Quality Assessment of Digital Television Signals

By LEONARD S. GOLDING

Procedures have existed for many years permitting the subjective evaluation of conventional (analog) television pictures. Subjective data has been amassed and objective test signals have been developed that can be used to predict with some certainty the quality of picture that will be presented to the viewer. It is noted that the digital case is substantially more complex than the analog case, but it is important and worthwhile to derive similar procedures for subjectively evaluating digitally encoded television pictures. A way is suggested for beginning such an effort, and it is foreseen that objective test signals could be developed for assuring a given subjectively evaluated quality level in the presence of various digital impairments. In this way, digitized television pictures could be validly compared with each other and with analog pictures. Topics discussed include: the various subjective grading scales; source coding and channel effects; PCM, DPCM and transform encoding; interframe processing; and encoding of composite and component video signals.

Determining Quality Objectives

Digital encoding and processing of the television signal is rapidly becoming of considerable importance to the broadcast industry, with the proliferation of new digital television equipment for the studio and for transmission which has become available in recent years. A problem which arises with the introduction of the new digital television equipment is the evaluation and the measurement of the quality of digital television signals. Quality objectives must be determined in order to decide what the proper bit rates and proper encoding techniques are that should be used in the new equipment being designed. Before quality objectives can be determined, one must address the following question, "How do we get to the quality objectives that are meaningful for digital television systems?" It is this question I will be discussing in this paper.

One way of getting to the quality objectives is to consider past history relating to the case of analog television. In the early days of analog television, setting up quality objectives and establishing test procedures for evaluating these quality objectives were problems similar to those presently being encountered with digital television. What we are seeking is a set of performance parameters, a standard you might say, similar to what we have for the analog case, which if met, would be considered to provide broadcast-quality television. In the analog case, for example, we have specifications on random noise, impulsive noise, linear distortion and nonlinear distortion. You may not agree with all the specifications for these impairments, but they are a representative set of numbers for these parameters, which most of the industry agrees represents broadcast-quality television. The point is that for the analog television case we have been able, over the years, to come to a point where we can specify quantitatively the parameters that define quality. We would like to reach a similar objective in the digital case. Let us examine how these parameters were derived and how we reached the present objectives for the analog case.

In the early days of television, there were many laboratories such as Bell Telephone Laboratories, R.C.A. Laboratories and others around the world that carried out subjective tests to evaluate such parameters as signal-to-noise ratio and differential phase. They determined just how these parameters related to a given subjective quality such as just-perceptible distortion or annoying distortion. In other words, quantitative parameters were related to a subjective measure of performance. Once the subjective measures were determined, a second phase took place where objective test patterns were developed which could measure these quantitative parameters by means of a vectorscope, waveform monitor, oscilloscope or some other piece of test equipment and relate them directly to picture quality based on the subjective test results. The use of objective test patterns eliminated the need for carrying out subjective testing every time one wished to evaluate the quality of a particular television system. In the development of the objective test patterns, the quantitative measures needed for each impairment, such as noise, were originally based on the subjective test results that had been obtained. As an example of this, Fig. 1 shows a block diagram of the experimental configuration used by Barstow and Christopher[1] to subjectively evaluate the effects of random noise on the analog television signal. Figure 2 shows the results of these measurements where a subjective rating or quality was related to a quantitative value of signal-to-noise ratio. This type of subjective test result forms the basis for all of the current analog specifications.

Three Kinds of Subjective Test

In the area of subjective testing, there are three main types of subjective test. The first is an impairment test where the observer is asked to judge the degree of impairment to the television signal that has been created. This is the type of test used by Barstow and Christopher. Second, there is a quality test where the observer is asked to rate the overall quality of the picture; and third, there is a comparison-type test where the observer is asked to compare the quality of a given picture against the quality of another picture. All three types of subjective test have been used in evaluating analog television signal quality and each has its own grading scale and test procedure. Table I lists typical grading scales that have been used for each type of subjective test. Table II lists common subjective test procedures which have been followed by various countries such as the U.S.A. and the United Kingdom and by several international organizations. The subjective test procedures must consider the number of observers, the type of grading scale used, the viewing conditions and the type of picture material used in the test. These are all referred to in Table III. After a number of years, there has been agreement within the CCIR as to a recommended subjective testing procedure for testing television signals. Table III lists the recommended subjective testing procedure now internationally accepted.

The importance of subjective testing is that the subjective grading scale (such as the impairment scale — which has grades of imperceptible, perceptible but not annoying, slightly annoying, annoying and very annoying — as given in Table IV) is a universal scale which allows one to compare different kinds of impairment in the television picture and therefore allows one to compare one television system with another and one type of signal processing method with another. So the subjective test provides a universal scale that can be used to measure all different kinds of systems and compare them with each other. The subjective testing scale also is directly related to picture quality as seen by the observer, and so permits one to easily define a broadcast-quality signal. In the case of commercial broadcast service, where the ultimate objective is to present a pleasing and high-quality picture to the observer, the subjective scale allows that picture quality to be evaluated directly.

Let us consider the impairment test in greater detail. In the impairment test, one adds different amounts of an impairment such as noise to the original signal and determines how the observer evaluates the visibility of the impairment as a function of the amount added to the television signal. Typically, one considers a single parameter such as the amount of noise or the power of the noise and relates that to a judgment on the subjective grading scale made by the observer as given in Table III.

Presented on 24 January 1976 at the Society's Winter Television Conference in Detroit, by Leonard S. Golding, Digital Communications Corp., 19 Firstfield Rd., Gaithersburg, MD 20760. The paper was subsequently revised for publication in *Digital Video* © 1977 by the Society of Motion Picture and Television Engineers, Inc. and is being reprinted here.

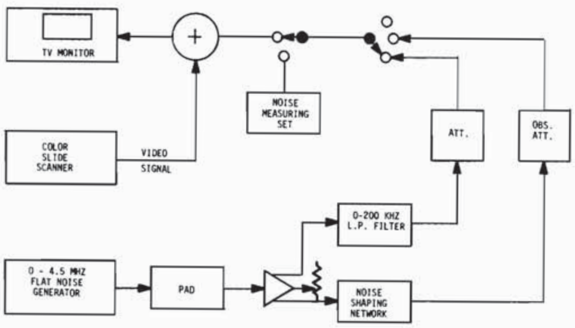

Fig. 1. Experimental configuration used by Barstow and Christopher for subjective evaluation of noise.

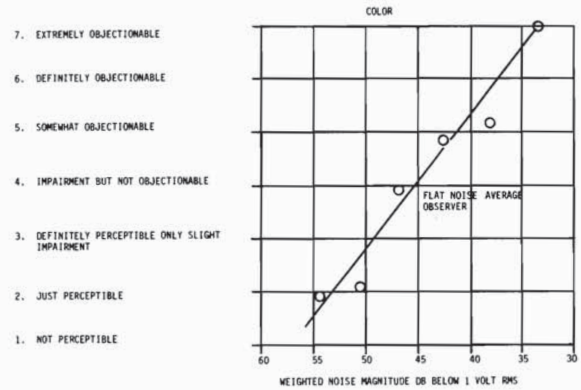

Fig. 2. Analog signal-to-noise measurements.

In the case of digital television, one could carry out similar types of subjective testing. For example, in the analog-to-digital conversion of the signal in a pulse-code-modulation system, one could vary the number of bits per sample (related to the number of sampling levels) used to quantize the signal and evaluate a subjective quality associated with varying that particular parameter. One could also vary the sampling frequency and determine a subjective quality related to the amount of impairment occurring in the picture due to different sampling frequencies. As in the analog case, where varying the amount of noise and comparing it to subjective quality allowed one to determine a suitable signal-to-noise ratio, for a broadcast-quality signal, one could determine the number of bits per sample and the sampling frequency, based on the subjective rating, that is required in order to provide a broadcast-quality signal.

In both these cases a broadcast-quality signal was determined by picking some value of subjective grade, such as "just imperceptible" (1.5 on the 6-point impairment scale — meaning that half the observers can perceive the impairment and half cannot) and using that as a measure of broadcast quality. Thus, just as analog broadcast quality was equated to a subjective grade of "just imperceptible" for a number of different types of impairment one can also similarly assess signal quality in the digital case. Table IV lists analog impairments that have been evaluated subjectively. The parameters are quite different, but the subjective test procedures could be quite similar. Subjective testing thus provides a basis for deriving specifications on picture quality for both analog and digital television systems. Furthermore, subjective scales could provide a means of relating analog television systems to digital television systems. There are, however, a number of significant differences in the digital case which makes the process more complicated than it was for the analog case.

Evaluating Digital Parameters

One of the desirable features of digital signal processing of the television signal is that there is a great deal of flexibility in being able to carry out a variety of different types of signal processing without seriously impairing the signal. While this feature is very attractive from a design standpoint, it makes assessment of quality more difficult, for it leads to many more parameters and cases to evaluate than was encountered in the analog case.

Table V lists typical sets of parameters that would be of interest to evaluate for different coding methods. If we consider an analog-to-digital-conversion process such as pulse code modulation, then such parameters as sampling rate, number of bits per sample, companding law, clock jitter, and the description of the filtering used, are the kind of parameters which are important in determining the quality of the reconstructed analog signal. When other types of digital encoding methods are used the set of parameters that must be evaluated will vary and be dependent on the type of digital processing carried out. So, for example, if we were to use differential-pulse-code modulation as the means of converting the analog television signal into a digitally encoded form, the parameters of interest would be different, e.g., the particular prediction algorithm used in the differential PCM coding, the number of bits per sample used in the feedback loop, and the number of previous samples used in predicting the next value of the signal. Other coding methods such as transform coding require yet another set of parameters to be evaluated, as indicated in Table V.

In the case of digital television, there are two classes of parameters to be evaluated which impact the quality of the picture. The first class of parameters relates to the conversion of the analog signal into digital form and the conversion of the digital signal back to analog form, a type of processing termed "source coding." Parameters associated with different methods of source coding that are to be evaluated are listed in Table V.

There are also impairments introduced into the picture after the signal is in digital form. They are typically called channel effects and are also listed in Table V. Such parameters are: random errors which occur on the bit stream, slips of the bit timing clock or jitter of the clock, burst errors, etc. These errors which are introduced into the bit stream after the television signal is in digital form will result in additional impairments appearing in the reconstructed analog signal and must also be evaluated.

Correlated Impairments

There is another complicating difference associated with the digital case, however, and that is that the nature of the impairment in the reconstructed analog signal, due to channel effects occurring on the bit stream, is related to the type of source coding that was used to convert the analog signal into digital form. Because the number and type of bit errors introduced into the bit stream cause different analog

Table I. Subject grading scales.

Impairment	Quality	Comparison
5-Imperceptible	A-Excellent	+2 much better
4-Perceptible but not annoying	B-Good	+1 better
3-Somewhat annoying	C-Fair	0 the same
2-Severely annoying	D-Poor	−1 worse
1-Unusable	E-Bad	−2 much worse
1-Imperceptible	1-Excellent	+3 much better
2-Just perceptible	2-Good	+2 better
3-Definitely perceptible but not disturbing	3-Fairly good	+1 slightly better
4-Somewhat objectionable	4-Rather poor	0 the same
5-Definitely objectionable	5-Poor	−1 slightly worse
6-Extremely objectionable	6-Very poor	−2 worse
		−3 much worse

Table II. Subjective test procedures.

Reference	U.K. (C.C.I.R., 1963–1966)		E.B.U., O.I.R.T., (C.C.I.R., 1963–1966)		U.S.A. (C.C.I.R., 1963–1966)	U.S.A. (C.C.I.R., 1966–1969)	U.S.A. (C.C.I.R., 1966–1969)	
Observers Category	Non-Expert				Non-Expert	Expert	Non-Expert	
Number	20–25				Approx. 200	>10	Approx. 20	
Grading Scale Type	Quality	Impairment	Quality	Comparison	Quality	Impairment	Comparison	Impmnt.
Number of Grades	5	6	6	7	6	7	5	5
Test Pictures Number	4–8		5		2–8	3–4	6	
Viewing Conditions: Ratio of viewing distance to picture height	6		4–6		6–8	4	6	
Peak Luminance on the screen (cd/m²)	50		41–54		70	170 (monochrome) 34 (color)	50	
Contrast range of the picture	Not specified				Not specified			
Luminance of inactive tube screen (cd/m²)	≤0.5		0.5		2		Approx. 0.5	
Luminance of backcloth (cd/m²)	1 Illuminant C							

Table II continued.

Reference	Fed. Rep. of Germany (C.C.I.R., 1963–1966B)		Japan (C.C.I.R., 1963–1966C and 1966–1969A)	
Observers Category	Non-Expert		Non-Expert	
Number	>10		20–25	
Grading Scale Type	Quality	Comparison	Impairment	Quality
Number of Grades	5	5	5	5
Test Pictures Number	>5		>3	
Viewing Conditions: Ratio of viewing distance to picture height	6		6–8	
Peak Luminance on the screen (cd/m²)	50		Approx. 400 (monochrome) 74–84 (color)	
Contrast range of the picture	Not specified		30/1 to 50/1	
Luminance of inactive tube screen (cd/m²)	≤0.5		Approx. 5 (monochrome) 0.7–2 (color)	

Table III. CCIR-recommended subjective testing procedures.

Viewing condition designation	Viewing condition description	Specifications: 50-fields/s systems	Specifications: 60-fields/s systems
a	ratio of viewing distance to picture height	6	4 to 6
b	peak screen luminance (cd/m²)	70 ± 10	70 ± 10
c	ratio of inactive-tube (cutoff) luminance to peak luminance	0.02	0.02
d	ratio of screen luminance displaying black level in completely dark room to that corresponding to peak white	approx. 001	—
e	ratio of luminance of background behind picture monitor to picture peak luminance	approx. 0.1	approx. 0.15
f	other room illumination	low	low
g	chromaticity of background	white	D_{65}
h	ratio of solid angle subtended by that part of the background which satisfies this specification to that subtended by the picture	≥9	—

Table IV. Picture impairments.

Analog case	Digital case
Additive independent noise	Sampling noise
Random	Quantization noise
Impulsive	Intersymbol Interference
Periodic	
Crosstalk	Bit error rate
Linear distortion	Bit error time Distribution
Field time	
Line time	Bit Timing clock jitter
Short time	
Chrominance/luminance	Phase and amplitude hits
Gain & delay inequality	Bit timing slips
	Impulsive noise
Gain/frequency	
Nonlinear distortion	
Differential phase	
Differential gain	
Chrominance/luminance	
Intermodulation	
Luminance nonlinear distortion	
Chrominance nonlinear	
Gain and phase distortion	
Synchronizing pulse nonlinearity	

impairments for different types of source coding, there is an interrelation that must be considered when evaluating the quality of a digital system.

Furthermore, the impairments introduced in the analog-to-digital conversion process are correlated with the television signal and are strongly dependent on the characteristics of the picture material being digitized. The correlated nature of the impairments can result in some peculiar subjective effects. For example, in the case of pulse-code modulation, when too few bits per sample are used, false edges or contours appear in the picture, demonstrating a type of noise that is not normally seen in analog television. (In the analog television case, most of the impairments are uncorrelated with the television signal and produce a more random type of noise impairment in the picture, but this is not necessarily true for the digital case.)

There are other differences in the digital television case which must also be considered. Because television signals can be stored in digital form, interframe or frame-to-frame coding of the signal is possible, and devices for accomplishing this have already been developed by certain manufacturers. This type of source coding required consideration of frame-to-frame subjective effects which must be tested with picture material that involves motion between one frame and the next. This in turn leads into more complex subjective testing procedures than have been used in the analog television case.

Table V. Parameters in digital impairment testing.

	Source Coding	Channel
PCM	Filter parameters	Random error rate
	Sampling rate	Imp. error rate
	No. of bits/sample	Burst duration
	Companding law	Clock jitter
	Clock jitter	Phase/amp. hits
		Bit clock slips
DPCM	Prediction algorithm	
	Companding law	
	No. of bits/sample in feedback loop	
	Loop filter parameters	
Transform	No. of coefficients used	
	Companding law per coefficient	
	No. of bits/coefficient	
	Filter parameters	

Composite vs Component Encoding

In the source coding area there is yet another basic choice to be made between coding methods. It involves choosing between direct analog-to-digital conversion of the composite color television signal and separate encoding of the components of the television signal (the luminance signal and the two chrominance signals). The impairments perceived when using these different analog-to-digital coding processes are quite different — especially if there is interaction between the chrominance and luminance signals.

Summarizing Digital Impairments

At this point it may be useful to summarize what we have found about digital

Table VI. Quality-assessment procedure.

- Determine source coding technique to be evaluated and channel conditions to be considered.
- Determine key parameters which are to be tested, and minimum range over which each parameter should be varied. Minimize number of combinations of different parameters, which must be tested.
- Carry out subjective impairment test, following internationally accepted practices for subjecting testing.
- Compare performance with other digital systems using subjective grading scale as common measure of performance.
- Develop objective test signals and procedures which permit evaluation of performance of given digital system, using quantitative measures on the television signal.

parameters, the impairments produced, and the difficulties of subjective testing. We have found that in general the digital case is a more complex case to develop standards for because:

1. The impairments are more varied as they are correlated with the television signal and are a function of both the source coding and channel effects.

2. Generally more parameters need to be subjectively tested to fully evaluate specific coding methods, and there are potentially a greater number of coding methods which may be useful and practical to consider.

3. Frame-to-frame signal processing is quite feasible with digital techniques, which means that subjective tests taking motion into account must also be considered.

4. After the signal is digitized and encoded, the primary effect of further sources of degradation is to increase the bit error rate and possibly change the error pattern.

5. If error coding is employed to reduce the bit error rate on the digital bit stream then the error coding process used also will affect how bit errors will appear in the reconstructed analog signals; thus the impairments are a function not only of the analog-to-digital coding process, but any signal processing carried out on the signal after it has been converted into digital form.

6. Chrominance/luminance impairments depend on whether the digitizing and encoding is done on composite video or on components.

Overcoming the Complications

While the digital case is more complicated than the analog case, I believe it can be handled quite successfully with some intelligent planning. For each analog-to-digital coding process one can specify a particular set of parameters such as number of bits per sample, sampling rate, etc., that have to be evaluated. Impairment or quality testing following recommended test procedures could be carried out to relate each of these parameters to an equivalent subjective quality grade. If a subjective quality of "just imperceptible" is selected, then through a series of subjective tests the

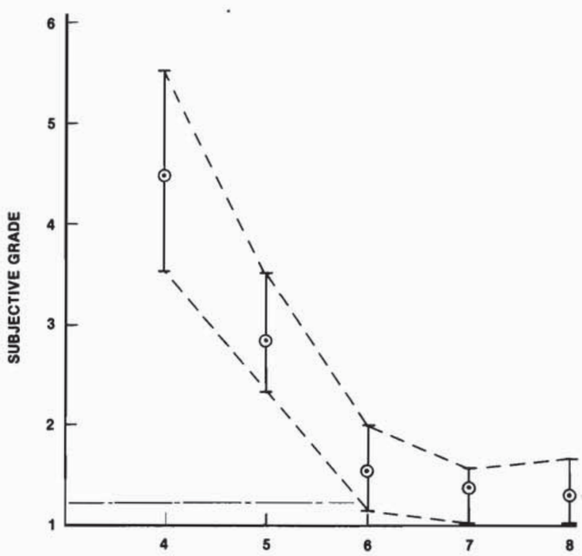

Fig. 3. Example of measured results for PCM coding: variation in subjective impairment at different numbers of bits per sample, using dither. (Vertical bars show variation in grade for different picture sources, and open circles denote mean grades for all picture sources; horizontal line at Subjective Grade 1.25 indicates the mean score for an unquantized picture.)

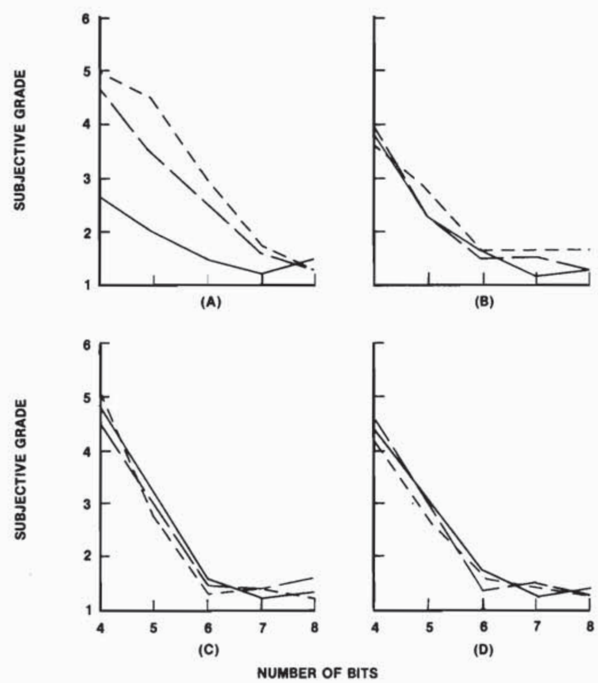

Fig. 4. Effect of different sampling frequencies (PCM coding) on critical picture color bars and on noncritical pictures taken off-air with a receiver. (A) Color bars; no dither. (B) Color bars; with dither. (C) Off-air pictures; no dither. (D) Off-air pictures; with dither. (Solid lines denote a sampling frequency f_s of three times color subcarrier; long dashes show $f_s = 851 \times$ line frequency; and short dashes show f_s unlocked.)

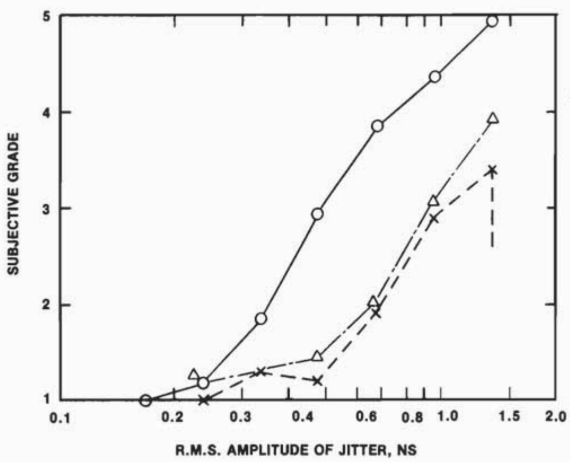

Fig. 5. Measured results for timing jitter. The impairment is caused by white Gaussian jitter on a display of 100% colorbars. Circles, triangles and crosses respectively denote maximum jitter frequencies of 20 kHz, 600 kHz and 6 MHz.

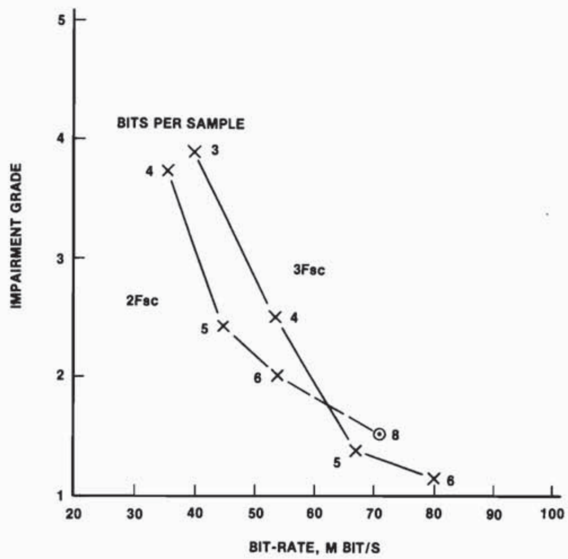

Fig. 6. A digital subjective test: impairment vs bit rate for DPCM coding of PAL signals with sampling frequencies of $2f_{sc}$ and $3f_{sc}$. Crosses show results for DPCM and circles for PCM.

correct value of the parameters to achieve this subjective quality could be determined by a series of impairment or quality tests. While the number of parameters may be large, by some careful planning and some preliminary screening the range of each of the parameters that have to be tested can be made relatively small, for a given subjective quality. Knowing the nature of the impairments introduced by the particular analog-to-digital coding method one could select a reasonably small set of picture materials that are effective at showing up these impairments and that could be used in carrying out the subjective tests. For each source-coding method the effects of different types of error patterns on the digitally encoded signal could be evaluated. In a very systematic way, the set of parameters, which give a specified subjective quality, could be determined for each analog-to-digital coding method in the presence of different bit error rates and bit error patterns that might be encountered in practice. The subjective quality scale would then provide the means of comparing different coding methods with regard to the bit rate and bit error rate needed to provide a specified subjective quality.

Objective Test Signals

Once this subjective test data had been compiled one could dispense with the frequent subjective tests (as has been done with analog television) and look into objective test signals which could be used to evaluate a given subjective performance. For example, the pulse and bar pattern commonly used in the analog case for measuring the short time distortion could be used in the digital case to measure edge busyness and background noise in constant gray level areas of the picture. For the PCM analog-to-digital coding technique a ramp signal would be quite useful in detecting contours and quantization noise. While much more work must be done to determine the correct objective test signals, the procedure for arriving at these test signals can follow along similar lines to those used to arrive at analog test signals. Table VI outlines what I believe to be a quality assessment procedure for the digital case that can be followed to arrive at quality objectives; the procedure is similar to that used originally to arrive at the analog quality objectives.

Some examples of subjective tests that have already been carried out successfully on different digital coding methods are illustrated in Figs. 3, 4 and 5. This data, provided by the British Broadcasting Corporation,[2,3] involved testing the analog-to-digital encoding of the PAL television signal. In Fig. 3, the number of bits per sample was varied and related to a subjective quality. In Fig. 4, different sampling frequencies were evaluated and related to a subjective quality as a function of different number of bits per sample. In Fig. 5, the effects of timing jitter were related to subjective quality for a PCM signal. The test procedure followed was similar to that recommended by the CCIR (Table III). Five or six different pictures were used as the subject material. The results, as illustrated in these figures, show how quantitative values can be determined for the set of parameters associated with PCM encoding of the signal in order to obtain a given specified subjective quality.

Figure 6 illustrates how the subjective rating scale can be used to compare different analog-to-digital coding methods. In this figure both pulse-code modulation and differential pulse-code modulation are related to a given subjective quality at a given bit rate. The subjective quality is shown for different numbers of bits per sample and different sampling rates for the DPCM method. The PCM performance, with 8 bits per sample and a sampling rate of twice the color subcarrier frequency is also plotted on the same graph. These test results, provided by the BBC, illustrate how a comparison can be made. As previously mentioned, different digital coding methods could also be compared to analog system performance by using one of the subjective grading scales such as the impairment grading scale, as a common basis of comparison even though the nature of the impairments may be different. One must be careful that the test results used, however, apply to a sufficiently large amount of picture material to make the comparison valid.

While the procedure to get to the quality objectives for digital television systems appears to involve a considerable effort (possibly a lot more effort than was originally needed to arrive at the analog television quality objectives), it is expected that this procedure would be carried out over a considerable period of time. Furthermore, with the type of impairments being much more varied in the digital case than in the analog case, it would appear that the subjective testing procedure would be the only way of getting a common measure of quality which could be used to determine the correct parameters for different digital encoding methods, different bit error rates and different digital impairments which might occur.

References

1. J. M. Barstow and H. N. Christopher, "The Measurement of Random Video Interference to Monochrome and Color Television Pictures," *AIEE Trans.*, No. 63, Nov. 1962.
2. V. G. Devereux, "Pulse Code Modulation of Video Signals: Subjective Study of Coding Parameters," BBC Research Report No. 1971/40.
3. "Pulse Code Modulation of Video Signals: 8-Bit Coder and Decoder," BBC Research Report No. 1970/25.