COMMITTEE T1 - TELECOMMUNICATIONS
STANDARDS CONTRIBUTION

DOCUMENT NUMBER: T1A1.5/94-124

DATE: April 19, 1995

STANDARDS PROJECT: T1A1.5 - Performance

SUBJECT: Combined A/V Model With Multiple Audio And Video Impairments

SOURCE: Bellcore

CONTACT(S): Bill Cotton
Bellcore, Room NVC 1E-332
331 Newman's Springs Road, Red Bank, NJ 07701
Tel. (908) 758-2510, Fax: (908) 758-4545
E-mail: billc@cc.bellcore.com

ABSTRACT: This is a proposed contribution to ITU-T Study Group 12 which addresses that portion of Question 22 dealing with global audio/video quality evaluation by subjective means. We are seeking review and comment from T1A1.5 prior to submitting it to ITU-T as a Bellcore Contribution. The contribution reports on additional work completed earlier this year on the combined audio/video subjective test model that was reported previously in 1993 and 1994 in Contributions T1A1.5/93-104 and T1A1.5/94-141 (and in corresponding ITU-T, Study Group 12 contributions). This most recent work validates the robustness of the model in the presence of multiple combinations of audio/video impairments.

DISTRIBUTION: T1A1.5

INTERNATIONAL TELECOMMUNICATION UNION

DOCUMENT NUMBER
Date
Original: English

TELECOMMUNICATION
STANDARDIZATION SECTOR

STUDY PERIOD 1993 - 1996
Geneva, 6-14 September 1995

Question:22/12

STUDY GROUP 12 - CONTRIBUTION

Source*: BELLCORE

Title: COMBINED A/V MODEL WITH MULTIPLE AUDIO AND VIDEO IMPAIRMENTS

Abstract: This contribution reports on additional work completed this year on the combined audio/video subjective test model reported previously in ITU-T, Study Group 12 Contributions, COM 12-20-E, December, 1993 and COM 12-37-E, September, 1994. This most recent work extends the application of the model to multiple audio and video impairments.

1. INTRODUCTION

Two previous studies were conducted to determine how audio and video impairments affect overall ratings of the quality of entertainment video services. The first study[1] was completed in 1993 and used random noise in the video channel and temporally correlated noise (T impairment) in the audio channel. The second study[2] was completed in 1994 and used blurring in the video channel and Q distortion in the audio channel. Results of both of these studies indicated that independent ratings of audio-only quality and video-only quality could be combined to predict overall audio/video quality. Multiplicative models using perceived audio-only quality and video-only quality parameters, rather than separate audio-only and video-only impairment levels, were derived in both studies. The models from the two studies were essentially the same and accounted for 98% of the variance in ratings of overall quality.

The purpose of this contribution is to document a third study completed earlier this year to determine the robustness of the model when multiple combinations of audio/video impairments are present. In this study the audio/video impairment combinations of the two previous studies were used along with a new impairment combination of blocking distortion in the video channel and random noise in the audio channel. The test method and test procedures were identical to those used in the previous studies. The goal for this work is to provide information useful in the creation of a new draft recommendation, Rec. P.AVQ - Global audio/video quality evaluation by subjective means, in support of Question 22 which addresses audio/video quality in multimedia services.

* **CONTACT PERSON** Mr. Bill Cotton

Tel.:1 908-758-2510

Fax :1 908-758-4545

2. SUBJECTIVE TEST PLAN AND PROCEDURES

With the exception of the type and number of levels of audio and video impairments tested, the test plan and procedures were essentially the same as used in the two previous studies. The reader is directed to the documentation of the first study[1] for a full discussion of the test plan and procedures; only a brief discussion is given here.

The basic approach used in the test plan was to have nonexpert subjects rate, first, the combined video/audio quality of two television sequences and then in separate test sessions, the audio-only and video-only quality of the same sequences. They judged quality using a nine-point discrete quality scale with alternating numerical categories labelled as: EXCELLENT, GOOD, FAIR, POOR and UNSATISFACTORY. The test subjects were instructed to use entertainment television as their criterion for judging quality.

2.1 Test Plan

The aim of the present study was to test the robustness of the A/V model in the presense of multiple audio and video impairment combinations. The audio/video impairment combinations from the two previous studies were used in the present study along with a new impairment combination for a total of three different audio/video impairment combinations. The impairment combination from the 1993 study consisted of random noise in the video channel and T impairment in the audio channel. The impairment combination from the 1994 study consisted of blurring in the video channel and quantizing noise (Q) in the audio channel. The new impairment combination added in the present study consisted of blocking distortion in the video channel and random noise in the audio channel. These new impairments are discussed in the following subsections. The reader is referred to the documentation for the two previous studies[1][2] for discussion of the other two audio/video impairment combinations.

2.1.1 Video Blocking Impairment

The video blocking impairment was created using a C program that adds simulated impairments to ITU-R 601[3] video images. The program was developed to simulate random noise in the first study[1] in this series of three studies . Simulated blurring distortion was added in the second study[2] and additional coding was added in the present study to simulate blocking distortion. The program is similar to and was originally derived from the VIRIS program[4] that was developed to add simulated impairments to Source Input Format (SIF) video images.

Blocking distortion in the present study was created in a similar manner as described in Reference 4 except that an ITU-R 601 digital image (720 X 480 pels) was used rather than a SIF image (352 X 240 pels). The amplitude levels of the pels in each distorted block of the ITU-R 601 image were also determined in a different manner in the present study than described in Reference 4 for SIF images. For both image types the average value of the 64 pels in the 8 X 8 block is first determined. For SIF blocks, the amplitude of each pel in the distorted block is adjusted to a value half way between its undistorted value and the undistorted average. In the present study, the amplitude of each pel in the distorted ITU-R 601 block is adjusted to the undistorted average.

2.1.2 *Audio Random Noise Impairment*

The new audio impairment added in the present study was random noise, characterized in terms of signal-to-noise-ratio (SNR), in dB. The active speech level of the sound portion of each of the tested television sequences was first adjusted to a level of -24 dBm using a British Telecom SV6 Speech Voltmeter. Gaussian noise was then added to adjust the SNR to desired test values.

2.1.3 *Test Conditions*

The two television sequences used in the previous two studies were again used in the present study. Each was approximately 18 seconds in length. The *Library* sequence is set in a school library and consists of conversation between two young men. The *Street* sequence is set on a busy street and consists of conversation between a young man and young women while walking

Four levels of each audio and video impairment were selected for testing. These levels corresponded roughly to expected quality levels of High, Medium, Low and Very Poor. All of the Video impairment levels were objectively characterized in terms of Peak-Signal-to-Noise-Ratio (PSNR), in dB, averaged across all of the frames of each picture sequence. (See References 1 and 4 for details on the PSNR calculation.) The audio quantizing distortion was objectively characterized by a parameter called Q, which denotes the ratio of speech power to speech-correlated noise power, in dB. The audio T impairment was objectively characterized in terms of T level, a dimensionless unit. The audio random noise was objectively characterized in terms of signal-to-noise-ratio (SNR), in dB.

The objective levels for each of the audio and video impairment types are given in Table 1, below. The audio objective levels for a specific impairment type and level were the same for both television sequences. The video PSNR values for a specific impairment type and level differed for the two sequences and the numbers in the table represent the average for the two sequences. The PSNR differences between the two sequences ranged up to 0.5 db for random noise, up to 1.3 dB for blurring and up to 0.8 dB for blocking.

Table 1: Video And Audio Impairment Levels

Impairment Level	(1) Video / Audio Noise / "T" Impairment		(2) Video / Audio Blurring / "Q"		(3) Video / Audio Blocking / Noise	
(Expected Quality)	PSNR, dB	"T" Level	PSNR, dB	"Q", dB	PSNR, dB	SNR, dB
1 (High)	66.9	100.0	43.3	35.0	55.1	49.0
2 (Medium)	44.4	40.0	32.9	20.0	42.1	44.0
3 (Low)	30.5	20.0	28.2	10.0	33.5	34.0
4 (Very Poor)	26.7	10.0	26.4	5.0	31.2	24.0

For each specific video/audio impairment combination, all combinations of the first three video and audio impairment levels were tested in the combined video/audio portion of the tests resulting in 54 test conditions (3 audio levels x 3 video levels x 3 combination types x 2 sequences). All four audio and video impairment

levels for each impairment type were tested in the separate audio and video portion of the tests. Repeating each condition resulted in 48 test conditions in each separate test (4 levels x 3 impairment types x 2 sequences x 2 repeats). Thus, a complete test consisted of 150 test conditions.

2.2 Test Procedures

Thirty-three nonexpert subjects participated in the tests. They had ages ranging from 18 to 58 and an average age of 36.6. The subjects viewed the test material on a 20-inch television receiver at a viewing distance of 6 times the height of the picture screen (about 6 feet). The audio portion of each sequence was played through an amplifier to an external loudspeaker. All screen and room lighting conditions were the same as in the previous two studies[1][2]. Two groups of three subjects participated at the same time in two different rooms (a few groups contained less than three because of no-shows).

The test consisted of three parts with an instruction and practice session before each of the parts. The subjects were given ample opportunity to ask questions concerning their task before each part began. The first part included 66 combined audio/video test conditions (12 practice (not used in data analysis) and 54 test conditions) and included a short, 3-minute "stretch" break in the middle. After completion of the first part, the subjects were given a 20-minute refreshment break. The second and third parts each consisted of 60 separate audio or video test conditions (12 practice and 48 test conditions) and included a short 3-minute stretch break in the middle. There was a 5-minute break between Parts 2 and 3.

The order of test condition presentation was randomized. The random orders were changed for every two groups of three subjects. In addition, the order of testing audio and video in Parts 2 and 3 was alternated for every two groups of three subjects.

3. TEST RESULTS

The data analysis procedures used in the present study were similar to those used in the previous studies. The mean opinion score (MOS) across the 33 subjects was calculated for each test condition for each of the two picture sequences and for the combined picture results. In general, the MOS differences between the two picture sequences were small for most of the test conditions. The MOS differences between the two picture sequences ranged from 0 to 0.7 with an average difference of 0.18 for the audio-only impairments. The corresponding differences for the video-only impairments ranged from 0 to 0.9 with an average difference of 0.21. For the combined video/audio impairments, the MOS differences between the picture sequences ranged from 0 to 0.9 with an average difference of 0.26.

MOS results, averaged across the two picture sequences and 33 subjects are shown for each impairment level (described in terms of expected quality - see Table 1) in Figures 1 and 2, respectively, for the audio-only and video-only impairment types. The average MOSs ranged from about 1.2 to 7.3 and from about 1.8 to 8.1 for the audio-only and video-only impairments, respectively.

The results of Figures 1 and 2 were used along with the results for the combined audio/video test conditions to develop a model relating the opinion ratings obtained in the separate audio and video impairment tests to the opinion ratings obtained in the combined audio/video impairment tests. As in the previous two studies, the model was based on a least-squares regression analysis of the combined audio/video MOSs as a function of a single multiplicative factor representing the interaction between the separate audio and video MOSs. The results are shown in Figure 3 and the relationship is defined in Equation 1.

$$MOS_{AV} = 0.912 + 0.114 \times (MOS_A \times MOS_V) \quad \text{(EQ1)}$$

Where MOS_{AV} = MOS for combined audio and video impairments,

MOS_A = MOS for separate audio impairments and

MOS_V = MOS for separate video impairments.

The R squared value for Equation 1 is 0.9853, which indicates that the model accounts for about 98.5% of the variance in ratings of overall audio/video quality and represents an excellent degree of fit to the data. Compare Equation 1 with the regression equation shown in Equation 2 (R squared = 0.9806) which was obtained in the most recent previous study using only the video blurring/audio "Q" impairment combination and then compare it with Equation 3 (R squared = 0.98) which was obtained in the first previous study using only the video noise/audio "T" impairment combination.

$$MOS_{AV} = 1.07 + 0.111 \times (MOS_A \times MOS_V) \quad \text{(EQ2)}$$

$$MOS_{AV} = 1.295 + 0.107 \times (MOS_A \times MOS_V) \quad \text{(EQ3)}$$

The regression equations from the three studies are remarkably similar, suggesting that the three models are essentially the same and that any of the models could describe the data from all three studies. The results from the present study combined with those from the previous studies verify that the overall quality of combined video and audio, represented by an MOS, can be expressed as a linear function of the interaction between the separate audio and separate video, represented by a product of their MOSs.

4. CONCLUSIONS

A model developed initially in 1993 with one video/audio impairment combination and refined further in 1994 using another single video/audio impairment combination has been extended in the present study to apply to multiple video/audio impairment combinations. Tests and data analysis in the present study using the two video/audio impairment combinations used in the previous studies in addition to a new, third video/audio impairment combination resulted in a model essentially the same as derived in the two previous studies. The model predicts overall ratings of combined audio/video quality based on independent ratings of audio-only or video-only quality. Its accuracy is such that it accounts for about 98% of the variability in combined audio/video ratings. The robustness of the model is such that it holds up extremely well for various single audio/video impairment combinations as well as for multiple audio/video impairment combinations.

The model is unique because it predicts overall audio/video quality from separate audio-only and video-only quality rather than from separate audio-only and video-only impairment levels. The model may prove to be extremely valuable to system designers since it allows a cost/benefit analysis of how changes in the relative quality of either the audio or video channel affect overall audio/video quality.

Use of the model is presently limited to entertainment television services. Further study would be required to apply the model to other services such as video teleconferencing. It should also be emphasized that the same two picture sequences were used in all three studies of this study series. This was done to maintain continuity from study to study and to also limit each study to a reasonable number of test conditions.

However, it would be desirable to verify the accuracy and robustness of the model with a wider range of source material.

REFERENCES

1. Contribution ITU-T, COM 12-20-E, "EXPERIMENTAL COMBINED AUDIO/VIDEO SUBJECTIVE TEST METHOD," Question 22/12, December, 1993
2. Contribution ITU-T, COM 12-37-E, "EXTENSION OF COMBINED AUDIO/VIDEO QUALITY MODEL," Question 22/12, September 1994.
3. ITU-R Recommendation 601-1, "Encoding Paramaters of Digital Television for Studios," 1982-1986.
4. Contribution ITU-T, COM 12-21-E, "VIRIS, AN EXPERIMENTAL VIDEO REFERENCE IMPAIRMENT SYSTEM," Question 22/12, December 1993.

FIGURE 1
Mean Opinion Scores for the Audio Impairment-Only Conditions

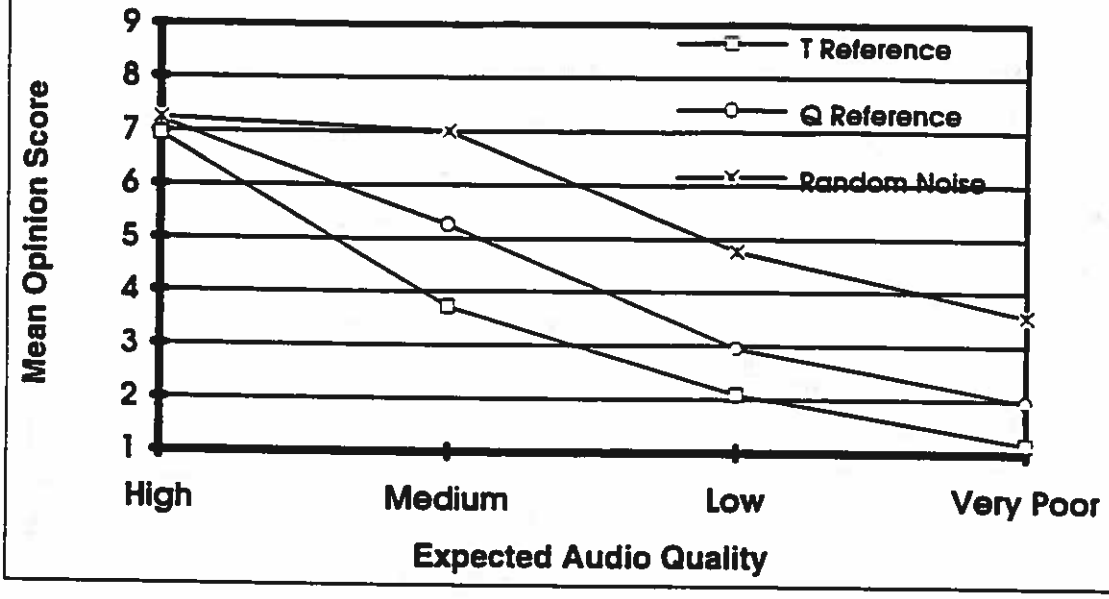

FIGURE 2
Mean Opinion Scores for the Video Impairment-Only Conditions

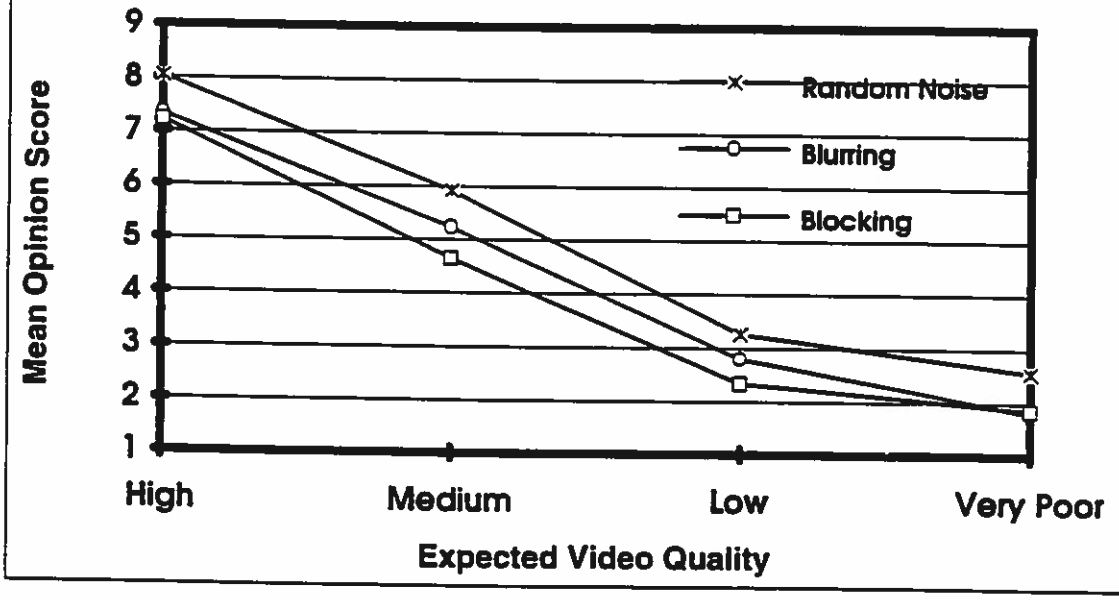

FIGURE 3
Observed Versus Predicted Mean Opinion Scores for Noise/T, Blurring/Q, and Blocking/Noise Combined Video/Audio Impairments