. When we mentioned messages, we did not specify what constitutes a message. We might consider a whole picture, a sequence of pictures, a single cell, or a group of cells as a message.

If the message unit is a picture, we must consider all possible combinations of the values of each element. In a television picture, there are about 500×500 or 250,000 elements. For a 32-level system, all combinations of values for each element would be $32^{250,000}$. Of course, most of these would be "snow" — not pictures at all; it would be impossible to catalog these probabilities.

At the other extreme, we might consider each picture element as a separate message.

This requires the measurement of only 32^1 probabilities for a 32-level system. Kretzmer (5) measured this first-order probability distribution for pictures. The results, insofar as finding an efficient coding scheme is concerned, are very disappointing. In order to achieve the ultimate goal of more efficient transmission, some of the messages must be much more probable than others. It can be shown that the entropy of a message set is greatest when the messages are equiprobable. For a particular picture, certain intensities may be more probable than others; for example, intensities in the near white region might occur frequently in a beach scene, but they would be less likely to occur in a night scene, in which near black intensities would be expected. The first-order case is dependent upon the particular picture chosen for measurement; hence an average over several pictures is likely to yield a near flat distribution and no reduction in entropy. The first-order statistics are not stationary in time, if a sequence of pictures is considered.

After the first-order case, the next most complicated one is the second-order or two-element case. Here there are two choices. The first choice considers the message to be pairs of elements. The corresponding set of probabilities that must be measured is the second-order or joint probability distribution of adjacent cells. The entropy per pair of original elements for this case is given by

$$H(X; Y) = - \sum_X \sum_Y P(x; y) \log P(x; y)$$

The second possibility considers the message to be a single element with a probability conditioned by the value of the preceding element. This is a more efficient system, since the process is continuous, whereas in the first case the first-order information is thrown away after each pair is specified. The conditional entropy is given by

$$H(X|Y) = - \sum_X \sum_Y P(x; y) \log P(x|y) = - \sum_X \sum_Y P(x; y) \log \frac{P(x; y)}{\sum_X P(x; y)}$$

The second-order joint probability distribution is all that is required for either case. The fact that $H(X; Y) \geq 2 H(Y|X)$, or that on a per element basis the conditional entropy is less than the joint entropy, is shown in Appendix I. For a 32-level system, the measurement of the second-order probability distribution involves the measurement of 32^2 or 1024 probabilities, which is not an impossible task.

We would expect, intuitively, that the second-order conditional entropy would yield a much better upper bound to the information content of the picture than the first-order entropy would. We expect adjacent cells in a picture to be much alike and this to be a characteristic that does not vary greatly between pictures.

Schreiber (6) measured the second-order probability distribution, using analog equipment at television rates. The Schreiber equipment is limited in accuracy by noise in the

extremely wideband system that was required. We offer a design of an analog equipment that works from the output of a facsimile transmitter for the purpose of measuring second-order probability distributions by digital means. Actually, the equipment is not restricted to measuring statistics of pictures and can be considered as a device for measuring the second-order probability distribution of any signal within certain limitations.

Our problem is the measurement of the second-order joint probability distribution of the quantized intensity of adjacent samples of the sampled signal by digital means. The sampling and quantization to 32 levels, as well as the generation of the video signal representing the intensities in a picture, is another part of the problem. From these data an upper bound to the information content of the picture will be obtained. It should be emphasized that the measurements apply only to the particular picture measured.

II. DESCRIPTION OF THE EQUIPMENT

Since this equipment was built specifically for measuring probability distributions of pictures, the intensity of the picture at a given point or the value of the video waveform at a given time will be referred to interchangeably in the ensuing discussion. We have shown how, by sampling and quantizing, a picture can be represented by a set of numbers. The function of this equipment is to sample and quantize the intensity of a picture and to measure the relative frequency of occurrence of all possible combinations of the 32 quantum levels in two adjacent samples.

The basic idea in measuring these probabilities for a picture is that of counting relative frequencies. Each time the desired event occurs, a pulse is generated and these pulses are counted by an electronic counter. A typical event might be that of level 25 followed by a level 24. We think that this event would occur more times in a picture than the occurrence of level 25 followed by level 1. There are 32^2 or 1024 possible combinations of the 32 levels in a picture. It is impracticable to build 1024 selection devices and the same number of counters. Sixteen counters are used with 16 adjustable selectors that permit the determination of a group of 16 of the 1024 event at one time. If the picture were perfectly reproducible from the facsimile transmitter source, there would be little difficulty. But the transmitter tends to drift — not very much during the time (8 minutes) it takes to scan one picture — and in the course of 1024 divided by 16 or 64 runs, a large problem is created. Our solution is to record the picture in sampled and quantized form on magnetic tape by means of a pulse code modulation scheme. Once the picture is in this form on the tape, little can happen to cause any lack of reproducibility.

A block diagram of the equipment used to sample, quantize, and record the picture on the magnetic tape is shown in Fig. 1. The functions of each block are discussed in this section; a detailed description of the electronic circuits with a complete set of schematic diagrams is given in Appendix A of the thesis (same title).

The source of video signal is a standard facsimile transmitter that has been modified

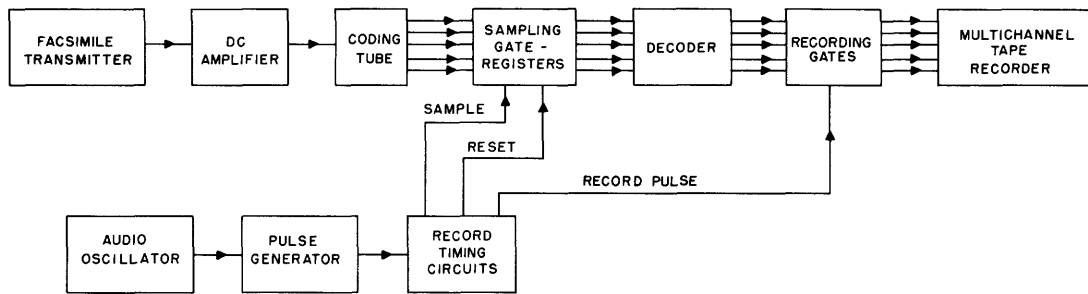


Fig. 1. Block diagram of recording equipment.

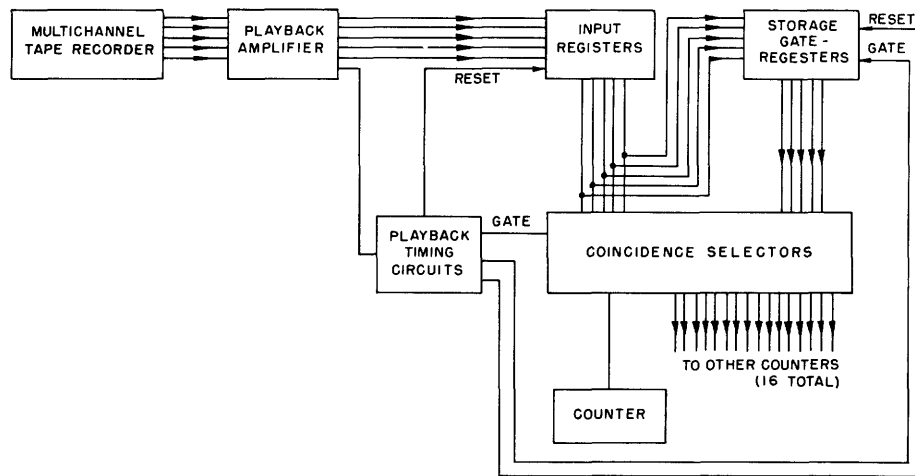


Fig. 2. Block diagram of playback equipment.

for this application. The picture is scanned by wrapping it around a revolving drum that moves slowly along its axis. Light reflected from the picture is picked up by a photomultiplier. Thus, the output of the photomultiplier is a video signal with frequency components from dc to approximately 600 cycles.

The video output of the facsimile transmitter is amplified by the dc amplifier to provide the necessary voltage and push-pull output for the analog-to-digital converter, the coding tube. The coding tube is a special experimental tube, developed at the Bell Telephone Laboratories, that was given to us through the courtesy of Mr. R. W. Sears (7). The output of the coding tube is a quantized representation of the signal input. Five of the seven binary outputs are used.

The outputs of the coding tube are connected to the sampling devices (indicated in Fig. 1 as sampling gate-registers). The outputs of the gate-registers are binary signals held at the value of the input at the sampling instant during the time between the sampling and clear pulses. These pulses are derived from the record timing circuits, which supply the other circuits with a properly delayed sequence of pulses for their operation. The outputs of the sampling gate-registers are connected to the decoder, which is a

switching device that converts the reflected binary code used in the coding tube to the conventional binary code. The output of the decoder controls the recording gates that cause pulses of current to flow through the recording heads in accordance with the binary output of the decoder. It should be emphasized that all the digit channels are parallel in operation, identically built, and independent — with the exception of the decoder.

The multichannel recorder requires some comment. The unit was converted from a conventional tape recorder to use tape that is 35 mm wide and the eight heads are arranged to record on eight parallel channels on the tape. Six heads are used in this application. Five heads are used for the five binary-signal channels and the sixth is a timing channel containing a recorded pulse for each sampling interval.

In the first part of the measurement process, the apparatus is arranged to record and run a picture signal from the facsimile transmitter through the equipment in order to record the sampled quantized signal on the magnetic tape. In the second part of the measurement process, the equipment is arranged for playback, and then the tape is played back repeatedly. The equipment counts occurrences of 16 of the 1024 possible joint occurrences, requiring a total of 64 playbacks.

A block diagram of the equipment for playback is shown in Fig. 2. The playback amplifier-limiters amplify the low-voltage signals from the tape and restore a proper waveform to the pulses played back from the tape. The input registers hold the presence or absence of a pulse from the tape for the time between the tape pulse and the reset pulse. The reset pulses, as well as pulses generated for other purposes, are generated from the played-back pulses from the timing channel of the tape recorder. This procedure keeps the variation of speed in the tape from affecting the operation of the equipment. The storage gate-registers transfer the output of the input registers to the storage registers so that the binary signals that represent the value of the previous sample are available simultaneously with the binary signals that represent the present sample for the input to the comparison device (the coincidence selector).

The coincidence selector is a device that gates out a pulse only when the desired binary signals representing the values of the signals in the present and previous samples are presented at its input from the outputs of the input register and the storage register. The particular design used here is a single unit with 16 outputs. The circuit is arranged so that any six of the ten digits representing the signal and the delayed signal can be chosen by switches in common for all 16 outputs, while all combinations of the remaining four digits are available at the 16 outputs.

The output pulses from each of the 16 outputs of the coincidence selector are counted by 16 electronic pulse counters. The counters consist of an electronic binary scale of 256, followed by an electromechanical register. These counters are of the kind used for radiation counting in nuclear science.

III. THE CODING TUBE

3.1 DESCRIPTION

The most unconventional part of the equipment is the coding tube. It is used as the analog-to-digital converter. There are several reasons why this tube offers advantages over other methods used for quantizing. Slicing circuits are often used in this application but unless the signal voltage is made so large that the steps are large enough for the operation of diodes these circuits become rather critical in operation. In any event, slicing circuits require a number of bias voltages that must be set and maintained with precision.

The coding tube, which we used, is an experimental tube supplied to us by the Bell Telephone Laboratories (7). It is essentially a cathode-ray tube with a special gun structure and a target plate instead of a screen. The gun structure focuses a horizontal line electron beam on the target plate. This line-beam is deflected up and down by the signal applied to the deflection plates. The target plate has apertures as shown in Fig. 3. Behind the apertures are current collecting electrodes, shown dotted in Fig. 3. The

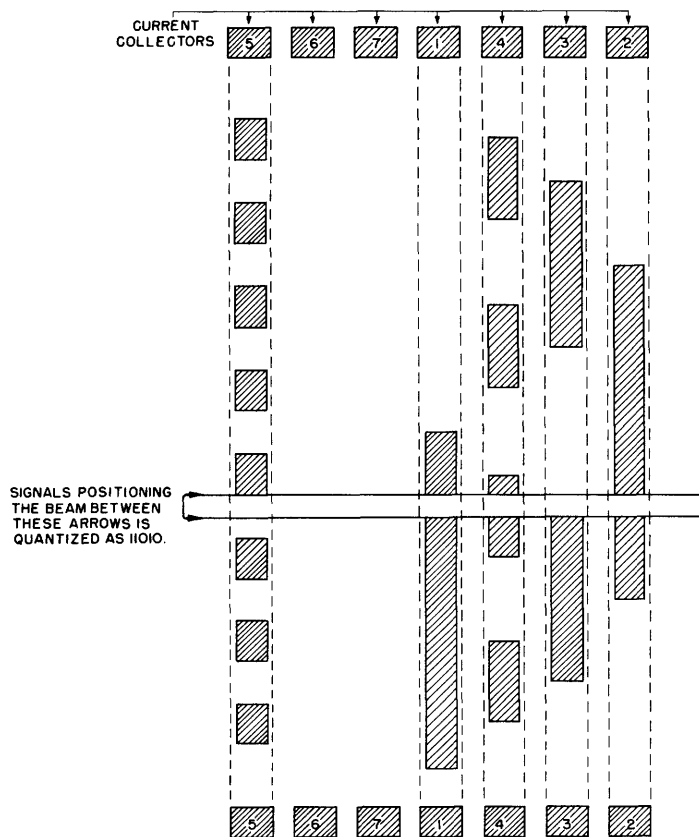


Fig. 3. Aperture plate of the coding tube.

tube has seven sets of apertures and collectors, but only the five that are used are shown in the figure. A particular collector will receive current from the beam if the signal has a value appropriate for positioning the beam so that the current falls on the aperture to the collector in question. The apertures are arranged so that the outputs of the current collectors are a binary representation of the signal.

For example, if the beam is in the region shown between the arrows in Fig. 3, current collectors 1, 2, and 4 will receive current, while collectors 3 and 5 will not receive current. Thus, the binary representation for the signal, if it positions the beam in this region, is 11010.

Although the tube has seven sets of apertures and current collectors, only the first five (heaviest weighted) outputs are needed to yield the required 2^5 or 32 quantum levels.

A second set of deflection plates labeled "tilt" is incorporated in the tube. These are nonparallel deflection plates and serve to compensate for any tilting of the aperture plate with respect to the line-beam in the manufacture. The coding tube and associated circuits are shown in Fig. 4. The output current from the collector digits is rather small (approximately 7-14 μ amp). Single-stage dc amplifiers are used to provide an output swing from -30 to +10 volts. A special power supply for the beam current and other electrodes was built. The circuit that supplies the high-voltage section uses a single series regulator of special design for the high voltages involved.

One precaution in using the tube that was not observed at first was magnetic shielding of the coding tube. The filament transformers on the coding tube chassis caused a 60-cycle deflection of the beam. A mumetal shield was made and completely eliminated the deflection from this source.

The apertures on the aperture plate are so arranged that as the beam is moved vertically by the signal, only one digit changes at a time. The arrangement shown yields the so-called reflected binary code in the digital output. The reflected code is used to minimize the errors in the output caused by tilting of the beam.

The base connections and operating conditions for the coding tube were supplied by the Bell Telephone Laboratories (8); they are shown in Table I. The recommendations were for the 1608 Flash Coding Tube; 7-digit, Experimental Model UK86 (not standard).

3.2 OPERATION

There are two primary factors that prevent ideal operation of the coding tube; focus and tilt. The beam cannot be perfectly focused; hence its width may be of the order of the physical separation of some of the apertures on the aperture plate. The focus is adjusted by looking at the output of one digit on an oscilloscope when a saw-tooth voltage has been applied to the deflection plates, and adjusting for the sharpest transitions in the observed waveform. The plot in Fig. 5 shows the actual current from the fifth-digit collector as a function of applied voltage to the deflection plates over a small region. If the beam had zero width, the transitions would be very steep, but with a finite width a linear

Table I. Coding Tube Specifications.

<u>Base Connections</u>	<u>Potential with Respect to Cathode</u>
1 and 14: Heater, 4.7 volts	90 volts maximum
2 Cathode	0
3 Mask	(60 volts with respect to A_2)
4 Anode #1	Approximately 525
5 NC	(Adjust for focus)
6 Anode #2 } connected	1000 volts
7 G_2 } see Note 1	1000 volts
8 NC	
9 Cathode	
10 NC	
11 NC	
12 NC	
13 NC	
 <u>Side Bulb Connections</u>	
Control Grid	Negative (never 0) or see Note 2
Tilt Plates	Adjust for alignment Note 3
Deflection Plates	Balanced signal input Note 4
 <u>Target Connections</u>	
Output digit electrodes #1 through #7	90 volts
Metal Envelope	180 volts
Aperture Plate	90 volts
Shield	0 volts
 Note 1. Normally operated with 2 nd anode grounded and cathode at negative potential.	
Note 2. Adjust grid bias for a total cathode current of 2.0 ma maximum (-37 for Model UK86 when $A_2 = 1000$ volts).	
Note 3. Apply dc potentials balanced to ground (2 nd anode) and adjust for fine adjustment of alignment between line focus and aperture plate.	
Note 4. Apply dc centering potentials and signal input balanced to ground. A deflection potential difference of approximately 105 volts is required to deflect the beam from the center to the end of the code plate.	

transition should be observed, assuming that the current density is constant over the width of the beam. An estimate of the effective beamwidth can be made from this plot. The aperture plate is about two inches long, requiring approximately 170 volts deflection voltage from end to end. The width of the beam requires about 6.7 volts deflection; therefore the beam is about $6.7/170 \times 2$ inch = 0.079 inch, which is not very good.

The digit amplifiers are designed to operate the sampling gate-registers at the half-current point, so that the "on" time will equal the "off" time. If this were not so, the quantum steps would not be spaced equally. Since the output currents of the various digit collectors are unequal, and the tolerance on the output dividing resistors is only 5 per cent, the load resistors for each digit are individually adjusted. The values of R_d (not shown in Fig. 4) are: digits 1 and 2, 91K; digits 3 and 4, 82K; digit 5, 240K.

By designing for enough gain, the sampling gate can be made to operate over a very small voltage range; therefore, over a small current range, the effect of poor focus is minimized.

The tube is supplied with a pair of nonparallel deflection plates which are supposed to correct tilt, but also cause deflection of the beam. The tilt control is adjusted so that the proper sequence of outputs with equal-voltage spacing is obtained when the signal is varied manually.

To facilitate calibration and adjustment of the coding tube, dc voltages were applied to the signal-deflection plates. The tilt voltage was not measured directly but by the amount of signal voltage necessary to center the beam. Because of misalignment and errors in the tube, perfect quantization could not be achieved. The graph shown in Fig. 6 shows the transfer function of the coding tube, that is, the transition points between quantum levels as a function of applied voltage. The tilt control was set to optimize the operation.

Actually, the variation in the size of quantum steps is not too much of a problem once it has been accurately determined. If each count of joint event is normalized by dividing by the product of the size of the two levels involved, not too great an error will result. Essentially, we desire to approximate a continuous probability density function by a staircase function. The value of the staircase function is the average value of the density function. If the steps are unequal in width but the discrete probabilities are normalized as above, then some of the steps in the staircase represent an average over a wider range of the continuous function than others.

The operational speed of the coding tube is limited by the digit-output circuits. The requirement on the speed in this application is that at a maximum rate of change, i. e. , a full rise from black to white or fall from white to black, the output fifth digit will swing enough to operate the gate-register into which it works. There are variations in the voltage required to operate the various gate-register units, but at full speed the swing is enough to operate any one. However, this variation between channels will cause a phase difference, equivalent in effect to a slight tilting of the beam. This is not troublesome because the reflected code makes an output error of only one or two levels in the

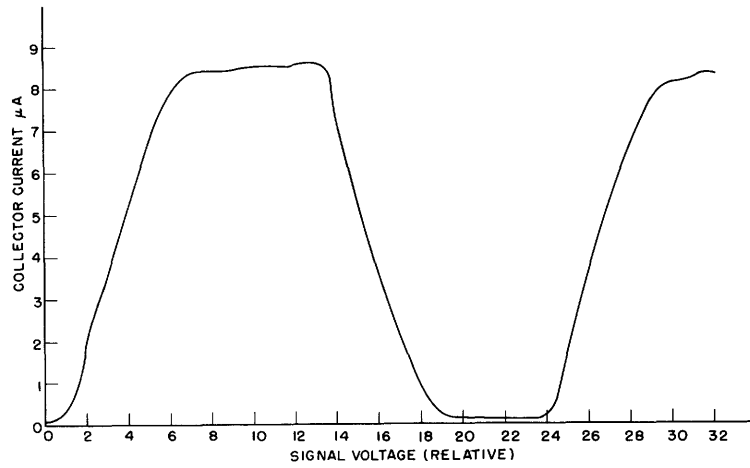


Fig. 5. Current from fifth digit versus signal voltage.

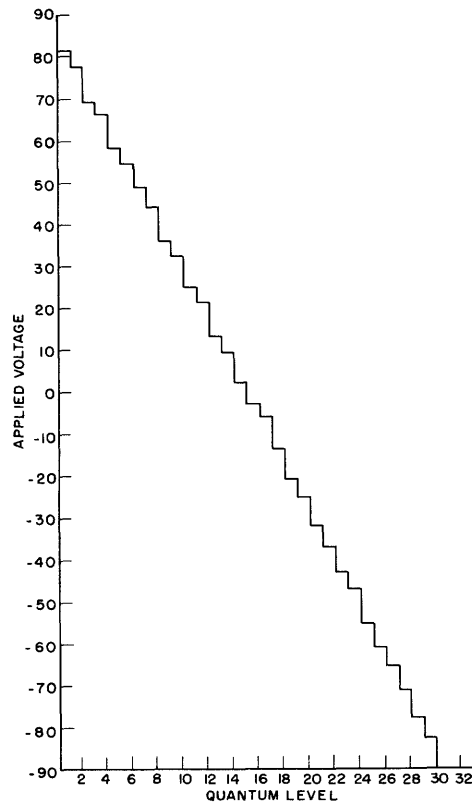


Fig. 6. Transfer characteristics of the coding tube.

representation of the signal after the most rapid input change, which is corrected in the next sample.

IV. TESTS AND EXPERIMENTAL RESULTS

4.1 TEST PROCEDURES

Because of the large number of tubes involved, a person operating the equipment should be very careful to make periodic checks to ensure that time is not wasted taking incorrect data. It is preferable to make rapid checks both before and after taking data.

A good check of the record system can be made by varying the voltage on the coding tube manually and observing the absence or presence of pulses from each channel at the connection to the tape-recorder heads. This should be done for each of the 32 levels.

The playback amplifiers can be checked for output from a tape recorded with pulses on all channels (11111); all the indicator lights on the input registers should then be on. The entire playback system can then be checked by noting the counts received in the counters. The number of counts in the counters other than the 1111 counter is a measure of the number of pulses "lost" from the tape. Typical results from this test are given below.

The coincidence selector and the counter operation can be checked from the tape recorded with (11111) by introducing zeros in any of the channels, which is accomplished by removing the plugs that connect the phase inverters to the playback amplifiers.

Accuracy of the counters can be evaluated by running a tape recorded with a saw-tooth signal through the system when it is set up to measure first-order probabilities. The results should be compared with results previously obtained from this kind of test, since the quantum levels are not equally spaced.

In case of trouble with the counters, it is useful to feed in pulses at a low repetition rate, say 40 cycles, so that individual stage operation can be observed from the flashing of the indicator lights.

4.2 TEST SIGNAL RESULTS

For testing and experimental purposes the equipment is arranged to run one half of a conditional probability, i. e., probability of the odd or even levels of the delayed sample, given the first sample. The first test was for the purpose of seeing how reliably the tape playback system operated. Before the rollers were adjusted to push the tape against the recorder heads, a considerable amount of trouble was encountered from the fluttering in amplitude of the signal from the heads, the signal completely disappearing at times. The playback amplifiers were designed to amplitude limit the signal pulses, but this action causes the amplitude differences to become phase differences. The effect

Table II. Data from Measurement of Probability of Saw-Tooth Wave.

Counter	Fifth Digit Zero				Correction Factor
	1	2	3	4	
1	85	85.5	88	85.5	85.5
2	59	58	59.5	57	58
3	45.5	42.5	43	43	43
4	65	61.5	63.5	63	63
5	53	53	52.5	52.5	53
6	66	66	66	65.5	66
7	59.5	61.5	61	61	61
8	56.5	55	54	56.5	55
9	54.5	53	53	53	53
10	60	60	64	62.5	62
11	56.5	48.5	52	49.5	53
12	69	62	65	77	65
13	62.5	65	64.5	64.5	64.5
14	68.5	66.5	69	68	68
15	59.5	62	60.5	61.5	60.5
16	46	32.5	40	42	42

Fifth Digit One					
1	27	27.5	27	27	27
2	35.5	35.5	35.5	35.5	35.5
3	39	30	29.5	29.5	29.5
4	47.5	47	47	46	47
5	23.5	23.5	23.5	23.5	23.5
6	38.5	38.5	38.5	38.5	38.5
7	22.5	22.5	22.5	22.5	22.5
8	42.5	42	42.5	42	42.5
9	26.5	26.5	26.5	26.5	26.5
10	43.5	43	43	43.5	43.5
11	37	37.5	37	31	37.5
12	60	58.5	58.5	59.5	59.5
13	28.5	28.5	28.5	28.5	28.5
14	53.5	53	53	53	53
15	34.5	34.5	34.5	34.5	34.5
16	47.5	44.5	42.5	37.5	42.5

Note: Numbers are counter readings.

of the phase differences between the channels was minimized by proper adjustment of the delay times in the playback timing circuits.

Two kinds of error can occur in the tape playback system. A pulse may be lost from one or more channels, that is, "ones" can be changed to "zeros"; or a pulse may be introduced into the system when no pulse was intended, that is, "zeros" may be changed to "ones". The loss of pulses in a channel may be caused by poor contact of the tape with the heads on record or playback which causes too great a phase difference between the outputs of this channel and of the timing channel. This, in turn, may cause the pulse to disappear entirely. Extraneous pulses may be clipped from hum in the playback amplifiers, or ineffective erasing of the tape may cause previously recorded pulses to be played back.

The system is well suited for direct test of both sources of error. A tape recorded with pulses on all channels is used to detect lost pulses. Four of the five binary channels can be tested at a time. The switches on the coincidence selector are set in their neutral positions, so that one of the counters counts the occurrence of pulses in channels 1-4, while the other counters count all the combinations of missing pulses in these four channels. Although the pulses occur simultaneously from all the channels under test, they will be referred to as a sequence.

The results of this test are very good. With new tape, one run of 250,000 correct sequences was made without error. The worst results were obtained from old, spliced tape; as many as 17 incorrect sequences in 250,000 correct sequences were counted.

The test for extraneous pulses was made in the same manner with the exception that tape was recorded with all "ones", erased once, and recorded with pulses on the timing channel only. Several runs were then made with new tape, indicating that there were no extraneous pulses in 250,000 correct sequences. If the timing pulse is absent, the system does not count the event occurring at that time.

The final test of the system was made by applying an electric signal of known waveform to the coding tube. A saw tooth was chosen because of the simplicity of its first-order quantized probability function, and its availability. The sweep from a Techtronics oscilloscope was used for the test signal. The period was set at 100,000 microseconds. The results from this test were variable. Four playback runs on the same recording are shown in Table II. The best explanation of the variation in the results is that the counters do not operate perfectly. Examination of the results shows that, if any three runs are considered and if the best two values are taken and the results averaged, not too many errors show up.

The corrections obtained from the static test of the coding tube do not make the numbers equal, as would be expected. Since the saw-tooth test is a dynamic test which includes the whole system, the correction factors used for the measurements on pictures will be those given in the last column in Table II.

4.3 MEASUREMENTS ON A PICTURE

Because of the occasional variation in counts in identical runs, at least two runs were made for each setting of the coincidence selector. If these two did not agree, additional runs were made until several of them agreed. This method was time-consuming, but it was felt that relatively accurate data on one picture were better than unreliable data on several different pictures. The picture used for measurement is shown as Fig. 7.

Levels 31 and 0 have little meaning, since the coding tube quantizes any signal larger than the one for the 30-31 transition as 31 and any signal smaller than the one for the 0-1 transition as 0. The bar that holds the picture on the transmitter was covered with black tape; therefore it was quantized blacker than anything in the picture, and as level 0, which was not counted. The facsimile transmitter was adjusted to give level 1 on the blackest part of the picture and level 30 on the whitest part.

Several calculations must be made to derive the probabilities from the raw data received from the equipment. First, all the numbers were converted from the binary count plus register differences to decimal numbers. Then, the correction factors shown in Table II were applied to each of these numbers. Each corrected number was divided by the sum of all the corrected numbers to obtain the actual probabilities that are shown in Table III. The data show that the probability of equal levels in adjacent cells is high. Many combinations of levels occur with nonzero probabilities, a result that would not



Fig. 7. Picture used for measurement.

necessarily be expected.

The entropy calculation was made directly from the probabilities. The fact that $H(X|Y) = H(X; Y) - H(X)$ was used to compute $H(X|Y)$. $H(X; Y)$ is the sum of $-P(x_{ij}) \log P(x_{ij})$ for all 900 probabilities. Tables of $P(X) \log P(X)$ for logarithms to the base two are given by Dolansky (9). $H(X)$ is $-P(x_i) \log P(x_i)$ of the column sums in Table III. The calculated values of entropy for the picture that was measured were: $H(X; Y) = 6.01$ bits, $H(X) = 4.77$ bits, and $H(X|Y) = 1.24$ bits. If all the 900 probabilities were equal, the value of $H(X|Y) = H(X)$ would be 4.907 bits. This saving in the average amount of information is true only for this picture. Measurements on other pictures will give a larger or smaller value depending upon the complexity of the picture. As was expected (see Sec. I) the value of $H(X|Y)$ was less than one half $H(X; Y)$.

V. CONCLUSION

Although the potential accuracy of measurements from this equipment is very great, repeatable results were difficult to obtain. The primary difficulty with this equipment is the large number of tubes involved and the servicing problems they create. The counters are a continuing source of trouble. A single unsoldered lead or weak tube in one of the counters can cause extraneous counting and lead to serious errors. The author feels that by making several runs for each set of levels, until repeatable runs are obtained, meaningful results can be achieved. The tremendous range in the measured probabilities is not obscured by relatively small errors in the individual probabilities.

Suggestions for further work include trying to make the results more repeatable so that time spent in taking the measurements could be cut down. The data taking for all 900 probabilities with three runs each takes more than thirty hours, working rapidly and barring breakdown of the equipment. The use of magnetic cores is being investigated for these applications at the Research Laboratory of Electronics; greater reliability might be obtained by redesigning the equipment so that the number of tubes is reduced.

From the point of view of information theory, the figures of entropy quoted do not seem to indicate that second-order probability measurements constitute the most fruitful approach to the whole problem of transmitting pictures. As mentioned in Section I, an investigation of the use and entropy of quantized derivatives of the picture signals is in progress. Which of the many possible approaches is the best one, will not be known until some measurements from each one are compared with previous estimates of the necessary reduction in channel capacity. One feels that potentially there is a great saving to be made, if one could find the right statistics to measure and could design a corresponding coding scheme.

Acknowledgment

The author wishes to thank Prof. Peter Elias for his help and suggestions, both theoretical and practical, and Prof. R. M. Fano for his interest in the project. Mr. W. A. Youngblood helped with the facsimile transmitter and associated circuits; Mr. A. J. Osborne and Mr. F. Tung with taking the data. Mr. Tung also helped with the calculations. Other members of the laboratory staff contributed helpful suggestions.

APPENDIX I. PROOFS

1. It is desired to show that: $H(X|Y^{(1)}, Y^{(2)}, \dots, Y^{(n)}) \geq H(X|Y^{(1)}, Y^{(2)}, \dots, Y^{(n-r)})$.
 First consider the difference: $D = H(X|Y^{(1)}, Y^{(2)}, \dots, Y^{(n)}) - H(X|Y^{(1)}, Y^{(2)}, \dots, Y^{(n-r)})$.
 From the definition of the entropies involved we obtain the relation

$$D = - \sum_{\mathbf{X}} \sum_{Y^{(1)}} \sum_{Y^{(2)}} \dots \sum_{Y^{(n)}} \times \left[P(x, y^{(1)}, y^{(2)}, \dots, y^{(n)}) \log \frac{P(x|y^{(1)}, y^{(2)}, \dots, y^{(n)})}{P(x|y^{(1)}, y^{(2)}, \dots, y^{(n-r)})} \right]$$

by combining logarithms and summations. If we let

$$w = \frac{P(x|y^{(1)}, y^{(2)}, \dots, y^{(n)})}{P(x|y^{(1)}, y^{(2)}, \dots, y^{(n-r)})}$$

and use the inequality that $\ln w = \log w \times \log e \leq w - 1$, we obtain the following relation:

$$D \leq \sum_{\mathbf{X}} \sum_{Y^{(1)}} \sum_{Y^{(2)}} \dots \sum_{Y^{(n)}} P(x, y^{(1)}, y^{(2)}, \dots, y^{(n)}) \times \left[\frac{P(x|y^{(1)}, y^{(2)}, \dots, y^{(n)})}{P(x|y^{(1)}, y^{(2)}, \dots, y^{(n-r)})} - 1 \right] \frac{1}{\log e}$$

The first term in the square brackets multiplied by $P(x, y^{(1)}, y^{(2)}, \dots, y^{(n)})$ and summed over $X, Y^{(1)}, Y^{(2)}, \dots, Y^{(n)}$ is unity, since the summation is over all the random variables. In the same way the second term in the square brackets multiplied by $P(x, y^{(1)}, y^{(2)}, \dots, y^{(n)})$ and summed over $X, Y^{(1)}, Y^{(2)}, \dots, Y^{(n)}$ is unity. Hence the difference D is zero, and we obtain the desired relation, $H(X|Y^{(1)}, Y^{(2)}, \dots, Y^{(n)}) \geq H(X|Y^{(1)}, Y^{(2)}, \dots, Y^{(n-r)})$. This derivation was given by Fano (2) in a slightly different form.

2. It is desired to show that: $H(X, Y) \geq 2H(X|Y)$ for the case of picture elements. That $H(X, Y) = H(X) + H(Y|X)$ can be seen from the definition of the comentropies involved. Starting with the definition of $H(X, Y)$, we obtain

$$\begin{aligned}
H(X, Y) &= - \sum_X \sum_Y P(x, y) \log P(x, y) = - \sum_X \sum_Y P(x, y) \log \frac{P(x) P(x, y)}{P(x)} \\
&= - \sum_X \sum_Y P(x, y) \log P(x) - \sum_X \sum_Y P(x, y) \log \frac{P(x, y)}{P(x)} \\
&= - \sum_X \sum_Y P(x, y) \log P(x) - \sum_X \sum_Y P(x, y) \log P(y|x) = H(X) + H(Y|X)
\end{aligned}$$

$H(X) = H(Y)$ because a difference of one cell does not affect the first-order distribution. The fact that $H(Y) \geq H(Y|X)$ is shown above by letting both n and r equal one. Therefore, $H(X, Y) = H(Y|X) + H(X) \geq 2H(X|Y)$.

References

1. C. E. Shannon, Communication in the presence of noise, Proc. IRE 31 (1949), pp. 10-21.
2. R. M. Fano, Statistical theory of information, Class Notes for Course 6.574, M. I. T., 1953.
3. W. W. Goodall, Television by pulse code modulation, Bell System Tech. J. 30 (1951), pp. 33-49.
4. E. H. Gibbons, The transmission of picture signals by pulse code, S. M. Thesis, Department of Electrical Engineering, M. I. T., 1950.
5. E. R. Kretzmer, Statistics of television signals, Bell System Tech. J. 31 (1952), pp. 724-750.
6. W. F. Schreiber, Probability distributions of television signals, Ph. D. Thesis, Harvard University, 1953.
7. R. W. Sears, Electron deflection tube for pulse code modulation, Bell System Tech. J. 28 (1948), pp. 44-57.
8. The specifications in Table I were received from R. W. Sears of Bell Telephone Laboratories, Incorporated, in a letter dated Jan. 8, 1954.
9. L. Dolansky and M. P. Dolansky, Table of $\log_2 \frac{1}{p}$, $p \cdot \log_2 \frac{1}{p}$ and $p \cdot \log_2 \frac{1}{p} + (1-p) \cdot \log_2 \frac{1}{1-p}$, Technical Report 227, Research Laboratory of Electronics, M. I. T., Jan. 2, 1952.

