The Perceptual Impact of ATM Cell Loss Errors

Philip Corriveau, Kathleen Myles, Philip Blanchfield, Gilles Gagnon, Ron Renaud and André Vincent
Communications Research Centre
Ottawa, Ontario Canada

Abstract

ATM Networks are rapidly becoming the medium of choice for efficient video transmission. Investigation of the implications of increasing ATM network usage, particularly those for the ever-important area of Quality of Service (QoS), is therefore of great importance. The purpose of this study was to investigate the perceptual impact of ATM cell loss errors, and viewer tolerance to such errors. Twenty-two non-expert viewers rated video material that had been processed using a simulated ATM network at varied cell loss rates (no loss, $2x10^{-6}$, $2x10^{-5}$, $5x10^{-5}$, $2x10^{-4}$). The viewers rated the video material using the Single Stimulus Continuous Impairment Evaluation (SSCIE) method for subjective assessment, which due to its continuous nature allows the use of longer clips while avoiding viewer fatigue. SSCIE is an impairment scale variant of the Single Stimulus Continuous Quality Evaluation (SSCQE) method. The issue of temporal pooling for continuous subjective data was also examined, using a weighted averaging method to achieve a single subjective score for each of the video sequences. Continuous evaluation was found to be a successful method for evaluating the perceptual impact of these types of ATM cell loss errors. Viewer tolerance was observed to drop below 60% beyond a cell loss rate of $2x10^{-5}$. The preliminary results from the temporal pooling suggest that the method has promise for the collapsing of continuous subjective data.

1.0 Introduction

With the advent of high quality digital video compression such as MPEG-2 for Standard Definition Television (SDTV) and High Definition Television (HDTV) as proposed by the Advanced Television Systems Committee (ATSC), as well as the availability of high capacity broadband networks such as Asynchronous Transfer Mode (ATM) networks, interest has grown in the delivery of high quality video services over broadband networks. These networks are becoming more and more available, and attract applications that require transportation of a large number of internet, audio and video services to end users at high bit rates of 155 Mbps and above. The rising network traffic, is already creating data delivery problems such as cell loss, which results from information overflow at network switching points (ATM switches). Investigation of the implications of increasing ATM network use, particularly those for the ever-important area of Quality of Service (QoS), is therefore of great importance.[4,5,7] This study investigated the perceptual impact of ATM cell loss impairments on viewer opinion scores, and the viewer tolerance for such errors. Twenty-two non-expert viewers rated video material that had been processed using a simulated ATM network at five cell loss rates (no loss, $2x10^{-6}$, $2x10^{-5}$, $5x10^{-5}$, $2x10^{-4}$). Cell losses have different perceptual impact, depending on the content of the video material, and can vary considerably over the duration of the sequence. Thus, sequences of several minutes in length were used, as opposed to clips of only 5 or 10 seconds. The use of longer sequences also allowed testing of lower cell loss rates than those examined in previous studies.[1] The viewers rated the material using the Single Stimulus Continuous Impairment Evaluation (SSCIE) method for subjective assessment, which due to its continuous nature allows the use of longer clips while avoiding viewer fatigue. SSCIE is an impairment scale variant of the Single Stimulus Continuous Quality Evaluation (SSCQE) method. A continuous evaluation method has not previously been used to examine these types of errors.

The issue of temporal pooling for continuous subjective data was also examined, using a weighted averaging method to achieve a single subjective score for each of the video sequences.

This paper is organized broadly into three sections; section 2.0 covers the methods of the test, section 3.0 covers the analysis and discussion of the results and section 4.0 offers a summary and conclusions.

2.0 Methods

2.1 Participants

A total of 22 non-expert viewers (15 female and 7 male) were employed for this study. The average age of the non-expert viewer was 22 years old. Non-expert viewers are test subjects who have no professional or extensive personal experience in dealing with video display systems or devices. These viewers were recruited through a local university (Carleton University in Ottawa, Canada) and were screened for visual acuity, contrast sensitivity, and color vision.

2.2 Viewing Room

The experiments were conducted in the Advanced Television Evaluation Laboratory (ATEL), a facility of the Advanced Video Systems Group (RAVS) at the Communications Research Centre (CRC). This room, illustrated in Fig. 1, was specifically designed to meet or exceed all requirements set out in ITU-R Recommendation 500.[2]

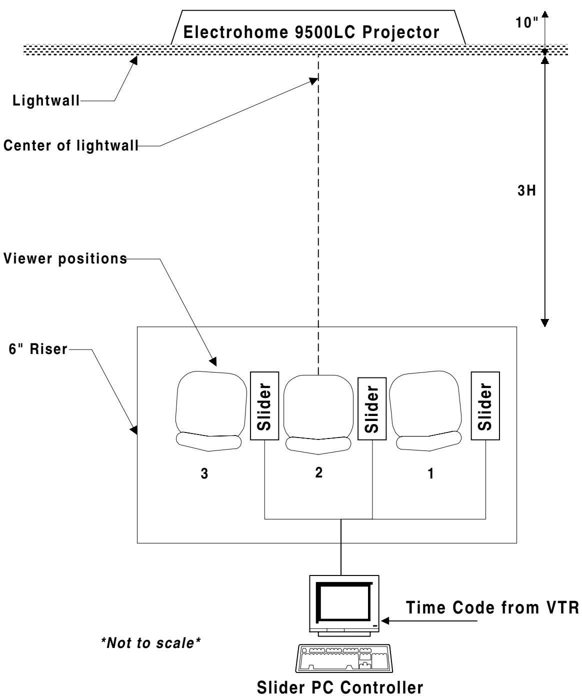

Figure 1. ATEL viewing room configuration

2.3 Video Sequences and Test Conditions

Two HDTV video sequences of the 1080i format (1080(V) x 1920(H) 60 Hz interlaced), encoded in MPEG-2 at a bit rate of 18.5 Mbits/sec using the ATSC standard were used for these evaluations. The first video sequence, called 'Over America" (OvAm), contained 3 minutes of Grand Canyon landscape shots taken from a helicopter. The second video sequence, called 'Mex", was 7 minutes in length. The content of this sequence was approximately 5 minutes of high detail, slow action, head-and-shoulders material followed by short clips of various sporting events (e.g. skiing, football and hockey etc.) and a film excerpt from 'The English Patient", completing the remaining 2 minutes.

2.4 ATM Cell Loss Generation and Block Construction

The method used to generate cell losses was as described by the ISO-IEC Test Model 4 (TM4).[8] The cell loss rates ranged from 2×10^{-6} to 2×10^{-4} in order to generate a minimum of one cell lost over the whole duration of the video sequences, while ensuring that the video sequence was still viewable at the maximum number of losses. The burst length of the losses was kept to a value of one, so that the loss of a cell would not cause the loss of the cells immediately following it. The MPEG-2 Transport Streams (TS) were arranged in an AAL5 Service Data Unit (SDU), as specified in the ITU-T recommendation J.82.[3] Two (2) TS packets are packaged into eight ATM cells, forming the Common Part Convergence Sublayer Protocol Data Unit (CPCS-PDU). Whenever a cell loss occurred, the whole CPCS-PDU was considered lost as per the recommended action from ITU-T Recommendation J.82.[3] Thus, the loss of a single cell causes the loss of not only the cell payload, but of the 376 bytes contained in 2 TS packets as well, resulting in the potential for more degradation of the video data and/or important header data.

The error rates used were not necessarily the most representative of ATM cell loss rates that would be found under normal transmission conditions. Error rates that occur in the real world environment would be 2x10-6 or less. The increased cell loss rates used ensured that the study trials would be of a reasonable duration, taking into consideration viewer fatigue. As cell loss rates increase, the errors increase in frequency, but remain of the same qualitative character.

Thus, an error loss of an identical frame in the same sequence at different cell loss rates would be exactly the same, but the frequency of errors at the increased cell loss rate would be higher. As well, the frequency and severity of the errors in the sequences at the higher cell loss rates could be considered comparable to those found in sequences subject to other types of transmission information loss. Thus, the results from these sequences are still of scientific value.

2.5 Procedures

A maximum of 3 viewers were brought into the viewing room at a time for any given test session. Viewers were asked to rate the visibility and severity of impairments they noticed in the image. They were asked to position a moveable slider at the level that they felt accurately represented the current visibility and severity of impairment. For the presentation of the video material to the subjects, the two sequences, Mex and OvAm, were processed at 5 different cell loss rates (no error, 2×10^{-6}, 2×10^{-5}, 5×10^{-5}, 2×10^{-4}). In terms of possible visible errors the rates are (no error, ~ 6/min, ~ 60/min, ~150/min and ~600/min). These sequences were then randomized and assigned block labels, forming a matrix of 10x2 sequences.

The blocks were then recorded onto an HDTV videotape recorder (HDD1000). The structure of the test sessions is presented in Fig. 2. Before the session began, the research assistant provided instructions to the viewers and explained the block structure. Following the instructions, a three-minute training block was presented to the viewers, to allow them to familiarize themselves with the use of the slider devices used to capture responses, as well as the types of errors that might be present in the actual test trials. Following the training block, each viewer watched a total of 10 test presentations, separated by a short break in between. The duration of the entire experiment was 1.5 hours and all viewers were paid for their participation.

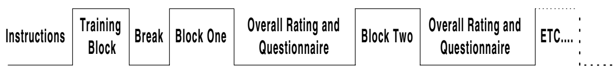

Figure 2. Session structure

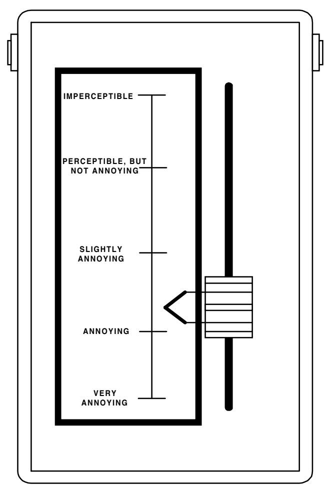

Figure 3. CRC slider device for continuous assessment

2.6 SSCIE

The premise behind continuous subjective assessment is to allow the subjects to rate and react to changes in image quality in a temporally continuous fashion. The standardized method for continuous evaluation is the Single Stimulus Continuous Quality Evaluation (SSCQE) method. Since the main focus of the experiment was the visibility of the picture impairments caused by cell losses, the impairment scale was judged to be more relevant than the traditional quality scale found in the continuous assessment methodology. Therefore, in this test, each viewer had a slider device fitted with the impairment scale shown in Fig. 3, as opposed to the quality scale normally used with SSCQE. We refer to the method as the Single Stimulus Continuous Impairment Evaluation (SSCIE) method. The 5-point impairment scale consists of 4 equally spaced intervals delineated by the adjectives "Imperceptible", "Perceptible but not Annoying", "Slightly Annoying", "Annoying" and "Very Annoying". Subjects were instructed to move the slider continuously during the video presentation so that the indicator was always at the point on the scale that currently and accurately reflected their judgment of the visibility and severity of the image impairments at any given instant. The computer controlling the slider devices polled each device for a position reading every 0.5-second and the VTR for time-code. This corresponds to a reading approximately every 15 frames, given a frame rate of 30 images per second. This frequency of sampling allows the tracking of impairments in the temporal domain, and therefore the observation of viewer response over time and the correspondence of increases in impairment ratings to specific cell loss occurrences.

2.7 Overall Subjective Scores ('Cognitive Summation')

Another important characteristic of the experiment was the demand on the viewer to provide one overall opinion of the presentation as a whole. Presumably, the viewers arrive at this value by mentally summing their opinion scores over the entire sequence duration, recalling from memory the impairments observed, and providing a single value that represents their overall opinion of the presented video sequence in its entirety. In this study, this process is referred to as 'cognitive summation'. This process is similar to that involved in the traditional Double Stimulus Continuous Quality Scale (DSCQS) methodology, where viewers are also asked to give one overall score per presentation. The largest and most fundamental difference between the two methodologies is the length of the sequences used. In this study the viewers were asked to give an overall impression for 3-minute and 7-minute sequences, a significant increase from the short 10-second presentations traditionally used in DSCQS. The comparison between the SSCQE continuous data gathered from the slider devices and the viewer's overall impression of the image is very important, thus this type of overall measure was instituted. In the study, participants were requested to provide this rating at the end of each block by placing the slider's indicator at the location on the scale that reflected their overall opinion of the visibility and severity of the impairments. Then by pushing a button on the side of the slider this value was recorded by the controlling computer. This rating is essential to the temporal pooling of continuous data, and in determining the ability of a viewer to accurately determine impact of errors in a sequence of long duration.

2.8 Viewer Tolerance Rating

The object of examining ATM cell loss errors was to determine the impact of the impairments over the changing video sequence. Therefore, it was decided that it would be of great interest to link this type of study to a personal decision for each viewer: would they watch a program exhibiting cell loss impairments like those they had observed? A questionnaire was designed and implemented. At the end of each block, the viewers were asked answer 'YES' or 'NO' to the following question.

"Based on your judgment of the perceptibility and severity of the impairments in the presentation you have viewed, would you watch a program exhibiting such impairments in your home, as a customer of digital video service?"

This allowed for the determination of a tolerance rating for each cell loss rate, and this was then translated into a failure characteristic, in percent, for each block.

3.0 Analysis and Discussion of Results

3.1 SSCIE Results

It was decided in the design of the study that the original ATSC compressed reference error free sequences, should be presented to the viewers as one of the blocks in the sessions to provide a hidden baseline. Figure 4 presents the results for best, mid-range and worst error rate conditions of both 'OvAm' and 'Mex' sequences. It can be seen from the figure that the viewers were stable in their ratings, with standard deviations comparable to those usually found in DSCQS experiments (between 5 to 20 points of 100).

In the 'OvAm' sequence, visible coding artifacts and softening of the image caused the increase in impairment perceptibility between 125 and 145 seconds. It can be seen based on the results of the sequences presented in Figure 4 that the perceptibility and severity of errors increased rapidly over the course of the five conditions presented.

Also, the difference between the straight average of the SSCQE scores and the overall subjective scores increased in the middle ranges and was tight at the extreme rates. This is caused by the variance in the responses, that is, individual differences with respect to what is considered an impairment, and the level of annoyance associated with it.

3.2 Results of Overall Subjective Scores ("Cognitive Summation") and Temporal Pooling

It is surmised that a subject will have difficulty in estimating the level of impairment of an entire sequence of long duration.[5] The cognitive summation score is analogous to the overall subjective impairment rating assigned to shorter sequences, such as the traditional 10 second sequences used in a DSCQS test. The concept in this experiment was to determine how the cognitive rating would be represented in comparison to the straight average of the SSCQE data, and also to establish a baseline for the development of a proper temporal pooling algorithm of the data which accurately reflected the cognitive score. The differences between the cognitive average and the straight average of SSCQE scores can be seen in Figure 4. The long dashed lines represent the cognitive averages and the straight averages by the short dashed lines. For the "Mex" sequence, the cognitive average is close to the average of the SSCQE data, however, for "OvAm" there is a slight disparity. This is due to the slightly higher level of impairment in the sequence, and the fact that these artifacts occur closer to the end of the sequence. In the OvAm sequence, the largest impairments occurred at the end of the sequence and its cognitive rating is significantly lower than its SSCQE average. This finding supports the hypothesis of recency effects.[5] Recency effects are observed in viewer scores when severe impairments occur in the last few minutes or seconds of a sequence, and the viewer rating is lower than for a sequence with comparable impairments in the middle sections of the presentation. A temporal pooling algorithm must take these effects into account. The greater disparity between the two averages for the OvAm sequence can also be observed in the mid-range condition. Again with increased variation in quality and more periods of lower quality, viewers drop their cognitive opinion scores. The cognitive score is very close to the SSCQE average in both the worst and best case conditions. This is expected, as there is little or no variation left in the data at these error rates because the impairments are either extremely visible and severe, or obviously nonexistant. However, overall correlations indicate that statistically, there is no difference between the "cognitive", overall scores and the averages of the SSCQE data. Using only a straight average of the SSCQE scores over time, a correlation of 0.98 was obtained.

However, a temporal pooling metric was then applied to each sequence, which effectively eliminated the disparity in the mid-range conditions of both the OvAm and the Mex sequences. This metric was devised using documented evidence of recency effects,5 and the fact that the negative scores, or drops in opinion of quality, have a larger bearing on the viewer's overall impression, especially those drops that are large in magnitude. An increased weighting factor was given to negative opinion scores relative to the mean (i.e. an increase in impairment perceptibility and severity), and an even larger weight was assigned to those drops of a magnitude of 10 scale points or more.

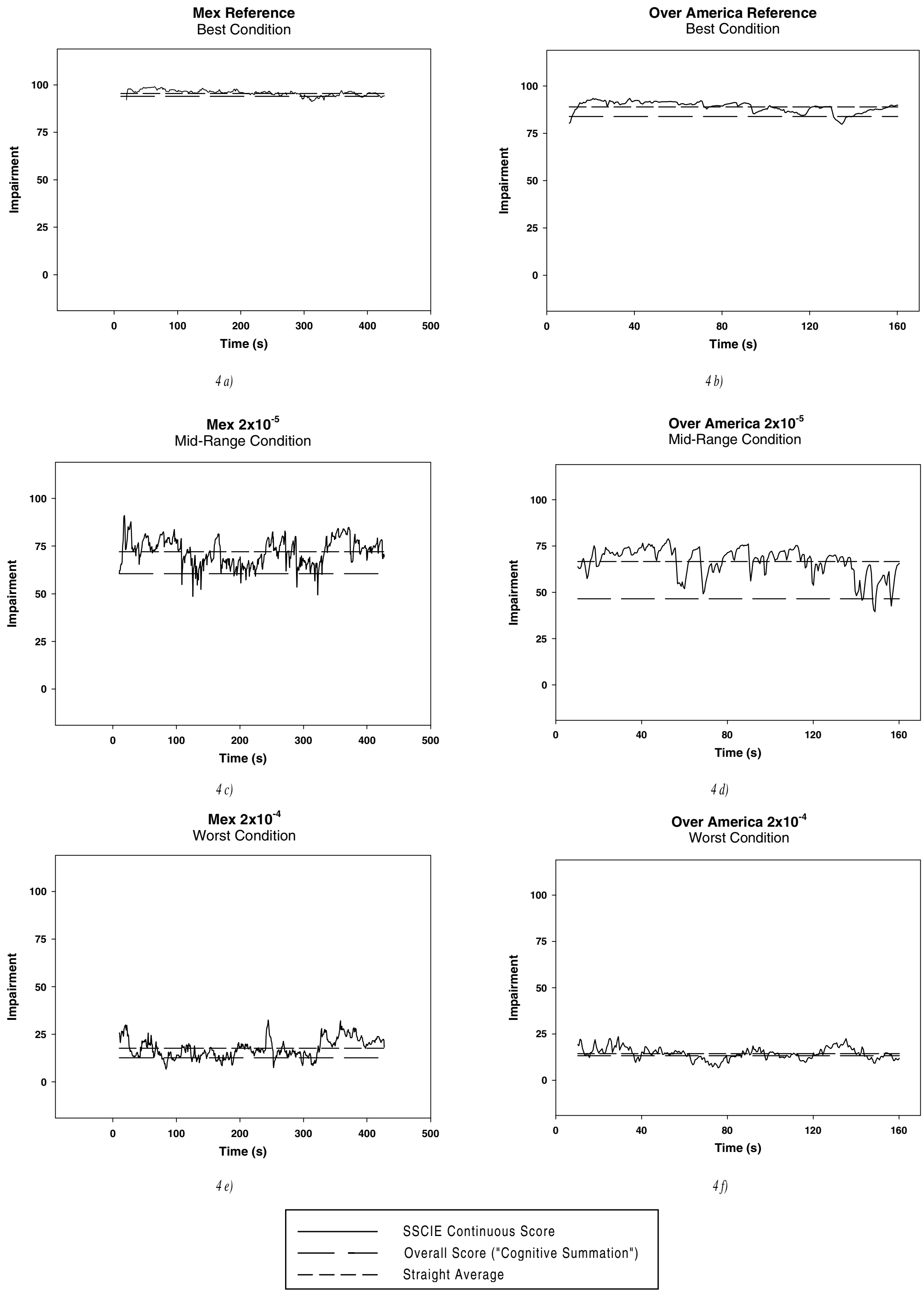

Figure 4, a – f. Subject responses to best, mid-range and worst error rate conditions.

A greater weighting factor was also given to drops that occurred in approximately the last 20 seconds of the sequence. The metric consists of the following steps:

1. Calculate the Mean Opinion Score (MOS) (averaged over N viewers) for each continuous Opinion Score (OS) data point.

$$MOS_j = \frac{\sum_{i=1}^{N} OS_i}{N} \quad (1)$$

2. Calculate the MOS, the average MOS for the sequence.

$$\overline{MOS} = \frac{\sum_{j=1}^{M} MOS_j}{M} \quad (2)$$

3. For each MOS, MOS – $\overline{MOS}$ = Delta (Δ). Then calculate a weighted delta (Δ_w) depending on the magnitude of the delta from the grand mean. This is indicated below using Boolean operators.

$$\begin{aligned} &\text{IF } \Delta \geq 0, \text{ THEN } \Delta_w = 1\Delta \\ &\text{ELSE IF } -10 \leq \Delta < 0, \text{ THEN } \Delta_w = 2\Delta \\ &\text{ELSE IF } \Delta < -10, \text{ THEN } \Delta_w = 6\Delta \end{aligned} \quad (3)$$

4. Then recalculate Δ_w depending on time (t) in seconds.

$$\text{IF } t_i \geq t_m - 20, \text{ THEN } \Delta_w = 2\Delta_w \quad (4)$$

5. Then reapply the weighted delta to its associated MOS.

$$MOS + \Delta_w = MOS_w \quad (5)$$

6. Then calculate a new weighted average for the sequence to compare to the subject's overall score ("cognitive summation").

$$\overline{MOS_W} = \frac{\sum_{j=1}^{M} MOS_{W\,j}}{M} \quad (6)$$

Using this metric, a perfect correlation of 1.00 was obtained. However, these results can only be considered preliminary, since the metric has been tested with this data set only, and thus requires more validation before it can be deemed a valid modeling of all SSCQE data.

3.3 Viewer Tolerance Results

A viewer's perception of picture quality does not necessarily provide an indication whether the viewer would watch the program or not. Yet, viewer tolerance to impairments is an essential component in the definition of Quality of Service. Figure 5 shows the percentage of viewers that would actually watch a hypothetical program that exhibited impairments generated by the varying cell loss rates associated with the test blocks, according to their responses to the viewer questionnaire. It is interesting to note that a loss rate of 2×10-6 is tolerated by viewers but a rate of 2×10-5 causes a rapid percentage drop to below 60 percent. Given a designated cut-off of 50 percent, people would still watch the "Mex" sequence but not "Over America". There is a slight to moderate decline in viewing tolerance between the reference and the first test condition, the slope then takes a cliff effect and continues on a severe downward trend.

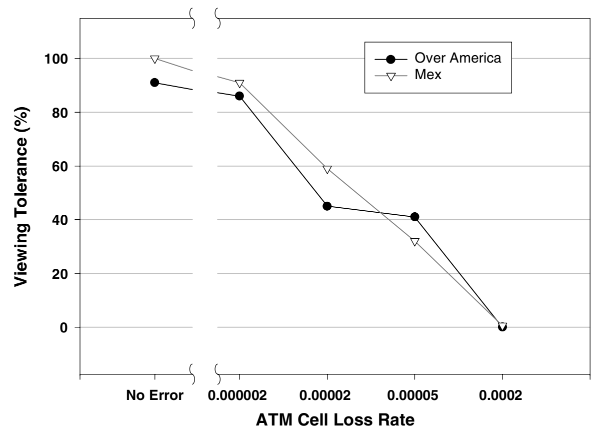

Figure 5. Viewer tolerance by cell loss rate.

The results obtained from the higher cell loss rates used in this experiment may be applicable to the study of other types of transmission errors, which might occur in digital video transmission, even if not due to ATM cell loss. The study of the impact of other types of bit errors or bit loss has not been covered here but may be addressed in another study of broadcast-type transmission errors. It shall then be possible to make a parallel between the impact of each type of error.

4.0 Summary and Conclusions

In this paper, we investigated viewer's reaction to ATM cell loss in a simulated digital video transmission over ATM network.

In this study a modified version of the Single Stimulus Continuous Quality Evaluation (SSCQE) subjective assessment method namely the Single Stimulus Continuous Impairment Evaluation (SSCIE) was implemented. SSCIE in conjuction, with long sequence length, was found to be a successful method for evaluating the perceptual impact of ATM cell loss errors.

Also the concept of asking viewers to rate the overall quality of longer sequences was introduced. Stemming from these results, viewer tolerance was observed to drop below 60% beyond a cell loss rate of $2x10^{-5}$. These results indicate that the majority of viewers will tolerate ATM cell loss up to an approximate rate of $2x10^{-6}$.

In order to further the investigation of a reliable pooling method for estimating the overall quality of longer sequences, a new method was explored. The preliminary results from the temporal pooling implemented here suggest it has promise for the collapsing of continuous subjective data.

Further studies need to be completed to characterize the effects of visual errors occurring on specific frames of the Group of Pictures.

Acknowledgements

The authors would like to thank the management of the Communications Research Centre for support during this research.

References

1. Seferidis, V., Ghanbari, M., Pearson, D.E., *Forgiveness effect in subjective assessment of packet video. Electronics Letters,* **28(21)**, 2013 (1992).
2. ITU-R Recommendation 500-8, *Methodology for the subjective assessment of the quality of television pictures*, 1998.
3. ITU-T Recommendation J.82, *Transport of MPEG-2 constant bit rate television signals in B-ISDN*, 1996.
4. Lu, J., Chatterjee, M., Schwartz, M.D., Ravel, M.K., Osberger, W. M., *Measuring ATM video quality of service using an objective picture quality model*
5. Aldridge, R., Davidoff, J., Ghanbari, M., Hands, D., Pearson, D., Recency effect in the subjective assessment of digitally-coded television pictures, *Proc. IEE Conference on Image Processing And Its Applications* (1995).
6. Hughes, C. J., Ghanbari, M., Pearson, D., Seferidis, V., Xiong, J., *Modeling and subjective assessment of cell discard in ATM video. IEEE Transactions on Image Processing*, **2 (2)** (1993).
7. Bjorkman, N., Latour-Henner, A., Hansson, U., The cell loss process in ATM networks and its impact on Quality of Service, *Proc. IEEE 20th Conference on Local Computer Networks* (1995).
8. Coded representation of picture and audio information, Test Model 4, ISO-IEC/JTC1/SC29/WG11, February 1993.

Biography

Philip Corriveau received his Bachelor of Science Degree with Honours in Psychology from Ottawa's Carleton University in 1990. That same year, he joined what is now Industry Canada's Communications Research Centre (CRC). He served as Senior Research Analyst at the CRC's Advanced Television Evaluation Laboratory (ATEL), evaluating high definition television systems until 1995. He then moved to the CRC's Group for Research in Advanced Video Systems (RAVS) and is currently a Researcher of Subjective Assessment. As a member of the ITU-R WP6Q standards body and as co-chair of the Video Quality Experts Group (VQEG), he investigates subjective and objective methods of picture quality assessment.