The *Picturephone*® System:

Service Standards

By T. V. CRATER

(Manuscript received September 25, 1970)

We describe in this paper the fundamental requirements of a visual telephone service. These lead to a basic physical configuration of the station equipment. The picture standards are chosen to provide the visual adjunct at no greater cost than is necessary to secure most of the possible enhancement of direct conversation. Transmission standards are established with the objective of limiting to an acceptable range the difference in quality between the image as viewed at the originating station and as received over the longest connection possible in the network.

I. THE FUNDAMENTAL REQUIREMENTS FOR *Picturephone*® SERVICE

The previous paper[1] described *Picturephone* service as primarily designed for face-to-face conversation. Put very simply, we are interested in improving telephone communication by making it possible for the two parties to see as well as hear each other. The notion of seeing what is going on at some distant point by means of electrical signals transmitted over wires is as old as the telephone itself.[2] The early work on television was in fact directed toward a wired point-to-point visual service.

Although television developed as a broadcast service, it brought into being most of the technology necessary to make visual telephone service technically feasible. Economically, however, the techniques developed for television are in many respects inappropriate for a face-to-face service. The subscriber's station must be equipped with a camera, a receiver and voice equipment, packaged for great durability and safety, and manufactured initially in relatively modest quantities; this suggests an equipment cost several times that of a home television receiver. The subscriber needs a private two-way channel to the nearest switching machine, but short-distance cable systems used for television today are costly, especially for the initial era of very light development of the service. Finally, long-distance television transmission today, requiring usually an entire microwave radio channel and attentive maintenance, is prohibitively costly for a face-to-face service to achieve wide public acceptance.

The key aspects of the design and implementation of a *Picturephone* system therefore pivot around questions of cost. Succeeding papers of this issue will describe the means chosen, in each part of the system, to minimize initial and maintenance costs while meeting standards of service. We have set those standards so that the quality will be no better than is really needed, particularly in the more cost-sensitive aspects. In doing this we have tried to ensure that the standards chosen are fully adequate for a face-to-face service in the foreseeable future, since it may be extremely costly to upgrade them later. In the beginning, the usefulness of *Picturephone* service will be limited by its cost; eventually, it will be limited by its inadequacy relative to something else. The choice of the standards will determine the length of its era.

1.1 *The Experimental Basis for the Standards*

The subjective factors involved in establishing video standards are sufficiently complex that a comprehensive theory is not available. Thus the weighting of economic and technical factors tends at every step to require subjective testing as well. Moreover, a set of standards, once tentatively established, must be tried under realistic conditions, to test the overall effectiveness and utility of the service they define. It is consequently the author's privilege to report on the results of a series of studies, test programs and trials, extending over a period of a decade and a half, and supported by the efforts of dozens of people. In what follows, we attempt to abstract an outline in a logical sequence of the bases for the separate but interrelated choices that make up the standards.

It is not possible in many cases to refer to the literature for a description of the test or trial alluded to, since most of the work at Bell Laboratories on visual telephone standards has not yet been published. As a partial substitute, it may be helpful to review here briefly the major study and test programs which have contributed to these standards.

The work may be considered to have started in 1954 with an investigation by W. E. Kock, F. K. Becker, R. L. Miller and others of the possibilities of a visual adjunct consisting of a series of snapshots of the distant party, produced at the rate, in one realization, of one every two seconds.[3–5] Although this approach was not pursued beyond the demonstration of its technical feasibility, the work excited new interest in the notion of a visual adjunct to telephony. In 1956 A. D. Hall, with J. D. Gabbe, H. Cravis, and later M. W. Baldwin, Jr., J. E. Abate, the author and others, began an engineering study of the economic and subjective factors involved, with the objective of establishing requirements for station equipment. At about the same time W. T. Wintringham, with R. L. Eilenberger, R. L. Miller and later R. C. Brainard, F. W. Mounts, E. F. Brown and others, began experiments with methods of efficient digital encoding of the video signal. This group also investigated standards in order to characterize the signal which might be encoded. Their work has continued to influence standards at each step.

By 1960 an initial service definition was available, calling for a 0.5-MHz video signal which could be transmitted and switched at acceptable costs. A station instrument development was initiated by a group under the direction of L. A. Meacham. The resulting instrument,[6] referred to as the Mod I *Picturephone* station set, was installed in an 8-station switched network as an exhibit at the New York World's Fair in 1964.[7,8] Visitors chosen by a random-sampling procedure were given an opportunity to make calls between booths of the exhibit, and their reactions to the experience, in respect to the utility of the service and the appropriateness of various features of the set design, were elicited by means of a questionnaire. Observations were also made of the learning process as visitors used the equipment, to determine whether any human-engineering improvements were indicated.

In 1965 Mod I sets were installed in 28 Bell Laboratories offices at Murray Hill and Holmdel, New Jersey, with switching and transmission equipment of preliminary design, and with telemetry equipment to record the duration of each phase of every call and all operations of electrical controls. Questionnaires were also used to determine subjective reactions. A separate trial was conducted at offices of the Union Carbide Corporation in New York and Chicago during the same year, in which 35 stations were installed and statistics of both local and intercity traffic were recorded.[8]

On the basis of results of these trials and of continuing laboratory experiments and economic studies, standards for an improved station set, referred to as Mod II, were established in 1966. This station set, described in another paper in this issue,[9] was tested in a trial in the Westinghouse Corporation's offices in New York City and Pittsburgh in 1969. It is also in regular experimental operation in the Murray Hill-Holmdel network, now expanded to include offices of the American Telephone and Telegraph and Western Electric Companies in New

York City. With modifications resulting in part from these tests, the Mod II set has been employed in commercial service since July 1, 1970, when *Picturephone* service was instituted in the Golden Triangle area of Pittsburgh, Pennsylvania.

1.2 *Basic Picture Requirements*

In view of the high cost of television transmission, it would be prodigal to provide the picture at broadcast television standards. The prospect of using fewer scanning lines and less bandwidth is very attractive. With good maintenance of station equipment and of wire transmission systems, picture quality can be kept close to the design objective; much of the possible quality of broadcast television is lost in the radio broadcast path and in poor maintenance of the home television receiver. The reduction of bandwidth which is possible is limited by a built-in human resistance to very fuzzy images and to images which flicker. As will be seen, a minimum bandwidth of several hundred kilohertz is found to be necessary.

A large reduction of bandwidth is possible if the display of the image in continuous motion is abandoned. In the visual adjunct experiments carried out in the years 1954 through 1956 by W. E. Kock, F. K. Becker, R. L. Miller and others, the picture elements were transmitted over a separate telephone circuit and stored at the receiving end until the snapshot was complete, and then displayed while information was accumulated for the next image. This adjunct permitted inspection of the other party, but the disjointed series of facial images, each one displayed while its successor was made, did not provide the enhancement of telephone conversation which occurs with full motion portrayal.

We have chosen to give the system sufficient bandwidth to provide full motion capability, adequate, for example, for lip reading, and resolution sufficient for a life-like image of the face.

Most of the visual enhancement of telephone communication requires a monochrome image only; the elements of it are smiles, frowns, averted glances, broad grins, expressions of shock, dismay, amusement, or sympathy. There are however additional values in the naturalness of an image transmitted in color. These have not been deemed sufficient to justify delaying the service until problems of cost of the color station instrument can be resolved. They do suggest the consideration of transmission capacity in the light of the needs of a future compatible color system.

1.3 *Basic Operational Objectives*

Since *Picturephone* service is to be an extension of telephone service, it must operate in typical telephone environments. To avoid duplication of equipment, visual telephone and ordinary telephone calls should be made from the same station instrument. To simplify operation as much as possible, the station address for *Picturephone* service should be the same as for telephone service, except for a prefix to indicate that a call is to include the visual adjunct.

On the other hand, the user will not want to give up other special telephone services which he may already have. This requires that the applique equipment for the visual adjunct be compatible with any of the many types of *Touch-Tone*® telephone station instruments now in use in the Bell System.

Telephone service should not in any way be diminished when the visual channel is added, and there is one respect in which it should be expanded. It is appropriate to supply hands-free speakerphone audio with *Picturephone* service because the handset becomes a barrier in the optical path and detracts from the feeling of presence. The handset, of course, remains available for use when privacy is important or room noise is disturbing. The speakerphone, on the other hand, is available for telephone calls as well.

In use, the subscriber should be able to make either a telephone or a video telephone call to another *Picturephone* station, or a telephone call to any telephone station, using in any case the same telephone instrument. If he elects to make a visual call, the video channel is provided at the beginning and is available throughout; this makes it possible for the system to select an available video channel and check to ensure that it and the called subscriber line are functioning before making the connection. It will not be possible to summon the video adjunct midway through a telephone call or drop it midway through a video call. Since any telephone line may become a *Picturephone* line, it may have telephone-only extensions; visual calls may be originated or received on these, although the picture will not be seen.

We have discussed only the use of the service for face-to-face conversation. It is also suitable for the transmission of graphic information, such as pencil sketches, diagrams, pictures and some printed material. The video display and broadband channel are ideal for interaction with a computer. The network of broadband channels will be useful for the transmission of data at very high speeds. These applications are described in other papers in this issue. The standards, however, are based on the face-to-face application, and only such minor modifications have been permitted on behalf of other applications as will not significantly affect the cost of the face-to-face use.

These elementary considerations lead to a concept of a station instrument which can be added to the existing telephone set, to operate as simply and with as little disruption of the normal telephone environment as possible. No more transmission capacity need be provided than is necessary to achieve the enhancement which vision adds to face-to-face conversation, without visual strain, nervous strain or discomfort.

II. CONFIGURATION AND PHYSICAL DIMENSIONS OF THE STATION EQUIPMENT

In setting standards to obtain fully adequate visual images at minimum cost, video parameters such as number of scanning lines, resolution, and the allowable degree of picture impairment of each kind are of primary importance because they determine the costs of transmission systems. These requirements must be evaluated in terms of a rather specific station configuration, with particular respect to viewing distance, picture size, and picture aspect ratio.

To continue with the discussion of standards, we therefore look next at the station configuration. Given the initial requirement that the equipment be designed for use at the close quarters provided by a desk or table, there are some constraints that must be accommodated.

2.1 *Imaging the Viewer*

For the parties to the conversation to enjoy a normal visual exchange, they should each remain reliably in view of the other, and they should be able to "look each other in the eye." The first of these needs might be satisfied by a camera arranged to follow the user so as to keep him centered in the field of view. This is presently considered unattractive on technical, economic, esthetic and psychological grounds. Some means is therefore desirable to help the user stay in view. The second ideally requires that the camera be located in effect where the screen is, at about the bridge of the nose of the image.

Both of these needs are quite well satisfied by putting the camera behind a half-silvered mirror which reflects the image of the display tube so it can be seen by the user.[10] The result is to box in the optical path to the screen so that it appears to be at the far end of a short tunnel. In order to see the entire image, the user stays within the camera field. Eye contact is also very good. A schematic diagram of such a station set is shown in Fig. 1; an instrument using this principle was constructed and tested at Bell Laboratories in 1963 by R. L. Eilenberger.

This arrangement makes it quite difficult for others in the room to see the image. While the resulting privacy is an advantage in some cases, it is a source of frustration when the user wants to introduce a second person to the distant party, or when he wants to demonstrate the service to others. The mirror also requires an increase in bulk. For these reasons the simpler configuration shown in Fig. 2 is preferred for general use. The viewer looks directly at the screen and the camera is located at as small an angle from the screen as possible. The eye contact requirement remains a factor in dimensioning the instrument; this question is discussed further in the paper on the station set in this issue. The eventual development of a station set using the split optical path is by no means ruled out by the standards we have chosen, and such a set would have marked advantages for the user who wants to exclude the distractions of a busy environment.

2.2 *The Camera Field of View*

Somewhat related to the choice of the open screen is the question of the field of view. Since the major visual clues in conversation are facial expression and movements of the head, eyes, and lips, the minimum bandwidth would be required for an image of the head only. With the open screen, the conscious effort required of the user to stay in so constrained a position would make him quite uncomfortable. Experiments conducted by J. D. Gabbe and others in 1956 and thereafter and by R. L. Eilenberger and others about the same time,

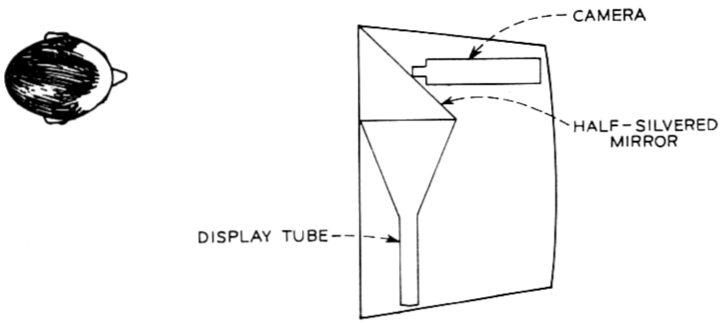

Fig. 1—Station set with coaxial display and camera optical paths—top view.

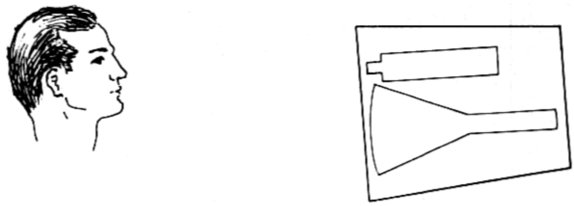

Fig. 2—Station set with open display.

quickly established the desirability of the head-and-shoulders view with the open-screen configuration. This view not only permits a necessary degree of freedom of motion, but also allows the distant party to see some of the surrounding environment, shows additional visual signals such as hand and arm motions and shrugs, and provides an esthetically pleasing picture. Experience with the Mod I set at the New York World's Fair in 1964 confirmed the advantages of the head-and-shoulders view.

The substantial savings in transmission cost might justify the head-only picture in the early years of the service when costs are paramount. In the long term, however, the provision of the head-and-shoulders view is considered necessary to assure the full adequacy of the service for normal face-to-face conversation.

2.3 *Picture Size and Viewing Distance*

The ratio of viewing distance to screen size is closely related to the number of scanning lines and to the horizontal resolution, and in fact to many other picture quality standards. This is because the user wants to be close enough to see all of the useful picture detail but far enough away to be undisturbed by the line structure and other visual effects due to the scanning process. In the visual telephone application viewing distance tends to be the independent variable, leaving picture size and scanning standards to be chosen in the light of the cost of bandwidth in transmission.

Since *Picturephone* service must operate in the relatively small spaces in which the telephone is used, the available viewing distance is limited. For a desk of ordinary size, a distance greater than about 0.9 meter is awkward. Even this is too much for many telephone situations. On the other hand, shorter viewing distances tend to degrade the image obtained by the camera, which must be located as near the display as practicable. Short camera ranges tend to increase the eye contact angle and to introduce distortion effects due to perspective. A short camera range also deprives the user of leeway for his normal forward and backward movements. A user who leans forward 0.3 meter from a normal range of 0.9 meter may still transmit an acceptable picture, but the effect of moving from 0.6 meter to 0.3 meter is found to be almost grotesque.

A longer camera range also reduces the height at which the camera must be located, and therefore leads to a more compact set. Since the head is to appear near the center of the picture, the camera must either be located near the level of the head or be tilted upward. Tilting the camera tends to bring overhead lights into the picture, and if carried to extremes, makes walls and bookcases appear to be leaning backward and produces a distorted view of the face. The further away the camera is, the lower it can be placed and still remain within a given tilt-angle limitation.

Because of these factors, confirmed by experience with the Mod I set, a viewing distance of 0.92 meter (36 inches) has been chosen, to place the instrument as far away as possible while keeping it on the desk or table at which the user is sitting. This standard has been retained through the several trials of the two station set models.

The choice of picture size is somewhat more complicated. A primary factor is the desirability of a compact subscriber's set. However, if the user's esthetic or psychological reaction is that the picture as seen at the design viewing distance is "too small," he may tend to compensate by moving forward. In experiments at Bell Laboratories in 1958, R. L. Eilenberger found that at a viewing distance of only 0.66 meter, a picture of about 0.013 square meter (20 square inches) was rejected by test subjects in favor of larger pictures.

An upper limit to the size is imposed by the fact that the desk is also used for other activities. It appears however that esthetic preferences restrict picture size more than desk space might. Mr. Eilenberger found preferred picture heights ranging from 0.14 meter (5.5 inches) to 0.16 meter (6.2 inches) at viewing distances from 0.66 meter (26 inches) to 1.07 meter (42 inches). His tests used a 525-line picture and the height preferences obtained are therefore based on esthetic preferences rather than raster visibility. In tests of pictures with limited bandwidth made at Bell Laboratories, J. D. Gabbe and A. C. Bandini found that with a 245-line picture of 0.20-meter (8-inch) height, subjects considered a 1-meter viewing distance too close, while with a 0.15-meter (5¾ inch) picture of the same number of lines, the 1-meter distance was entirely satisfactory.

Optimizing the picture at minimum bandwidth leads to a similar picture height. The picture height should be chosen to make the transmitted information fully available to a person with normal vision. The viewer tends to approach the picture until the scanning structure becomes sufficiently obtrusive that there is no advantage in moving closer. If the distance at which this happens is 0.92 meter (36 inches), the optimum viewing range will coincide with the chosen camera range. It will be shown in Section III that this leads to a picture height of about 0.13 meter (5 inches).

The choice of picture height cannot be entirely divorced from the choice of aspect ratio. Studies of preferred aspect ratio for a head-and-shoulders view, based upon viewing the picture, were not in complete agreement, a 3:4 aspect ratio (width to height) being found in one study and about 1:1 in another. The Mod I station set used the 3:4 ratio, with a 0.11-meter (4-3/8 inch) by 0.15-meter (5-3/4 inch) picture. Trials of this equipment at the New York World's Fair in 1965, at the Union Carbide Corporation in New York and Chicago in 1965, and at Bell Laboratories branches in Murray Hill and Holmdel, New Jersey, in 1965 and later, called attention to the user's problem of staying within the camera field of view. A choice of the 4:3 aspect ratio of entertainment television would solve this problem quite completely, but the bandwidth would then be increased by a factor of 16/9 for a given resolution of the actual user image.

These factors were compromised in the Mod II set by the choice of the 1.1:1 aspect ratio; this frame is more consistent with test results on optimum aspect ratio for viewing, provides adequate freedom for the user and is economical of bandwidth. At the same time, the picture height was reduced from 0.15 meter to the 0.13 meter mentioned above. This reduced the visibility of the line structure. It also reduced the increase in the contribution of the display tube to bulk, from 47 percent to 10 percent, and permitted reduction of the eye contact angle. The Westinghouse and Bell Laboratories trial results with the Mod II design indicate that these dimensions are entirely satisfactory.

2.4 *Audio and Controls*

In a video telephone call, the telephone handset is not only a visual obstruction, concealing important parts of the face, but also burdens the hands, interfering with normal gesturing and the manipulation of objects to be displayed. For this reason a microphone and speaker are considered essential for every station. The handset must also be available for those occasions when privacy or background noise makes the speakerphone arrangement undesirable.

Provision for viewing oneself has been found very desirable. Use of this feature is typically high in the first few months as a new user, not yet confident in the use of other visual cues, checks frequently to see whether he is properly framed in the picture. To secure proper framing, it must be possible for the user to adjust both the azimuth and elevation of the camera. Provision for shutting off the outgoing picture is necessary for obvious reasons; it is desirable to provide an electronic means for this rather than require the user to obstruct the camera view. A means of alerting the user, while the station is being rung, that the call is a *Picturephone* call, is also desirable.

These features imply a number of controls which is close to the minimum. Their realization is described in the paper on the station set in this issue.[9]

III. OPTIMIZATION OF PICTURE PARAMETERS

The choices to be made in selecting picture standards are so numerous, and interact in so many ways with each other and with the cost of providing the service, that it is difficult even to propose a logical study sequence which would lead to a unique set of standards. We have already seen that the picture size cannot be established without some reference to the scanning standards. In what follows, the bases of the major choices will be outlined.

3.1 *Brightness and Field Rate*

In a *sequential* scanning pattern, the picture is scanned line-by-line from top to bottom. The resulting field of horizontal lines contains all of the information to be transmitted about one frame. In a *two-to-one interlaced* pattern, the picture is scanned twice, the second time between the first set of scanning lines, so that two *fields* are required to complete one *frame*. With either pattern, the gross visual effect of a field is that of a moving pulse of light, decaying in accordance with the characteristics of the phosphor while the scanning beam traverses the picture. If the field rate is too low the picture appears to flicker. The lowest frequency at which flicker disappears depends on several factors but varies approximately as the logarithm of the highlight luminance. Under typical television conditions the threshold of perception of flicker at 50 fields per second occurs at a highlight brightness of about 30 foot-lamberts, while at 60 fields per second it occurs at 180 foot-lamberts.[11] In tests at Bell Laboratories in 1966, E. F. Brown found a highlight brightness of about 80 foot-lamberts to be preferred for *Picturephone* service. This would permit a field rate of

56 Hz under typical television conditions, but to allow for variable circumstances it is desirable to establish a somewhat higher field rate.

Interference from power lines at 60 Hz and its harmonics has usually been a factor in the choice of field rate for any video service. Interference at a frequency differing from the field frequency by 10 Hz requires about 12 dB more suppression than one differing by only 0.5 Hz.[12] Recent tests at Bell Laboratories by D. B. Robinson, Jr., show that if the two frequencies are very nearly equal, the suppression required may be as much as 20 dB less than at 0.5-Hz difference. With the advent of solid-state electronics and the development of improved clamping circuits, however, these results are of less significance than formerly.

Both fluorescent and incandescent lights have a fluctuating component at twice the power frequency. This interacts with the camera to produce a flicker if the power and field frequencies are different. Means of mitigating this effect have not been investigated, since flicker at the field rate provides sufficient motivation to retain a field rate near 60 Hz.

To obtain 250 active lines per frame a line repetition rate somewhat larger than 30 × 250 is required because of the need for vertical retrace time. The resulting line repetition frequency is in the neighborhood of 8 kHz. There is no reason why it should not be made exactly 8 kHz, and there may be some future advantage in doing so, in the digital transmission plant, for example, where the sampling frequency for voice signals is 8 kHz. Dividing 8000 by 30 gives a number close to 267 for the total lines per frame, including those lost in blanking. To make it exactly 267 (the odd number is necessary for interlacing), the frame rate has been made 29.9625 Hz.

3.2 *Interlaced Versus Sequential Scanning*

Although 60 fields per second are necessary to eliminate flicker, a much lower rate is sufficient for motion. Indeed, under ordinary conditions of room lighting and screen brightness, 30 pictures per second are indistinguishable from 60 if flicker is suppressed, for example, by displaying each picture twice.[13] Two-to-one interlace is therefore suggested. By this means the bandwidth required for a given horizontal resolution, vertical resolution, and field rate is reduced by half. However, the bandwidth saving in this case comes at the cost of a loss in overall subjective quality, because interlace introduces some undesirable visual effects. These are more noticeable in *Picturephone* service than in broadcast television because the angular subtense of the spacing between lines is larger in the former at normal viewing range.

If all but two of the raster lines are masked, only one is scanned in each field, so that a single line may be seen jumping back and forth at a 30-Hz rate. This effect, called interline flicker, occurs at horizontal brightness boundaries in the picture material and in small areas of high brightness.

If an object in the picture moves up or down at a rate such as to pass one scanning line every sixtieth of a second, the raster appears to break down into the lines of a single field, moving up or down at the same rate. The effect, called subjective line pairing, may be quite striking. In *Picturephone* service, object motion at just the right rate is unusual, although the effect can be invoked voluntarily at close viewing ranges by scanning the screen slowly upward or downward. Momentary eye motions up or down, however, can cause subjective line pairing to occur long enough that the alternate-line pattern emerges, although the apparent motion may not be seen. If the receiver is experimentally turned on its side so that the scanning is vertical, the effect is enhanced and the picture may appear to break up at almost every glance, apparently because involuntary horizontal eye movements are more frequent than vertical ones.

The net result of these effects is to give the appearance of a somewhat "busy" or noisy picture, compared to a sequentially scanned picture. It is appropriate to ask whether the interlaced picture is in fact subjectively better than a sequential picture transmitted over the same bandwidth. Test results indicate that it is, but for face-to-face *Picturephone* applications, not as much as might be expected. In a series of carefully designed experiments, E. F. Brown found that the actual bandwidth advantage was surprisingly low.[14] He used as a reference an interlaced 225-line, 0.13-meter (5-inch) square picture at various brightness levels and determined the bandwidth required for a subjectively equivalent sequentially scanned picture at the same brightness levels. The bandwidth reduction with interlace ranged from 37 percent at 40 foot-lamberts of highlight luminance to only 6 percent at 100 foot-lamberts. At the preferred highlight luminance of 80 foot-lamberts, the reduction under the conditions of the experiment was 16 percent.

Although the advantage is unexpectedly small, it is nonetheless desirable to accept it. As described in the previous paper in this issue, the video transmission from subscriber to central office utilizes telephone wire pairs with equalized amplifiers at regular intervals. The cost of this link depends largely on the number of amplifiers required, which in turn is closely related to the bandwidth. The possible advantage in sequential scanning is in long distance transmission, where future digital encoding systems may take advantage of the similarity between successive frames to economize on bit rate. These may require less storage with sequential scanning. However, the cost savings in trunks fall short of the additional cost in subscriber loops. We have therefore retained the interlace.

3.3 *The Optimum Raster with Interlace*

With interlaced scanning, a minimum of about 250 lines has been found to be necessary for adequate portrayal of the head-and-shoulders image; about six or seven lines then portray the eyes and eight to ten the lips. We want to choose the line spacing so that the user will be able to see all of the detail easily at 0.92 meter but will find the scanning structure annoying at closer range. The appropriate angular subtense, with interlaced scanning, for the distance between centers of adjacent scanning lines, is found to be about 2 minutes of arc. This leads to about 20 lines per cm at 0.9 meter, so that the 250-line picture requires a 0.13-meter height.

The Mod I station set, developed for experimental trials, was designed to hold bandwidth to a minimum. Since the picture was relatively narrow in width, the height was set at 0.15 meter ($5\frac{3}{4}$ inches) to keep the overall size adequate at 0.92 meter (36 inches), but to conserve bandwidth the number of visible scanning lines was kept at about 250. The line spacing then substantially exceeded 2 minutes of arc. Subjective line pairing and interline flicker effects were objectionable. With the change to the 1.1:1 aspect ratio for Mod II, and the reduction of the height to 0.13 meter, the angular subtense was reduced to slightly less than 2 minutes of arc. The resulting picture is more nearly optimized at the design viewing distance.

3.4 *Bandwidth and Horizontal Resolution*

With the number of visible lines fixed at about 250, the choice of bandwidth affects only the horizontal resolution. Since the vertical resolution has been optimized for the chosen viewing distance, horizontal resolution better than the vertical would be to some extent wasted. The viewer would have to "look through" the distractions of the scanning pattern discussed in the previous section in order to observe the fine-grained horizontal detail. This effectively sets an upper limit on the bandwidth with 250 lines, since additional bandwidth would appropriately be spent on increasing the number of lines. For a horizontal resolution equal to the vertical, and a 1.1 : 1 aspect ratio, we need, to a first approximation, 275 picture elements in a scanning line, or 137.5 cycles of the highest frequency. Allowing for total horizontal and vertical retrace time of about 23 percent of scanning time, the upper band limit frequency is given by

$$f \cong 250(137.5)(1.23)(30) \cong 1.27 \text{ MHz}.$$

The actual frequency for 1:1 horizontal-to-vertical resolution ratio is less, about 1 MHz. This is because the full vertical resolution corresponding to 20 lines per cm is not realized, by the ratio of the Kell Factor.[15]

At the other extreme, it is undesirable to make the horizontal resolution less than half the vertical resolution, because the overall subjective sharpness of a given horizontal-vertical resolution product decreases when the ratio is less than 0.5 or more than 2.[16] For this reason bandwidths less than about 0.5 MHz would also tend to be inefficient.

Within this range, the value of picture resolution must be weighed against the cost of bandwidth in transmission. The considerations involved cannot be examined here. It is perhaps worth pointing out, however, that in long distance digital trunks, the cost dependence on bandwidth is less severe than might be expected. This is because it is desirable for efficient digital encoding to take advantage of the correlation between successive samples, by differential feedback or other means.[17] When the bandwidth is reduced, for example from 1.0 MHz to 0.5 MHz, and the sampling rate with it, the more widely spaced samples are more nearly independent, and the sample differences are a larger fraction of the sample amplitudes. This requires an increase in the number of bits per sample to obtain the same signal-to-noise ratio (S/N). In addition, in the 1.0 MHz case the noise in the upper half of the bandwidth offers little impairment, as shown by the subjective weighting curve described in the next section, whereas in the 0.5 MHz picture most of the noise contributes to impairments. Therefore the S/N for the 0.5 MHz systems must be higher than for the 1.0 MHz system, for equal subjective noise impairment. The net effect is that the digital transmission rate with the 1.0 MHz bandwidth is reduced only about 25 percent with the 0.5-MHz bandwidth.

We have chosen the 1.0-MHz bandwidth and the nominal 1:1 resolution ratio. In return for the higher cost, the wider bandwidth provides assurance of resolution adequate for face-to-face service for the foreseeable future. With the addition of color, for example, the bandwidth should still provide adequate horizontal resolution.

3.5 *Receiver Roll-Off Characteristic*

The choice of the 1-MHz bandwidth implies that the amplitude and phase response of analog transmission channels will be controlled only within that band. Since the cathode-ray tube is inherently capable of displaying signals at much higher frequencies, it is essential to prevent either components of the camera signal or interference at frequencies above 1.0 MHz from reaching the display. Although the high-frequency signal components could be suppressed equally well either at the camera or the receiver, for maximum noise and interference suppression it is desirable to put all of this "roll-off" attenuation at the receiver. Figure 3 shows the circuit configuration.*

To keep signal energy beyond 1 MHz sufficiently unnoticeable in the picture, we have found it sufficient if the overall frequency response, from visual scene to receiver screen, is down 20 dB at 1 MHz, and more at higher frequencies. One might expect that the maximum resolution within the 1-MHz band would be obtained by using a phase-equalized sharp-cutoff filter to get 20-dB suppression at the band edge. Unfortunately the ringing thus produced is subjectively unacceptable in the picture. To obtain a rapid roll-off in frequency response without ringing, a filter whose impulse response is approximately a gaussian density function may be used.

With sufficiently large delay τ, a filter can be designed so that for values of t in the neighborhood of τ the impulse response $g(t)$ is approximated quite well by

$$g(t) = \frac{\omega_0}{\sqrt{2\pi}} \exp\{-\omega_0^2(t - \tau)^2/2\}. \tag{1}$$

The corresponding frequency response, $G(\omega)$, is given by

$$G(\omega) = \exp(-i\omega\tau - \omega^2/2\omega_0^2). \tag{2}$$

For the present purpose the delay implied by the linear-phase term may be ignored.

---

* As is discussed in the paper on the transmission plan in this issue,[18] pre- and de-emphasis networks are included in the circuit to suppress interference within the band. Although these circuits are physically located in the station set, they may be regarded in this discussion as part of the transmission channel and therefore are not shown in Fig. 3.

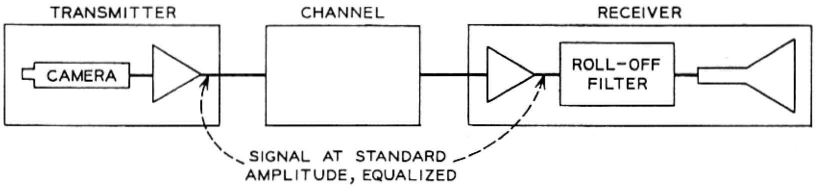

Fig. 3—Circuit configuration with roll-off filter.

To use this filter as the roll-off filter in Fig. 3 the value of ω_0 is to be chosen so that

$$G(2\pi \times 10^6) = 0.1G(0). \tag{3}$$

Assume now that any roll-off in the camera due to aperture or other effects is compensated in the camera circuitry, and that the display is similarly compensated in the receiver electronics if necessary. Assume further that the transmission channel has unity gain and linear phase over the band; departures from this ideal are considered in the next section. Consider the response $h(t)$ of the filter to the camera output signal when a vertical black-white boundary in the picture is scanned. This is

$$h(t) = \frac{\omega_0}{\sqrt{2\pi}} \int_{-\infty}^{t} \exp\left(-\omega_0^2 x^2/2\right) dx. \tag{4}$$

The received picture will therefore shade monotonically from black to white. However, it is not the most pleasing picture that can be transmitted within a given bandwidth. A subjective improvement is obtained by "crispening," that is, introducing an overshoot preceding and following the transition.[19] This not only shortens the actual rise time at the transition, but also provides an effect of greater resolution by increasing the contrast across the boundary. It is done by subtracting the second derivative of the signal from itself. Figure 4 shows the roll-off filter with this addition. It is convenient to represent the crispened signal with the expression

$$s(t) = h(t) - \frac{k}{\omega_0^2} h''(t). \tag{5}$$

With this representation the overshoot peaks are found to be at

$$t = \pm\sqrt{\frac{1+k}{k}} \left(\frac{1}{\omega_0}\right). \tag{6}$$

The amount of the peak overshoot, as a fraction p of the step amplitude, is given by

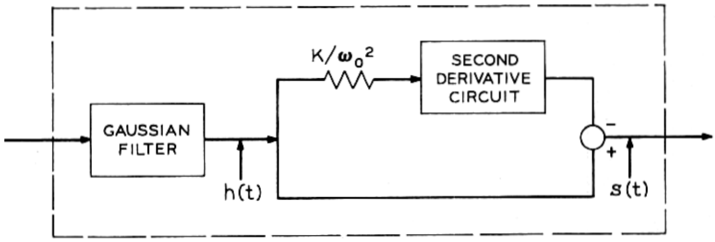

Fig. 4—Gaussian roll-off filter with crispening.

$$p = \frac{1}{\sqrt{2\pi}} \left\{ \sqrt{k(1+k)} \exp\left(-\frac{1+k}{2k}\right) - \int_{-\infty}^{-\sqrt{(1+k)/k}} \exp(-u^2/2)\, du \right\}. \tag{7}$$

The amount of the overshoot in this representation therefore depends only on k. We find k to give the desired overshoot and then choose ω_0 so that the frequency response $S(\omega)$ corresponding to $s(t)$ and given by

$$S(\omega) = 1 + k\left(\frac{\omega^2}{\omega_0^2}\right) \exp(-\omega^2/2\omega_0^2), \tag{8}$$

will be equal to 0.1 at 1 MHz.

Since the crispening technique provides more gain at higher frequencies within the band, it enhances noise. Nonetheless subjective tests carried out by E. F. Brown indicate a preference for about 12 percent overshoot in the presence of noise.[20] The effects of equalization error in transmission, however, make a smaller value desirable. This is because gain changes due to temperature variation on the telephone pairs used for connections to the central office tend to be greatest at the higher frequencies. When the cable is at a higher temperature than the one for which it was equalized, an additional loss increasing with frequency is imposed; when it is colder, there is an incremental gain increasing with frequency. The amount of overshoot for which the picture is about equally impaired with the maximum permissible high-frequency loss and gain deviations was found in studies by M. W. Baldwin, Jr., and H. G. Suyderhoud at Bell Laboratories in 1967 to be about 4 percent. They also found this to be the amount of overshoot which makes the effects of high frequency response deviations most tolerable.

Four percent overshoot corresponds to a value of k equal to 0.5292. For 20-dB attenuation at 1 MHz, the appropriate value of ω_0 is

$2\pi(355, 920)$ radians/second. The resulting time-domain response at an abrupt brightness boundary is shown in Fig. 5; the frequency response is shown in Fig. 6. This is therefore the scene-to-screen response of the system, exclusive of transmission channel effects.

To allow tolerances for the design of the filter, the maximum and minimum overshoot values are set at 4.5 and 3.5 percent respectively. The corresponding values of k, when substituted into equation (7), give upper and lower values at each frequency within which the frequency response is required to lie. The station set receiver is also required to meet the equalization requirement (echo rating) given in the paper on the transmission plan in this issue, as explained in the next section.

IV. PICTURE TRANSMISSION STANDARDS

The parameters given in the previous section define the quality of the picture seen when a station transmitter is connected to a station receiver through a connecting link that is free of noise, interference, and distortion. A complete specification of the picture quality requires a statement of the impairment which is permitted in transmission.

In establishing the basic picture standards, we were free to consider

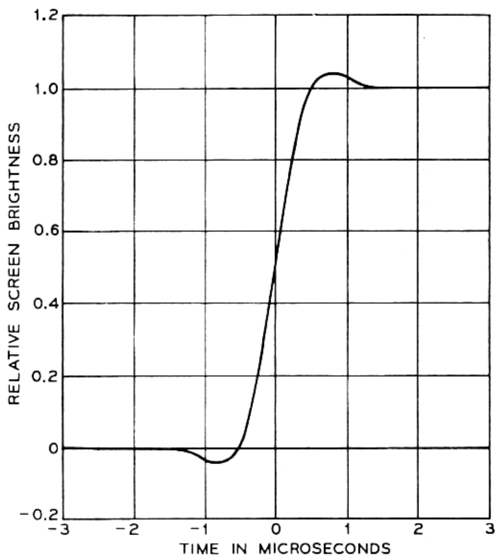

Fig. 5—Response of display at an abrupt brightness boundary.

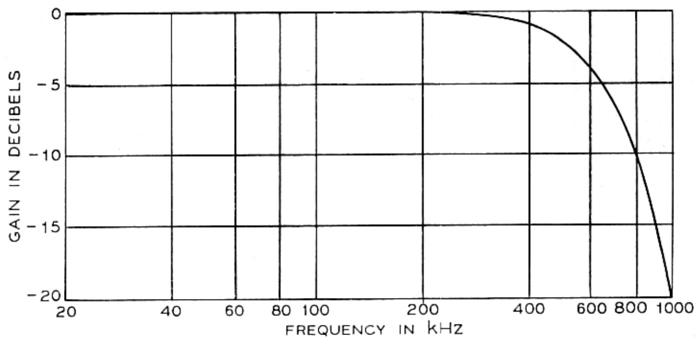

Fig. 6—Overall frequency response, visual scene to receiver screen.

a wide range of values of each, in order finally to establish a service quality sufficient to secure most of the possible enhancement of communication at no greater cost than necessary. With these standards established, the latitude remaining for transmission impairment is limited. This is because the customer will compare the quality of the picture at any given time with the quality of the best *Picturephone* visual telephone image he has had occasion to see. The objective in setting transmission standards is therefore to ensure that the quality of the picture transmitted over the longest connection is unobjectionable compared to the unimpaired picture.

4.1 *Analog Transmission*

Analog transmission may produce identifiable effects in the picture due to several different types of linear distortion and of interference. Since baseband amplifiers must ordinarily be ac-coupled, there is in any connection an accumulated low-frequency roll-off. Clamping in the receiver mitigates this,[21] but there is a residual effect known as "tilt." The accumulated effects of imperfect equalization in many links produce distortions of various types depending upon the value of the net equalization error at each frequency. Single-frequency interference, such as a radio broadcast carrier, tends to produce moving wave patterns. Connections through central offices may pick up brief high-amplitude impulses due to switching transients in telephone lines. Crosstalk from one video channel to another may cause a second image to appear, usually moving relative to the first, and often slowly enough to be identifiable if the degree of coupling were not controlled. Interference from power lines, at 60 Hz and its harmonics, produces moving bars in the picture. Thermal noise in amplifiers, and the accumulation of all other interferences which are individually too low in amplitude to be identifiable, produce random noise on the screen, its appearance depending on the average power spectrum of the noise.

The approach taken to these analog impairments is to characterize them, define the method of measurement, and determine by subjective testing the amount of each impairment which is acceptable in a connection of the maximum number of links. These amounts may then be allocated among the various switches and analog transmission systems as described in the paper on the transmission plan in this issue. The station transmitter and receiver also get an allocation of some types of impairment, as do the analog portions of the digital encoder and decoder.

The basis of this subjective testing is a comment scale. With the 7-point comment scale, for example, the subject is asked which of the following comments applies to the amount of a specific impairment in the picture he is seeing:

(*i*) Not perceptible.
(*ii*) Just perceptible.
(*iii*) Definitely perceptible but only slight impairment to picture.
(*iv*) Impairment to picture but not objectionable.
(*v*) Somewhat objectionable.
(*vi*) Definitely objectionable.
(*vii*) Extremely objectionable.

The subjects chosen are technical people who are not involved in video communications work. The resulting data are processed to estimate the amount of impairment at which 95 percent of the user population would rate the picture comment three or better. The transmission objective will be to make the overwhelming majority of circuits meet such a 95 percent point or better with respect to each impairment.

4.1.1 *Equalization Error and Echo Rating*

As explained in the preceding paper in this issue, the subscriber's connection to the central office utilizes telephone pairs with equalizers at intervals of the order of a mile, as do short interoffice trunks. Since equalization at each amplifier is not perfect, and since temperature changes in the cables introduce additional equalization error, departures from zero gain and linear phase occur and accumulate link by link. Those due to cable loss variations resulting from temperature changes may add systematically. As a result equalization error is commonly the limiting factor in baseband analog transmission.

Since equalization error has different subjective effects depending on its distribution in frequency within the band, it is desirable to define a figure of merit for equalization which increases monotonically with increasing subjective acceptability. This requires relating subjective impairment to particular features of the response as characterized by measurements. The figure of merit will be of greatest usefulness if it is so defined that the figure of merit for two links in tandem can be determined, or at least estimated, by combining in some way the figures for the two links considered separately.

A means of doing this was originally proposed at Bell Laboratories in 1949 by S. Doba, Jr., in a study which was not published. A formulation following his approach, as elaborated recently by H. G. Suyderhoud and R. Piater, may be outlined as follows.

Suppose $f(t)$ to be the response to a unit impulse of an ideal circuit having zero gain and linear phase throughout the band $(0, f_0)$, and arbitrarily large attenuation elsewhere. The actual response $h(t)$ of a given circuit to a unit impulse may be written as:

$$h(t) = \sum_{n=-\infty}^{\infty} C_n f\left(t - \frac{n}{2f_0}\right), \tag{9}$$

where C_n is proportional to $h(n/2f_0)$. The series may be truncated at suitable values of n denoted by $\pm M$. Since a small amount of overall time delay is not an impairment and an amplitude error may be corrected at the station by automatic gain control, we may choose the time origin and the amplitude of $h(t)$ so that $C_0 = 1$, $C_n < 1$, $n \neq 0$. Then the error response is given approximately by:

$$h_e(t) = \sum_{n=-M}^{M} C_n f\left(t - \frac{n}{2f_0}\right) - f(t). \tag{10}$$

Equation (10) expresses the difference between the desired response to a unit impulse and the actual response as a sum of preceding and following "echoes" of the desired signal of amplitude C_n. If we think of the unit impulse as representing, by its amplitude, an element of the picture, $h_e(t)$ represents the extent to which this element is spread to the left and right in transmission. Some of the echo amplitudes may be negative, so that the corresponding displaced picture information undergoes a change in sign.

The impairment introduced into the picture by a single echo has been studied in connection with television by S. Doba, A. D. Fowler and H. N. Christopher,[22] and P. Mertz.[23] The impairment in the case of visual telephone images was studied by R. Piater at Bell Laboratories in

1969. These studies show generally that the impairment increases with displacement from the picture, that is, with increasing $|n|$, as well as with amplitude, that is, with the magnitudes of the coefficients C_n. This suggests weighting the echoes with coefficients W_n and summing to determine the total power of the weighted error transient.

To secure additional freedom in matching the weighted error to subjective test results, each echo may also be frequency weighted. Conceptually this is done by passing impulses of amplitudes W_nC_n separately through filters with impulse responses denoted by $g_n(t)$, and combining the responses. The resulting weighted transient is given by

$$h_w(t) = \sum_{n=-M}^{M} W_nC_ng_n\left(t - \frac{n}{2f_0}\right), \quad \text{with} \quad C_0 = 0. \tag{11}$$

The total error power, relative to the desired signal power, is obtained by squaring and integrating this signal. We get:

$$P_e = \sum_{n=-M}^{M} \sum_{m=-M}^{M} C_nC_mW_mW_nk_{nm} , \tag{12}$$

where

$$k_{nm} = \int_{-\infty}^{\infty} g_n\left(t - \frac{n}{2f_0}\right)g_m\left(t - \frac{m}{2f_0}\right) dt. \tag{13}$$

The problem of defining P_e so that it increases monotonically as impairment increases is therefore reduced to choosing the matrix elements $A_{nm} = W_nW_mk_{nm}$. This must be done by analysis of the results of subjective tests of pictures impaired by circuits for which the values of C_n are accurately controlled. Although a preliminary result has been obtained, this problem is still under investigation. The description above is somewhat simplified, particularly with respect to the method of normalizing the amplitude and of choosing the time origin.

Since the desired received signal power, in response to a unit impulse input, has been set at one unit by adjusting the received level so that $C_0 = 1$, the power P_e represents the relative error power. The desired figure of merit is the level in dB of the error power relative to the signal power. This has come to be called the echo rating, in reference to the procedure of weighting the echoes, although this term is somewhat misleading since individually identifiable echoes are unusual in *Picturephone* transmission. We have:

$$ER = 10 \log P_e . \tag{14}$$

The echo rating objective presently established is −26 dB.

If the equalization errors of several links are random and independent, the values of P_e for the links can be added to determine the overall error power and hence the overall echo rating. Where the equalization error is systematic, the square roots of the separate error powers must be added to determine the square root of the overall error power. On these bases it is possible to allocate echo rating to every analog link in the network, including the station set and analog portions of the codec.[18]

The echo-rating figure of merit is designed to rate circuits in which the error transient associated with the transmission of a unit impulse (which is to say a picture element) is less than one scanning line in duration. Equalization error which is confined to a few kilohertz of bandwidth and is sufficiently severe, or which varies cyclically across the bandwidth, with a period of a few kilohertz, will not be correctly evaluated. In the analog transmission facilities planned for *Picturephone* service there are no networks or network components which should generate these effects. They should therefore occur quite infrequently. If necessary however the echo rating concept can in the future be extended to include them.

4.1.2 *Low-Frequency Cutoff and "Tilt" Impairment*

Baseband amplifiers and the baseband portions of encoders and modulators are usually ac-coupled through capacitive coupling circuits. The impairment due to the resulting low-frequency cutoff, known as "tilt," accumulates approximately linearly in a connected circuit, so that if one link has one percent tilt and a second has two percent, the two together will have three percent. The tilt allowable in a maximum-length connection is therefore allocated among subscriber loops, switching machines and trunks on the basis of linear addition.

The gross effects of low-frequency cutoff are removed by clamping the signal at the horizontal synchronizing pulses in the station receiver; tilt is the residual impairment. Suppose for example that the picture consists of a white rectangle against a grey background. The average signal voltage of the all-grey scanning lines is lower than those which scan the rectangle. A plot of average voltage across a line therefore shows rectangular pulses during the field scanning time. The coupling circuit transmits these pulses with the characteristic decay toward the average value. The result is that the baseline of the synchronizing pulses wanders as shown in Fig. 7. This would produce

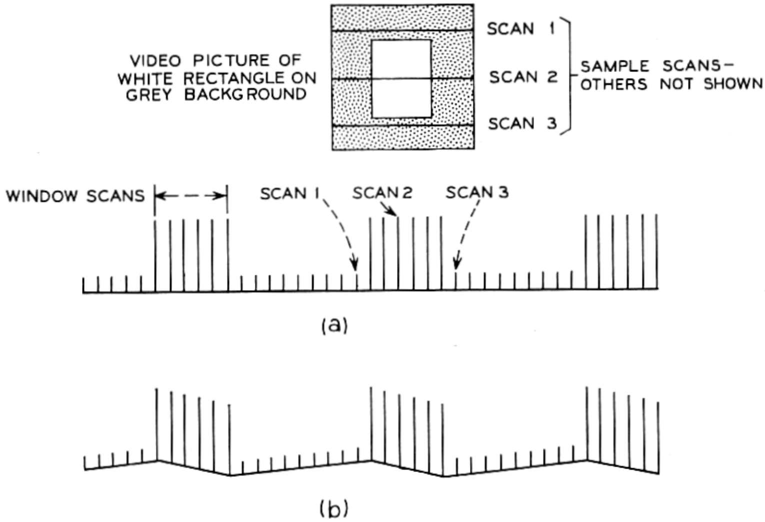

Fig. 7—Effect of low-frequency cutoff on composite video signal, before clamping. Simplified composite video signal (a) before low-frequency cutoff and (b) after low-frequency cutoff.

shading from top to bottom in the picture, except for the clamp, which restores the baseline. The slope or "tilt" remains in the video during each line, however. Figure 8 shows the signal plotted for scanning above, through, and below the white rectangle, assuming the signal clamped at the beginning of each line. The grey value drifts downward when scanning through the rectangle; this means that a shadow will appear on either side of the white rectangle, but more visibly to the right of it.[21] The effect is most noticeable when a white object is moved about in the picture.

Tilt is defined as the decay in the response to a voltage step, measured over the first 100 microseconds, expressed in percent of the step height. Tests made at Bell Laboratories by W. Ohnsorg in 1969 indicate that pictures transmitted over circuits with 10 percent tilt will be rated comment 3 or better by 95 percent of the user population.

4.1.3 *Random Noise*

Random noise interference consists of the sum of thermal noise and all those other interferences appearing in the signal which are at too low a level to be separately identified on the screen. The amplitude distribution is nominally gaussian within the dynamic range of the channel. The spectrum may vary widely depending on the characteristics of the transmission systems through which the signal has passed and other circumstances. The appearance depends on the spectrum. White noise peaks tend to be more visible than black; as a result the appearance is sometimes likened to falling snow. If the noise is preponderantly at low frequencies the "flakes" appear as horizontal streaks; at high frequencies as instantaneous white dots.

The S/N is referred to a point at which the signal is correctly equalized but has not passed through the roll-off filter in the receiver (see Fig. 3). It is defined as

$$S/N = 20 \log (p/n) \tag{15}$$

in which p is the peak-to-peak composite signal voltage and n is the rms value of the noise voltage.

The S/N at which 95 percent of the user population are expected to rate the picture comment three or better depends on the spectrum. The roll-off filter in the receiver suppresses noise at the higher frequencies, and at higher frequencies noise is less visible anyway. An "eye-weighting" curve, found by M. W. Baldwin, Jr., at Bell Laboratories in 1967, is shown in Fig. 9; this represents the relative impairment

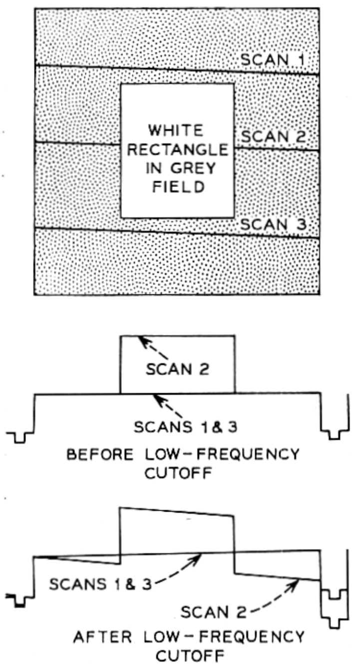

Fig. 8—Effect of low-frequency cutoff on video signal representing one scanning line, after clamping.

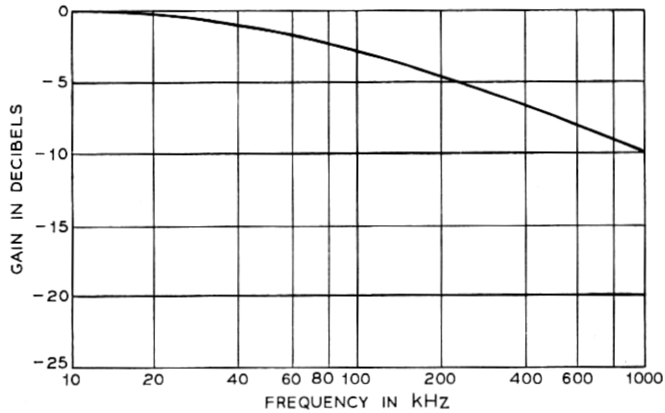

Fig. 9—Subjective noise weighting curve for video signal.

due to noise as a function of its position in the spectrum. Figure 10 shows the total weighting when the receiver roll-off is also taken into account.

The weighting curve of Fig. 10 is used in the same way as in television practice.[24] To determine the weighted S/N for random noise having a given spectrum, the noise may be passed through a filter whose amplitude response approximates the weighting curve of Fig. 10, and the weighted value of n measured with a true rms voltmeter. The equivalent operation may be performed numerically. The 95 percent point for the signal-to-weighted-noise ratio is 52 dB. For example, a noise flat over the 1-MHz band with S/N = 47 dB can be shown to have a signal-to-weighted-noise ratio of 52 dB.

4.1.4 *Switching Noise*

The use of telephone pairs for subscriber loops and short trunks exposes the video signal to the interference sources common in telephone switching of offices. Chief among these is switching noise. The opening of contacts attached to a telephone pair carrying direct current may produce a train of transients of large amplitude containing energy distributed over several MHz. The inevitable crosstalk coupling allows some of this energy to be transferred from telephone pairs to video telephone pairs in the same cable. The resulting interfering transient may reach amplitudes of the order of a volt and this together with its brief duration has led to the use of the term "impulse noise". The transient is typically fairly complex, consisting of a series of separate rapidly decaying oscillations. The duration of each may be 5 to 50 μs,

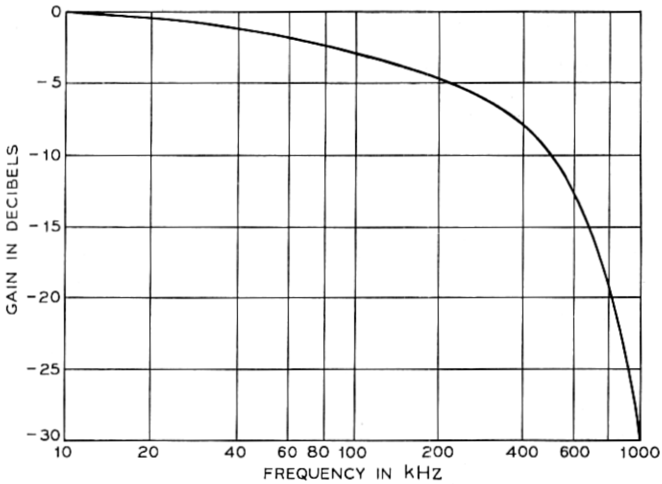

Fig. 10—Noise weighting curve, including effect of receiver roll-off.

they may occur at intervals of 20 to 200μs, and the entire train may last on the order of a millisecond. On the screen the appearance is typically that of a scattering of white dots and small blotches.

Statistically, these transients are very infrequent. With their short duration they contribute very little to total random noise power in spite of their high amplitude. They must therefore be subject to a separate set of requirements. The method of objective measurement must be devised to be representative of their ability to impair the signal.

The means used is to sample the noise on the idle channel at a 10-MHz rate and detect and count samples of amplitude greater than a given threshold. Since the dc content is zero the maximum count is 5×10^6 per second. The actual count, as a fraction of the maximum, over a sufficiently long period, is the estimated probability P of the noise exceeding the threshold.

Switching noise coupled onto a cable pair is subject to the frequency shaping of the equalizing amplifier at the central office. The gain characteristic of this amplifier is adjusted to equalize the cable section behind it, which may range in length from a few hundred feet to a mile or more. To take into account the resulting differences in switching noises as received at the line terminal, the noise is weighted by passing it through a weighting filter before making the measurement. The weighting which is applicable was found by R. M. Lund at Bell

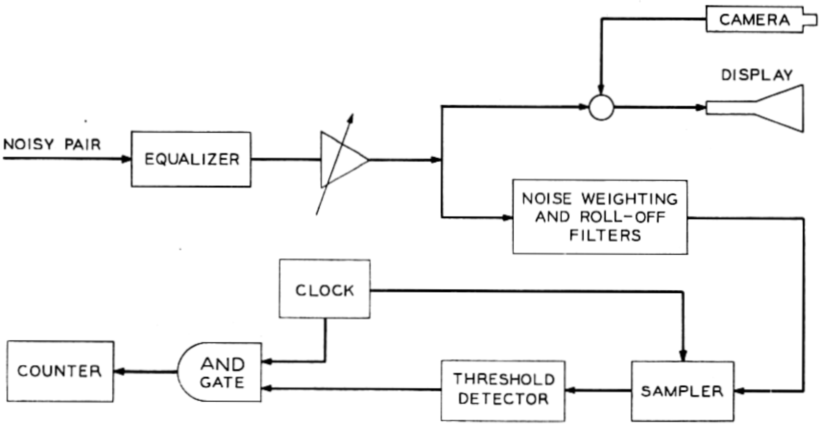

Fig. 11—Measurement arrangement for switching noise.

Laboratories in 1967 to be very closely the curve of Fig. 10, for random noise. The measurement arrangement is shown in Fig. 11. Lund found that noise counts at a threshold level 33 dB below the peak-to-peak composite signal amplitude were best correlated with subjective evaluations of this impairment.

The user may not evaluate the noise he sees over a long period. His attention is more likely to be drawn to switching noise when it is particularly bad, during an interval of the order of a minute. The number of samples which exceed the threshold during such a short interval is a random variable.

Lund has found that the logarithm X of the number of samples which exceed the threshold in one minute is normally distributed, with mean and standard deviation related to the value of P. This may be combined with the probability that a user will rate the impairment denoted by X comment three or better, to determine the probability that a user chosen at random, viewing the picture during an interval chosen at random on a pair whose weighted threshold probability is P will rate it comment three or better. With $P = 6 \times 10^{-5}$ the comment three score is 95 percent.

Because of the difficulty of obtaining and using long-term averages in a changing environment, and some uncertainty resulting from basing noise characterizations on a very limited sample of cables and offices, the requirement has arbitrarily been reduced to 1.5×10^{-5}, representing a comment three score of 99 percent on the basis of available data.

This is then the requirement on the long-term average fraction of weighted switching noise samples which may exceed the −33 dB threshold.

4.1.5 *Single-frequency Noise and Power Hum*

The single-frequency S/N is defined as in equation (15) for random noise,

$$S/N = 20 \log (p/n) \tag{15}$$

in which p is the peak-to-peak composite signal voltage at a point at which the signal is correctly equalized and before it has passed through the receiver roll-off characteristic, and n is the rms value of the sinusoidal interference.

Single-frequency interference at low frequencies which are exact multiples of the field rate of the particular camera and receiver under test produces a fixed bar pattern. At low multiples the bars are horizontal. As the frequency of the interference is made to depart from a multiple of the field rate the bars begin to move. The interference is most impairing when its frequency is about 10 Hz different from the field frequency.

Figure 12 shows the estimated S/N at which the impairment will be rated comment three by 95 percent of the user population, for the range of power hum frequencies, as determined recently by D. B. Robinson, Jr., at Bell Laboratories. Robinson's studies show that frequencies 10 Hz different from the field or frame harmonics continue to provide the locally most severe impairment throughout the band. However, these impairment maxima vary cyclically. They are about 10 dB more severe at multiples of the line scanning rate than at odd multiples of half the line rate.

The curve of Fig. 12 was obtained using a receiver without clamping. The improvement due to clamping cannot be determined by taking into account the measured clamping effectiveness, because effects due to low-frequency interference remain in the picture even with perfect clamping, much as in the case of low-frequency roll-off.

Figure 13 shows the envelope of maxima of the single-frequency S/N corresponding to comment three for the Mod II station set. This curve was obtained by fitting a curve to points at frequencies about 10 Hz different from harmonics of the field or frame frequency and near harmonics of the line rate. At lower frequencies the requirement is diminished by clamping, at higher frequencies by the roll-off filter of the station set and some additional eye weighting.

4.1.6 *Crosstalk*

The telephone cables used for transmission of the *Picturephone* video signal typically exhibit crosstalk coupling from one pair to another. The more important coupling effects are described elsewhere,[25] but may be briefly summarized here. In "far-end" crosstalk the desired and undesired signals are subject to the same amplifier gain and transmission loss, except for the crosstalk coupling loss. The coupling is randomly distributed along the cable. The net effect is that of a single capacitor coupling the transmitter in one circuit to the receiver in the other. "Near-end" crosstalk occurs only when both directions of transmission are in the same cable sheath. The undesired signal is coupled through many paths, each involving a different transmission loss down the cable and back. The resulting frequency characteristic, aside from the receiving amplifier gain, is a random variation about a 4.5-dB per octave trend line of loss decreasing with frequency, compared to a consistent 6-dB per octave for far-end crosstalk with no additional amplication involved.

Tests conducted at Bell Laboratories in 1966 by J. H. Gentry and others indicated that subjective effects of the two types of crosstalk

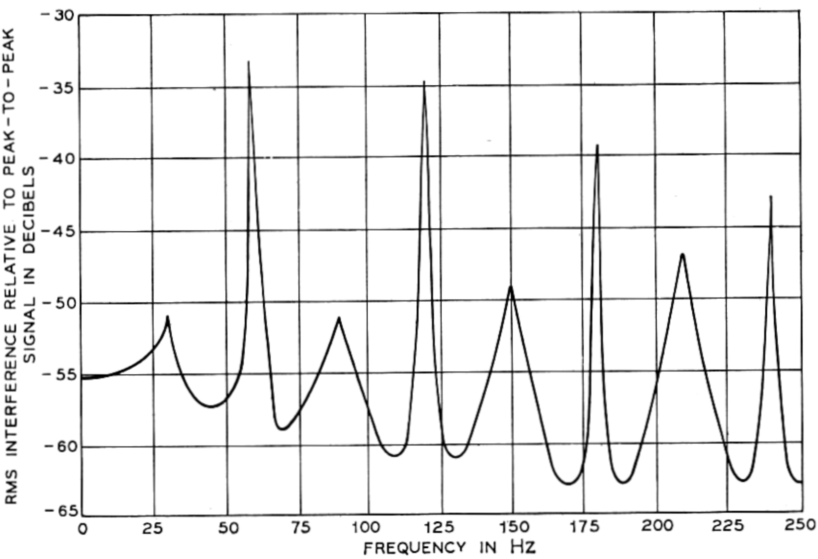

Fig. 12—Single-frequency interference for comment three impairment at ac-power harmonic frequencies.

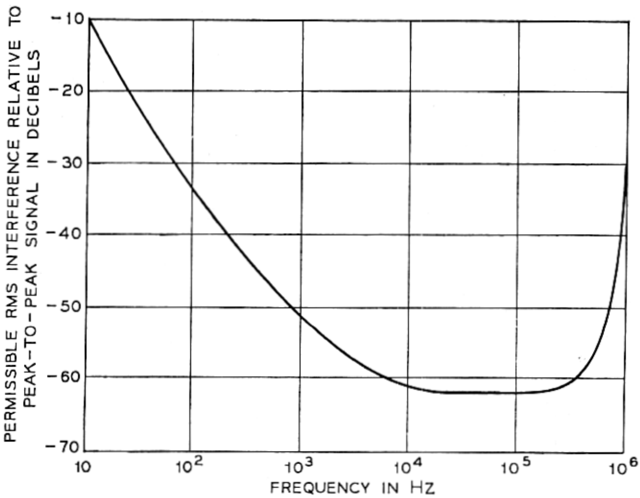

Fig. 13—Envelope of minima of permissible single-frequency interference.

coupling may be equated if the loss at the 150-kHz point of the near-end trend line, less the amplifier gain, is equal to the 150-kHz far-end coupling loss. Therefore the 150-kHz point is used to evaluate any coupling characteristic.

Since the unwanted image has in effect been differentiated by the crosstalk coupling, vertical boundaries between black and white areas tend to be accentuated and appear as lines in the picture, moving horizontally. The number of times per second that the coupled image passes across the screen is equal to the difference between the horizontal scanning frequencies of the connected and the interfering transmitters. The most objectionable rate, determined by J. Orr in tests made at Bell Laboratories in 1969, is about 0.2 passes per second. At this rate, the estimated coupling loss at 150 kHz required to get a rating of comment three or better from 95 percent of the user population is 45 dB, and this is therefore the requirement for any single interferer. Crosstalk also contributes to random noise, and the sum of all crosstalk interference, together with all other random noise sources, must meet the weighted random noise requirement given in Section 4.1.3.

4.2 *Digital Transmission*

The impairment introduced by the use of digital transmission facilities consists of quantization noise, pulse jitter, and the effects of regeneration errors in transmission. (The analog portions of the coding terminals may also contribute to the impairments described in Section 4.1.)

Quantization noise occurs in the differential feedback encoder because the video-sample differences, whose amplitudes occupy a continuum of values over the dynamic range of the signal, must each be assigned one of a small number of values. The resulting quantizing noise has an appearance similar to random noise except at vertical or diagonal brightness boundaries, where close examination usually reveals a slightly pulsating appearance.[17] This "edge busyness" must be traded off against the random noise effect in the encoder design. It is difficult to quantify objectively, and the optimum encoder design is best obtained by visual comparison. An *a priori* requirement has therefore not been placed on quantization noise, although an allotment has been made for the random noise component.[18]

Pulse jitter is an effect in which the pulse rate is alternately speeded up and slowed down, accordian-like. It occurs because the regenerators are timed from the incoming pulse train and are therefore to some extent affected by the information content. Jitter may be removed to any desired degree by buffering and retiming the signal, and the requirements therefore do not affect the basic system design. At present, jitter requirements have not been formulated.

A regeneration error in transmission occurs when noise, interference, and overlapping adjacent pulses combine to operate the regenerator when no pulse was transmitted, or cause it to fail to regenerate a transmitted pulse. The differential feedback decoder stores the resulting noise pulse in its feedback loop, so that its effects may be extended over a substantial part of a scanning line. Since the white errors are more visible than the black, the subjective effect is that of an occasional horizontal white streak along a scanning line. The majority of these are missed in observation.

The subjective effect of pulse regeneration errors depends to some extent upon the encoding algorithm used. With the differential feedback encoder[17] preliminary observations indicate that an error rate of 10^{-6} would introduce negligible impairment. A requirement of 3.3×10^{-7} can be met by the proposed network.[18] This allows some margin for changes in the coding algorithm and for the possibly more stringent requirements of network applications other than visual telephone.

V. SUMMARY

The basic standards of *Picturephone* service have been established with the objective of providing the visual adjunct at no greater cost than is necessary to secure most of the available enhancement of direct conversation, in an instrument which can be associated with an ordinary telephone and used with a minimum of rearrangement of the home or office environment. The standards do not preclude the future application of color or the use of the network for a wide variety of services other than face-to-face conversation. Transmission standards are based on the principle that the quality of the picture after transmission through the longest possible connection should not be objectionable by comparison with the unimpaired picture.

VI. ACKNOWLEDGMENTS

Contributors to this work are very numerous. It is a pleasure to call specific attention, however, to the leadership of Mr. A. D. Hall, the author's original mentor in this field, and to the contributions of Mr. W. T. Wintringham to some of the earlier work described here.

REFERENCES

1. Dorros, I., "The *Picturephone®* System: The Network," B.S.T.J., this issue, pp. 221–233.
2. Garrett, M. A., and Mumford, A. H., "The History of Television," Proc. IEE, London, *99,* No. 17, Part IIIA (1952), pp. 25–42.
3. "Telephones That Can See," The Reporter, September, 1956, pp. 10–11.
4. Kock, W. E., and Miller, R. L., "Two-Way Television Over Telephone Lines," Patent No. 2.895. 005, July 14, 1959.
5. Clark, M. A., Miller, R. L., and Sears, R. W., "Two-Way Television Over Telephone Lines," Patent No. 2,922,843, January 26, 1960.
6. Meacham, L. A., "The *Picturephone,*" IEEE Trans. Broadcasting, *BC-12,* No. 1 (June 12, 1966), pp. 37–52.
7. Hall, A. D., "Experiments with *Picturephone* Service," Bell Laboratories Record *42,* No. 4 (April 1964), pp. 114–120.
8. Carson, D. N., "The Evolution of *Picturephone* Service," Bell Laboratories Record *46,* No. 9 (October 1968), pp. 284–291.
9. Cagle, W. B., Stokes, R. R., and Wright, B. A., "The *Picturephone®* System: 2C Video Telephone Station Set," B.S.T.J., this issue, pp. 271–312.
10. Rosenthal, A. H., Patent No. 2,420,198, issued May 6, 1947.
11. Fink, D. G., *Television Engineering Handbook,* New York: McGraw-Hill Book Company, 1957, p. 1–27.
12. Fowler, A. D., "Low-Frequency Interference in Television Pictures," Proc. IRE, *39,* (October 1951), pp. 1332–1336.
13. Brainard, R. C., Mounts, F. W., and Presada, B., "Low Resolution TV: Subjective Effects of Frame Repetition and Picture Replenishment," B.S.T.J., *46,* No. 1 (January 1967), pp. 261–271.
14. Brown, E. F., "Low Resolution TV: Subjective Comparison of Interlaced and Non-Interlaced Pictures," B.S.T.J., *46,* No. 1 (January 1967), pp. 199–232.
15. Fink, D. G., Editor, *Television Engineering Handbook,* New York: McGraw-Hill Book Company, 1957, pp. 8–6 to 8–23.
16. Baldwin, M. W., Jr., "The Subjective Sharpness of Television Images," B.S.T.J., *19,* No. 4 (October 1940), p. 563.
17. Millard, J. B., and Maunsell, H. I., "The *Picturephone®* System: Digital Encoding of the Video Signal," B.S.T.J., this issue, pp. 459–479.

18. Brown, H. E., "The *Picturephone*® System: Transmission Plan," B.S.T.J., this issue, pp. 351–394.
19. Brown, E. F., "A New Crispening Circuit for Television Images," J. SMPTE, *72*, No. 11 (November 1963), pp. 849–853.
20. Brown, E. F., "Television: The Subjective Effects of Filter Ringing Transients," J. SMPTE, *78*, No. 4, (April 1969), pp. 249–255.
21. Doba, S., Jr., and Rieke, J. W., "Clampers in Video Transmission," Trans. Amer. Inst. Elec. Eng., *69*, Pt. 1 (1950), pp. 477–487.
22. Fowler, A. D., and Christopher, H. N., "Effective Sum of Multiple Echoes in Television," J. SMPTE, *58*, No. 6 (June 1952), pp. 491–500.
23. Mertz, P., "Influence of Echoes on Television Transmission," J. SMPTE, *60*, No. 5 (May 1953), pp. 572–596.
24. Barstow, J. M., and Christopher, H. N., "Measurement of Random Monochrome Video Interference," Trans. A.I.E.E., Communication and Electronics, *72*, Part I (1953), pp. 735–741.
25. Bunin, B. J., Hirsch, R. B., and Olsen, R. E., "The *Picturephone*® System: Crosstalk Considerations in the Transmission of Analog Signals on Paired Cable," B.S.T.J., this issue, pp. 427–457.