T1A1.5/96-105

COMMITTEE T1–TELECOMMUNICATIONS

STANDARDS CONTRIBUTION

STANDARDS PROJECT:	Analog Interface Performance Specificationsfor Digital Video Teleconferencing/Video Telephony Service
SUBJECT:	An Objective Model for Video Quality Performance
SOURCE:	Bellcore
CONTACT(S):	Bill Cotton Bellcore 331 Newman Springs Rd Red Bank, NJ 07701 (908) 758-4510
DATE:	March 11, 1996
DISTRIBUTION:	T1A1.5 Technical Subcommittee Working Group Members
KEYWORDS:	video quality, objective models
ABSTRACT:	This is a proposed contribution to ITU-T Study Group 12 which addresses that portion of Question 22 dealing with the evaluation of video quality by objective means. Although T1A1 has not yet reviewed this proposal, Working Group T1A1.5 discussed a similar contribution at their December 1995 meeting. We are seeking review and comment prior to submitting it to ITU-T as a Bellcore contribution. The contribution describes several measures that might be useful as a basis for comparing the present operational readiness of a video system with the same system's past performance. This contribution provides the results of a study conducted to determine the feasibility of developing a video performance model using the T1A1.5 measures to allow the monitoring of video service performance across time. Based on the results of this study it is concluded that the performance measures can be used to develop an objective model of video service performance.

An Objective Model For Video Quality Evaluation

Bellcore

1.0 Introduction

A primary concern of many video service providers is how to evaluate the transmission performance of compressed digital video systems and how to maintain these systems after installation. For example, if an end customer of a compressed digital video service complains about poor picture quality, there is, presently, no measurement method available to a technician to determine what part, if any, of the RBOC's transmission system has deteriorated or changed in transmission performance. The difficulty in developing performance measurement techniques for compressed digital video systems is due to the fact that the compression algorithms used in many of these systems dynamically allocate bits between intra-frame and inter-frame information, resulting in a trade-off in the use of the bits to convey information describing detail and motion in the scene. Therefore, the system's performance is dependent upon the amount of motion and detail within the encoded video scene. Unlike the assessment of analog or uncompressed digital video, static test signals and current standard test methods may be inappropriate or inadequate for quantifying the performance of many compressed digital video systems. Until standard transmission parameters and methodology for measuring those parameters on compressed digital video systems is developed, it would be highly desirable to develop an interim method that would at least be able to reliably detect changes in transmission performance.

Since 1984, T1A1.5 and NTIA/ITS have actively pursued a project to specify a system for measuring impairments associated with compressed digital video signals. The purpose of this effort was to develop an objective measure of video quality that would agree with judgments obtained from a panel of viewers. If successful, such a system would provide a reliable, unbiased, and cost-effective means of characterizing video quality without the necessity for expensive and time-consuming subjective testing. After extensive testing of algorithms developed over the last decade, in 1995 T1A1.5 concluded that while the objective measures now identified could not reliably characterize the quality provided by different video systems, they might be "useful as a basis for comparing the present operational readiness of a system with the same system's past performance [1]."

This contribution provides the results of a preliminary study conducted to determine the feasibility of developing a video performance model using the T1A1.5 measures to allow the monitoring of video service performance across time.

2.0 Video Material

2.1 Video Source Material

Two video test patterns were used as source material.[1] One was high in spatial detail with no motion, while the other exhibited a moderate amount of both detail and motion. These patterns were obtained from the Standards Committee T1A1.5 master tape source of video test scenes for subjective and objective video performance assessment. The stationary pattern, *Balls*, was a little over 3 seconds in length and consisted of a singular consistent pattern of approximately 500 white circular balls on a black background. The dynamic test pattern, *Rotating Wheel*, was about 5 seconds in length and consisted of a wheel rotating at a speed of 90 degrees per second over a stationary background. The wheel consisted of 10 black spokes, each 18 degrees wide and emanating from the center to the outside edge of the picture. The stationary background consisted of 10 color spokes, each 36 degrees wide, emanating from the outside edge of the picture and ending at one side of a solid white decagon figure located at the center of the screen and with a width of about half the screen width.

2.2 Stimulus Material

The two video source test patterns were transferred from D1 digital tape, via a video disk system to computer files on a Sun Spark 20 workstation for video processing. Each of the test pattern files was first MPEG encoded at different bit rates. Simulated digital video coding impairments were then added to each of the resulting MPEG sequence files.

[1] We would like to thank John Ellis of NYNEX for his suggestion that we use dynamic test patterns as source reference material.

2.2.1 MPEG Encoding

Each of the two test patterns was encoded with MPEG1 at 1.5 Mbits/s and with MPEG2 at 2, 4, 6 and 8 Mbits/s. The encoding was performed in software on the Sun Spark 20 workstation using Bellcore implementations of the MPEG1 and MPEG2 algorithms. This resulted in 5 source files for each of the test patterns. The source files correspond to initial video service performance as measured at service installation.

2.2.2 Addition of Simulated Coding Impairments

Simulated digital video coding impairments were added to each of the source files. Five levels each of random noise (RN), signal correlated noise (SCR), blurring (BLR) and block distortion (BLK) were added to each of the source files for a total of 100 test conditions (1 test pattern pair x 5 MPEG bit rates x 4 impairments x 5 impairment levels).

The video coding impairments were simulated using a software algorithm called VIRIS. The VIRIS impairment levels were selected to correspond to specific quality levels based on results from previous subjective tests with VIRIS. In these previous tests, various types of entertainment television material (i.e., popular movie clips) were degraded with VIRIS impairments [2]. Using entertainment television as their judgment criteria, non-expert participants rated the overall quality of the impaired test sequences using the nine-point rating scale shown below:

The VIRIS impairment levels chosen for this study were ones that corresponded to mean opinion scores (MOS) of 9, 8, 7, 5 and 3 in the previous studies. Table 2 shows the VIRIS data input level that corresponds to these specific quality levels for each of the VIRIS impairments used in this study.

3.0 Parameter Computation

The NTIA procedure defines twelve parameters based on an analysis of a video sequence on a frame by frame basis. For use in monitoring service performance, an initial set of measurements would be performed after system installation. Changes in system performance across time is characterized by later repeating the measurements and then computing parameters representing the differences between the two sets of measurements.

Each video sequence is defined by three types of measures that describe the information content present. Five temporal domain measures describe the motion energy in the sequence which are computed by comparing the difference in information content between successive frames in each sequence. Five spatial domain measures describe the amount of spatial information in the individual frames of the sequence. Two Fourier transform measures characterize the changes in frequency content of the sequence between the initial and subsequent measurements.

The algorithms used to process and extract the parameters listed in Table 3 are discussed in reference 1. Each parameter was computed for the rotating wheel and ball test patterns for all 100 test conditions.

4.0 Single Stage Model

Stepwise multiple regression procedures were used to develop a model that characterized the functional relationship between the parameters computed for each type of test pattern and the impairment levels. The independent variables in this analysis were the 12 objective parameters computed for each of the rotating wheel and ball test patterns, and the dependent variable was the impairment level associated with each of the 100 test conditions.

4.1 Single Stage Regression Model

The results of the regression analysis found that impairment level could be most reliably predicted by five of the independent variables.

$$\begin{aligned}\text{Level} = 8.417 &- 70.276 \times \text{MEL\_W} - 3.165 \times \text{MELN\_W} + 0.0096 \times \text{FSA\_W} \\ &+ 19.026 \times \text{EEA\_B} + 0.147 \times \text{FSA\_B}\end{aligned}$$

The single stage regression model accounts for 89%[2] of the variance in impairment level. Figure 1 shows the predicted impairment level for each condition plotted as a function of the amount of impairment present. Although the correlation between the predicted and observed values is relatively high (.95), the model is not sufficiently sensitive to detect relatively small changes in video quality. This is evidenced by the amount of overlap in the distributions of predicted scores for impairment levels of 7, 8, and 9.

[2] The actual percentage of variance accounted for by the model was slightly higher. All of the variances reported in this paper have been adjusted to estimate the proportion of variance that would be accounted for if the model were to be cross-validated on a different sample.

The insensitivity of this model may result from using a single set of parameters and weights to characterize the relationship between video quality and different types of digital impairments. For example, the correlation between predicted and observed values ranged from .98 for the blocking conditions, to only .92 for the random noise impairment conditions. This suggests that an improved model would be obtained if the type of impairment present could be determined, and then a model applied that appropriately describes the relationship between that impairment and customer perception of video quality.

5.0 A Two-Stage Model

A two-stage approach was applied to the same data to develop a model to characterize video quality. First, a model was developed to predict the class of impairment present (i.e., blocking, blurring, signal correlated noise, and random noise) as a function of the profile of the pattern of parameters obtained for the two test patterns. Then regression procedures were used to specify the functional relationship between the measured parameters and amount of impairment present.

5.1 Stepwise Discriminant Analysis

Discriminant analysis is a statistical technique that is used when the criterion variable is categorical and the predictor variables are interval-scaled data. The procedure identifies a profile of independent variables that optimally classifies a set of measurements into mutually exclusive and exhaustive groups. When many independent or predictor variables are available, stepwise statistical selection procedures are used to determine which variables should be included in the discriminant function [3].

In the analysis the criterion variable was impairment type, and the independent variables were each of the twelve parameters computed for the rotating wheel and ball test patterns.

5.1.1 Discriminant Analysis Model

The discriminant analysis model is specified in terms of a classification function for each type of impairment. The four functions are:

$$
\begin{aligned}
\textbf{Blocking} = -4.396 &+ 274.5\times \mathrm{MED\_W} - 44.67\times \mathrm{EEL\_W} + 4.19\times \mathrm{HVD\_W} \\
&+ 0.226\times \mathrm{FSA\_W} + 11.17\times \mathrm{MEA\_B} - 20.63\times \mathrm{MED\_B} \\
&+ 1.012\times \mathrm{MELN\_W} - 17.67\times \mathrm{FEE\_B} - 0.683\times \mathrm{FSA\_B} \\
&- 0.091\times \mathrm{FSL\_B}
\end{aligned}
$$

$$
\begin{aligned}
\textbf{Blurring} = -131.22 &- 1636.2\times \mathrm{MED\_W} - 926.0\times \mathrm{EEL\_W} - 4.843\times \mathrm{HVD\_W} \\
&- 0.787\times \mathrm{FSA\_W} + 64.20\times \mathrm{MEA\_B} - 87.89\times \mathrm{MED\_B} \\
&+ 208.42\times \mathrm{MELN\_W} + 372.3\times \mathrm{FEE\_B} + 5.70\times \mathrm{FSA\_B} \\
&+ 10.93\times \mathrm{FSL\_B}
\end{aligned}
$$

$$
\begin{aligned}
\textbf{Correlated Noise} = -7.404 &- 519.9\times \mathrm{MED\_W} + 88.87\times \mathrm{EEL\_W} \\
&+ 14.27\times \mathrm{HVD\_W} - 0.245\times \mathrm{FSA\_W} - 13.03\times \mathrm{MEA\_B} \\
&+ 30.26\times \mathrm{MED\_B} - 17.24\times \mathrm{MELN\_W} + 9.20\times \mathrm{FEE\_B} \\
&+ 0.158\times \mathrm{FSA\_B} - 0.031\times \mathrm{FSL\_B}
\end{aligned}
$$

$$
\begin{aligned}
\textbf{Random Noise} = -3.061 &+ 20.32\times \mathrm{MED\_W} - 12.974\times \mathrm{EEL\_W} \\
&- 3.23\times \mathrm{HVD\_W} + 0.0226\times \mathrm{FSA\_W} + 6.786\times \mathrm{MEA\_B} \\
&+ 3.205\times \mathrm{MED\_B} + 16.04\times \mathrm{MELN\_W} + 5.933\times \mathrm{FEE\_B} \\
&+ 0.174\times \mathrm{FSA\_B} - 0.0486\times \mathrm{FSL\_B}
\end{aligned}
$$

5.1.2 Discriminant Model Application

The type of impairment present on a test pattern is determined by substituting for the ten parameters in the equations above, and then selecting the impairment index with the highest positive value.

5.2 Stepwise Regression Analysis

Stepwise multiple regression procedures were used to develop separate models for each type of impairment that specified the functional relationship between the parameters associated with one type of test pattern and the five different levels of that impairment. The parameters used for the random noise impairment were obtained from the ball pattern, while the models for the other three impairments were based on the rotating wheel pattern.

5.2.1 Blocking Distortion Model

Three parameters computed from the rotating wheel video sequence were required to characterize the amount of blocking impairment present.

$$\mathbf{Level_{Block} = 8.476 - 7.688 \times MELN\_W + 0.692 \times FSA\_W - 4.004 \times FSL\_W}$$

The blocking regression model accounts for 96% of the variance. Figure 2 shows the relationship between the amount of blocking added and the amount predicted by the regression model.

5.2.2 Blurring Model

Five parameters computed from the rotating wheel video sequence were required to characterize the amount of blurring impairment present.

$$\begin{aligned}\mathbf{Level_{Blur}} &= \mathbf{10.23 - 37.62 \times MEL\_W + 2.125 \times MELN\_W + 82.93 \times EEA\_W} \\ &\quad \mathbf{- 4.444 \times FEE\_W - 0.1212 \times FSL\_W}\end{aligned}$$

The blurring regression model accounts for 99% of the variance. Figure 3 shows the relationship between the amount of blocking added and the amount predicted by the regression model.

5.2.3 Signal Correlated Noise Model

Five parameters computed from the rotating wheel video sequence were required to characterize the amount of signal correlated noise impairment present.

$$\begin{aligned}\mathbf{Level_{Corr}} &= \mathbf{7.499 + 44.39 \times MELN\_W - 8.888 \times HVD\_W - 1.11 \times FEE\_W} \\ &\quad \mathbf{- 0.0572 \times FSA\_W + 17.49 \times FSL\_W}\end{aligned}$$

The signal correlated noise regression model accounts for 96% of the variance. Figure 4 shows the relationship between the amount of signal correlated noise added and the amount predicted by the regression model.

5.2.4 Random Noise Model

Four parameters computed from the ball video sequence were required to characterize the amount of random noise present.

$$\begin{aligned}\mathbf{Level}_{\mathbf{Random}} = 9.004 \;&-\; 8.483 \times \mathrm{MELN\_B} \;-\; 48.02 \times \mathrm{HVD\_B} \;+\; 0.1706 \times \mathrm{FSA\_B} \\ &-\; 6.154 \times \mathrm{FSL\_B}\end{aligned}$$

The random noise regression model accounts for 96% of the variance. Figure 5 shows the relationship between the amount of random noise added and the amount predicted by the regression model.

5.2.5 Regression Model Application

Once the appropriate regression model is selected, the parameters computed for the rotating wheel or ball test pattern are inserted and video performance calculated.

5.3 Video Performance Characterization

After computing the parameters specified by the draft standard [1], a two-stage process is employed to determine if video performance has changed over time. First, the parameters are substituted into the four equations and a discriminant index computed for each of the impairment types. The type of impairment present is indicated by the index with the most positive value. Then the parameters are substituted into the regression model for that type of impairment, and the video performance level computed

6.0 Results and Discussion

The discriminant function model was able to correctly identify the impairment present 93% of the time.[3] Of those instances where an incorrect classification

---

[3] When the jackknife procedure in the BMDP 7M [4] program was employed as a cross-validation measure, the discriminant model identified impairment type correctly 92% of the time.

occurred, the impairment added was blocking or signal correlated noise, and the predicted impairment was random noise. In each of these cases, the impairment level was 9. For this level, the amount of impairment added is non-zero, but negligible. It is much less than would be detected by panels of observers in laboratory subjective testing.

Based on the classification provided by the discriminant function model, the appropriate regression impairment model was used to predict the amount of impairment present as a function of the computed test pattern parameters. The amount of predicted impairment present is plotted as a function of the amount of added impairment in Figure 6. The two-stage video performance model is very sensitive to changes in video quality, accounting for 96% of the variance in the data.

The performance levels included in this study were selected based on the results of earlier studies that characterized the relationship between video impairments and customer opinion of quality. What has been referred to as performance levels (9, 8, 7, 5, and 3) nominally correspond to the mean opinion score for the impairment levels on a nine-point rating scale. In subjective tests conducted in our laboratory using typical telco customers we find that the quality of today's commercial broadcast television is equivalent to a MOS of about 8.5, while VCR VHS-SP is rated at a MOS of about 5.2. Although under laboratory conditions people can make very fine judgments about relative video quality, few customers are probably aware of the differences between broadcast and VCR quality under typical viewing conditions. These results suggest that even a very preliminary objective performance model is capable of discriminating differences in video quality performance that are likely to go undetected by many telco customers watching television in their homes.

7.0 Summary and Conclusions

Based on the results of this study it is suggested that with additional research it will be possible to use the T1A1 performance measures to develop an objective model to assess the quality of video services. However, additional testing is recommended before the preliminary model presented in this contribution is implemented in testing equipment. Of particular concern is how well the model will predict the effect of impairments exhibiting qualitatively different perceptual effects (e.g., horizontal lines) than those included in this study.

Despite this caveat, there is reason to believe that an even better objective model of video performance is possible. Improvements in the preliminary objective model could be obtained by including more dynamic test patterns in the study. This would allow additional sophisticated statistical techniques to be employed to develop a more powerful model of video performance.

Finally, the work discussed in the paper has been presented within the framework of determining whether the objective measures specified in the T1A1.5 recommendation might be useful for comparing past and present video performance. However, it was the original goal of the T1A1 standards group to develop a model for the objective assessment of video quality. Still, this goal has remained elusive, and no model of video quality has been proposed that is sufficiently reliable. Additional work toward the original goal now may be warranted. The two-stage modeling procedure used in this study is a promising new approach that may overcome some of the limitations of past objective models of video quality.

8.0 References

1. S. Wolf, "American National Standards for Telecommunications - Digital Transport of One-Way Signals - Parameters for Objective Performance Assessment," contribution to the ANSI Accredited Standards Committee T1, Working Group T1A1.5, document number T1A1.5/95-107R1, May 1, 1995
2. "VIRIS, An Experimental Video Reference Impairment System," ITU-T Contribution COM 12-21-E, Question 22/12, December 1993.
3. C. J. Hurbety, *Applied Discriminant* Analysis, John Wiley & Sons, 1994
4. J. Dixon, *BMDP Statistical Software Manual*, University of California Press, 1992.

Table 1. Nine-point rating scale used in customer perception studies.

Video Quality Rating Scale	
9	EXCELLENT
8	
7	GOOD
6	
5	FAIR
4	
3	POOR
2	
1	UNSATISFACTORY

Table 2. VIRIS Data Input Level Corresponding To Specific Quality Levels

Level	RN	SCN	BLR	BLK
9	1	1	1	1
8	50	12	2	3
7	300	17	3	10
	1600	27	4	30
3	8333	40	5	70

Table 3. T1A1.5 Video Performance Assessment Measures

Wheel Parameter	Ball Parameter	Computed Parameter description
MEA_W	MEA_B	Maximum added motion energy
MEL_W	MEL_B	Maximum lost motion energy
MED_W	MED_B	Average motion energy difference
MELN_W	MELN_B	Average lost motion energy with noise removed
EEA_W	EEA_B	Maximum added edge energy
EEL_W	EEL_B	Maximum lost edge energy
EED_W	EED_B	Average edge energy difference
HVD_W	HVD_B	Maximum horizontal and vertical (HV) to non-HV edge energy difference
RF_W	RF_B	Percent repeated frames
FEE_W	FEE_B	Added edge energy frequencies
FSA_W	FSA_B	Maximum added spatial frequencies
FSL_W	FSL_B	Maximum lost spatial frequencies

Figure 1. Five Levels of Four Impairments Added to Test Patterns Coded at MPEG1 and Four MPEG2 Bit Rates

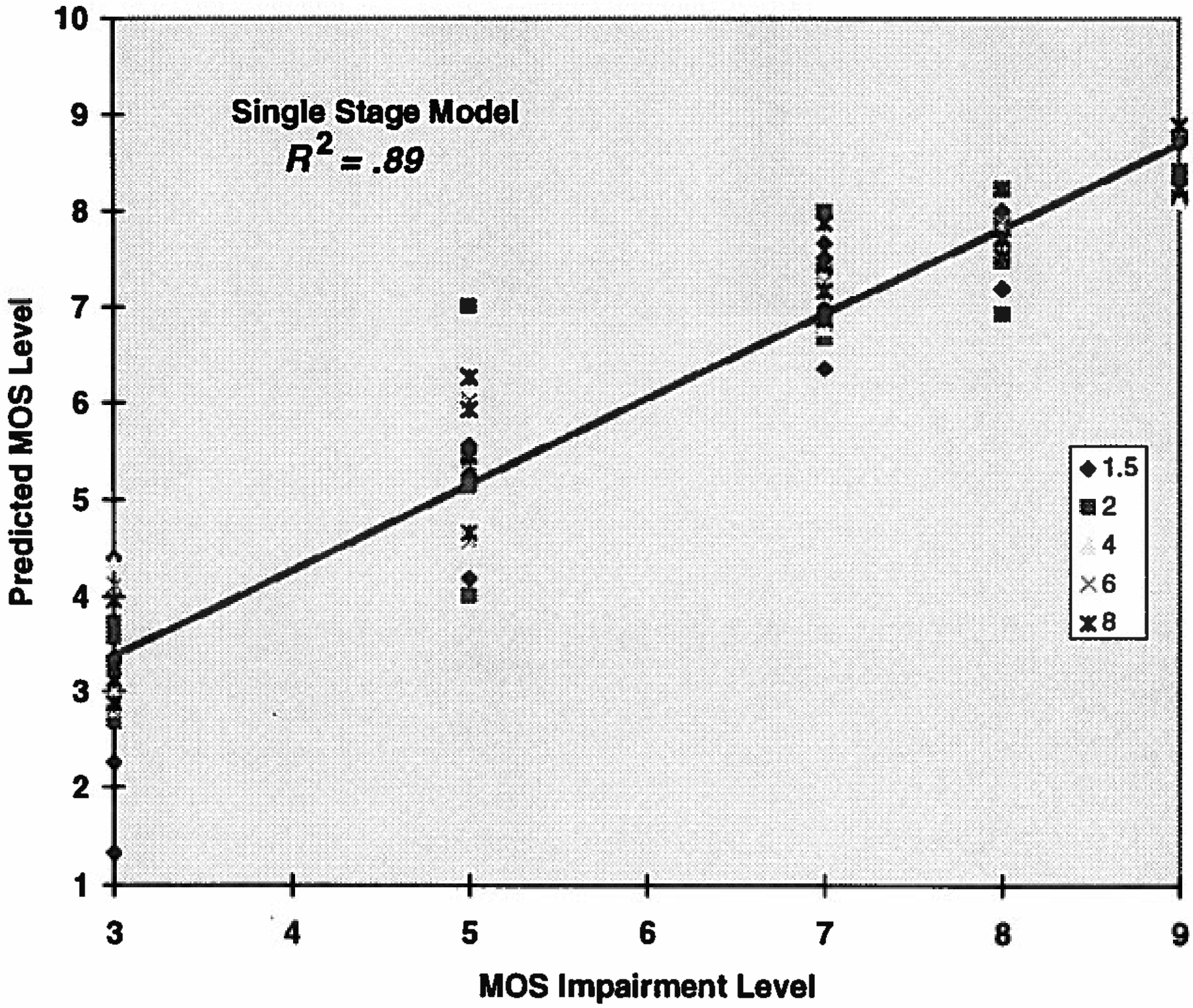

Figure 2. Five Levels of Blocking Added to Test Patterns Coded at MPEG1 and Four MPEG2 Bit Rates

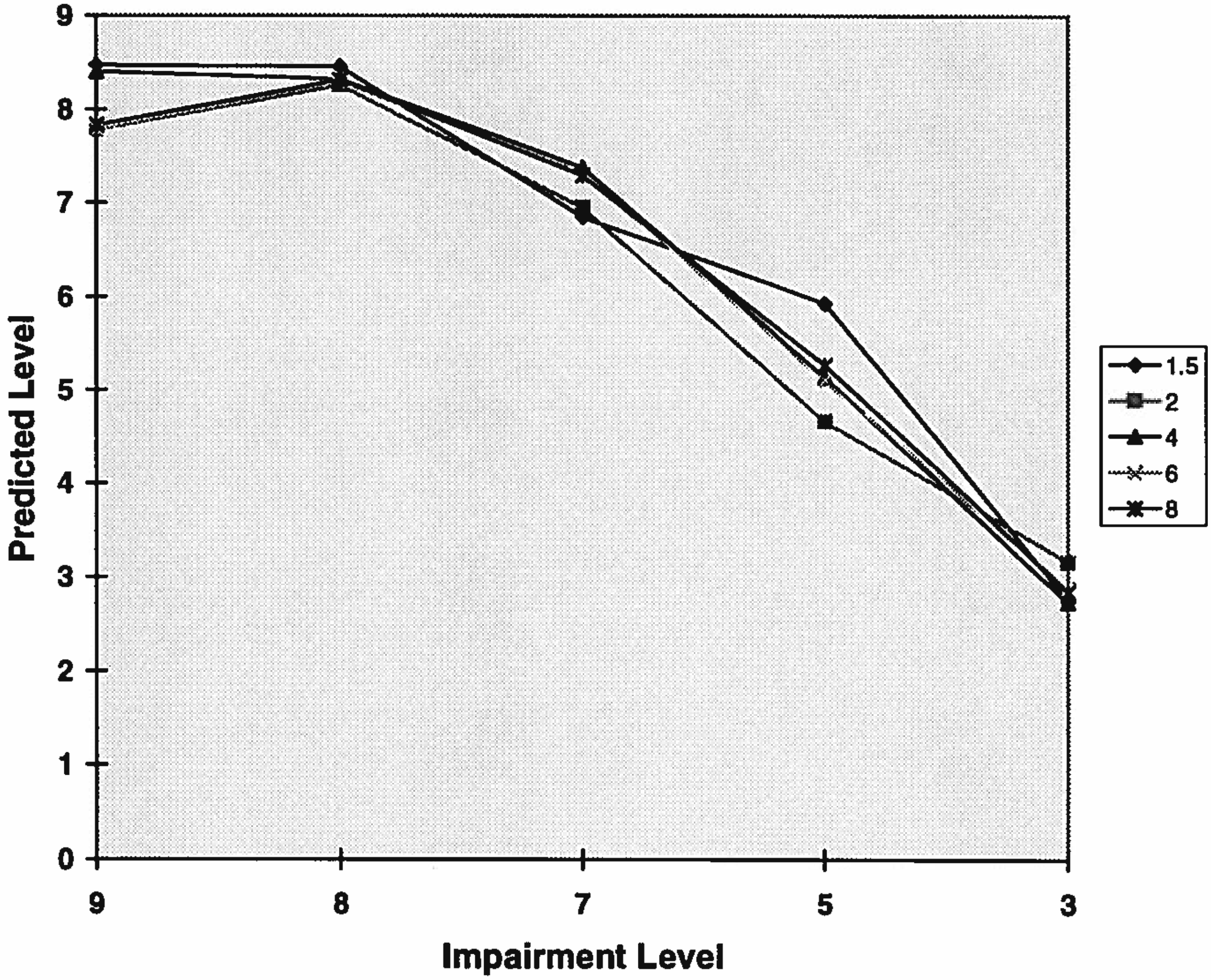

Figure 3. Five Levels of Blurring Added to Test Patterns Coded at MPEG1 and Four MPEG2 Bit Rates

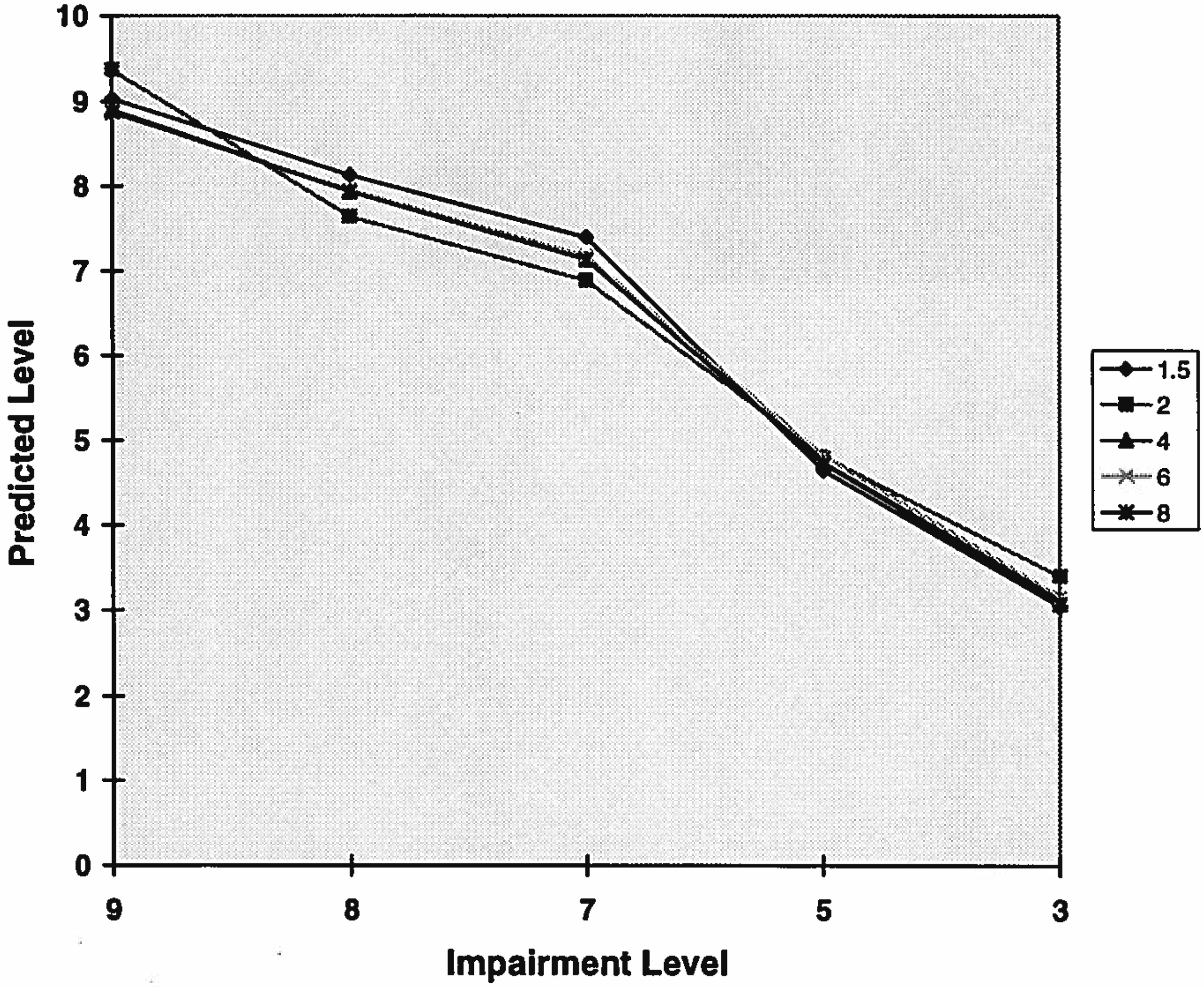

Figure 4. Five Levels of Signal-Correlated Noise Added to Test Patterns Coded at MPEG1 and Four MPEG2 Bit Rates

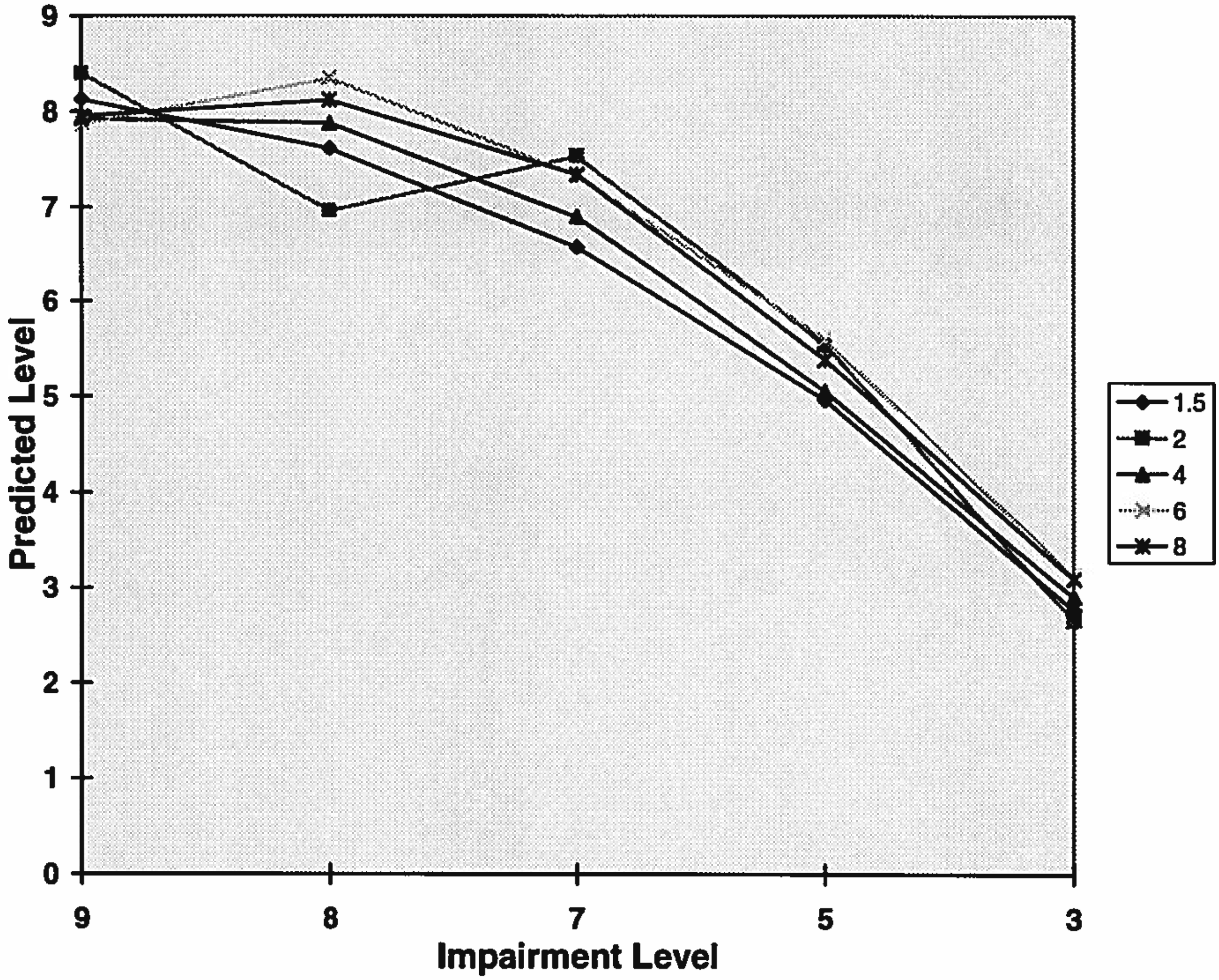

Figure 5. Five Levels of Random Noise Added to Test Patterns Coded at MPEG1 and Four MPEG2 Bit Rates

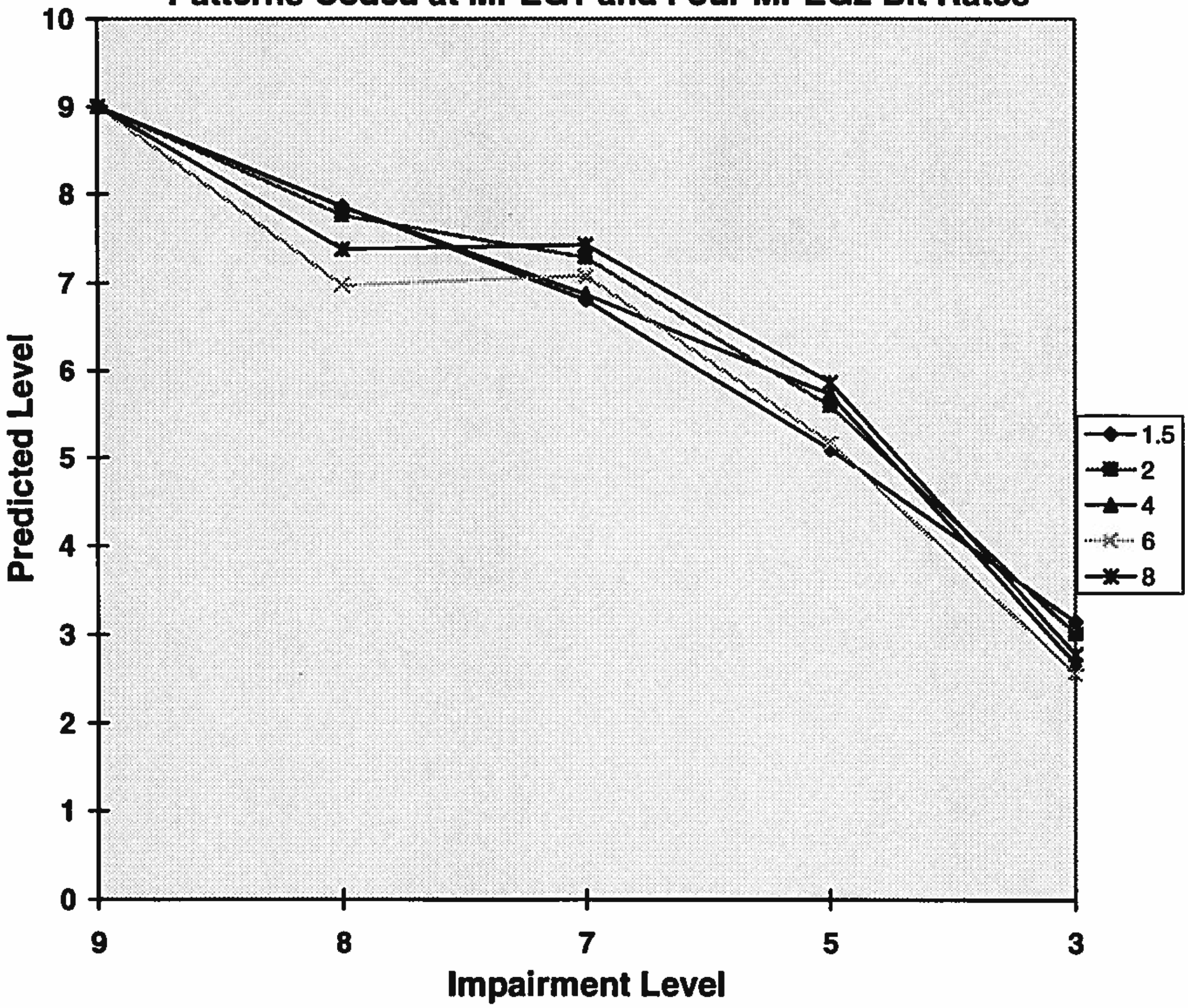

Figure 6. Five Levels of Four Impairments Added to Test Patterns Coded at MPEG1 and Four MPEG2 Bit Rates