

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/253255632>

# Sporadic frame dropping impact on quality perception

Article in *Proceedings of SPIE - The International Society for Optical Engineering* · June 2004

DOI: 10.1117/12.525746

CITATIONS

80

READS

499

4 authors, including:



Catherine Colomes

Orange Labs

17 PUBLICATIONS 789 CITATIONS

[SEE PROFILE](#)



Hocine Cherifi

Université Bourgogne Europe

319 PUBLICATIONS 4,043 CITATIONS

[SEE PROFILE](#)

# Sporadic Frame Dropping Impact on Quality Perception

Ricardo R. Pastrana-Vidal<sup>\*°</sup>, Jean C. Gicquel<sup>\*</sup>, Catherine Colomes<sup>\*</sup>, Cherifi Hocine<sup>°</sup>

<sup>\*</sup>France Telecom R&D, DIH/EQS/MAI, 4 rue de Clos Courtel, 35512 – Cesson Sévigné, France

<sup>°</sup>LIRISIA, Université de Bourgogne, BP 400, 21011– Dijon cedex, France

## ABSTRACT

Over the past few years there has been an increasing interest in real time video services over packet networks. When considering quality, it is essential to quantify user perception of the received sequence. Severe motion discontinuities are one of the most common degradations in video streaming. The end-user perceives a jerky motion when the discontinuities are uniformly distributed over time and an instantaneous fluidity break is perceived when the motion loss is isolated or irregularly distributed. Bit rate adaptation techniques, transmission errors in the packet networks or restitution strategy could be the origin of this perceived jerkiness. In this paper we present a psychovisual experiment performed to quantify the effect of sporadically dropped pictures on the overall perceived quality. First, the perceptual detection thresholds of generated temporal discontinuities were measured. Then, the quality function was estimated in relation to a single frame dropping for different durations. Finally, a set of tests was performed to quantify the effect of several impairments distributed over time. We have found that the detection thresholds are content, duration and motion dependent. The assessment results show how quality is impaired by a single burst of dropped frames in a 10 sec sequence. The effect of several bursts of discarded frames, irregularly distributed over the time is also discussed.

**Keywords:** Frame dropping, quality, streaming video, motion loss, freeze, thresholds, temporal analysis, perception

## 1. INTRODUCTION

Recently there has been an increasing interest in real time audiovisual services over packet networks. The performance of such communication systems is often related to the quality of service (QoS) parameters or to a measure of signal fidelity. However, these approaches are not well adapted to the end-user perception that is the main element of quality assessment in the transmission chain. When considering quality, it is essential to quantify user perception of the received sequence. Subjective assessment methods are used in order to quantify the impact of signal distortion on human quality perception.

The video signal may suffer several degradations at any stage of the transmission chain. Severe motion discontinuities are one of the most common impairments in video streaming. The end-user perceives a jerky motion when the discontinuities are uniformly distributed over time and an instantaneous fluidity break is perceived when the motion loss is isolated or irregularly distributed. Bit rate adaptation techniques, cell loss in the packet networks or restitution strategy could be the origin of this perceived jerkiness. At the source coding stage, temporal down-sampling is one of the most employed techniques for bit rate adaptation; the sequence undergoes a frame dropping operation that affects the motion information. Packet loss or jitter could cause a sporadic or non-uniform frame discarding at the decoding process because of the buffering time limit<sup>1</sup>. The last picture that was received is then displayed until a new image has been reconstructed. The end-user then perceives a frozen image followed by an abrupt displacement of the objects. In the following text, we will use the term *temporal discontinuity* as a perceptual synonym of a dropped picture burst.

In this paper we present a psychovisual experiment that was performed to quantify the effect of sporadically dropped pictures on the overall perceived quality. First, we measured the perceptual detection thresholds of generated temporal discontinuities. Then, we estimated the quality function in relation to a single burst of dropped frames for different durations. Finally, we performed a set of tests to quantify the effect of several impairments distributed over time.

Four to six original video contents were selected: sport, entertainment, news, advertisements clips and two classic test sequences. The sequences were all 10 sec long to avoid the forgiveness effect<sup>2</sup>. In our approach, a 10sec presentation time is long enough to deal with time varying quality because we have inserted artificially-controlled frame dropping covering a broad range of impairments.

A four-alternative forced choice method<sup>3</sup> was employed to measure the thresholds. This psychophysical method, also called identification, achieves a high efficiency accruing from minimizing the possibility of blind guessing. The observers are asked to identify which, of 4-alternative sequences, contains the impairment.

The quality ratings from 20 subjects were gathered using the new MSCQS<sup>4</sup> (Multi Stimuli using a Continuous Quality Scale) method developed by the EQS laboratory, France Telecom R&D. This method was submitted to the ITU-R study group 6Q (Quality assessment) and it was standardized by The European Broadcasting Union (EBU). The MSCQS shows a high intra-inter laboratory repeatability and it is well adapted for multimedia quality assessment.

The psychovisual study has been organized in two main parts: visibility and quality impairment. The study of quality was organized in three experiments that are detailed in section 3.

## 2. VISIBILITY EXPERIMENT

### 2.1. Apparatus

The test was carried out in a luminance-controlled environment. The background luminance conditions were based on the ITU-R BT.500<sup>5</sup> recommendation. All videos were stored and displayed on the PC station. The computer CRT was a NEC Multisync FE750+, 17 inches. The maximum luminance of the screen was set to 70 cd/m<sup>2</sup> using the pluge test signal following the ITU-R BT.814<sup>6</sup> recommendation. The distance between subject and computer screen was approximately 60cm. This is a distance frequently found in home and professional computer viewing situations<sup>7</sup>. The SEOVQ<sup>8</sup> (Subjective Evaluation and Optimization of Video Quality) software tool for quality evaluation of multimedia images was used to perform the detection test. The software tool can play audiovisual sequence using proprietary or non proprietary formats and gather participant's answers by means of an evaluation interface.

### 2.2. Stimuli

Six original video contents were selected: sport, entertainment, news, advertisements clips, and two classic test sequences: Barcelona and Mobile & Calendar. The sequences were all 10 sec long. The image format was CIF, RGB 24 bits and 25fps. The consecutive dropped pictures were placed at different motion contexts. The five selected sequences cover a broad range of video contents and motions in terms of complexity and speed terms. Only the advertisements clip presented scene cuts but the dropping process was done far from the scene changes to avoid a temporal masking effect<sup>9,10</sup>. The temporal distance from the scene change was greater than 280ms.

News content is a scene with very low motion consisting of a talking head. For a short moment a man is crossing the news room in the background plan. The Sport content is a basketball play with fast players and camera motion. The sequence contains only one shot. For the Entertainment content we have selected a motorcycle race; the camera follows a group of motorcycles and tracks one of them at the end of the scene; medium and fast motion characterize the sequence. The advertisements clip is a camera moving down in a panoramic scene, there is also a camera tracking a snow man walking on the street followed by an other shot with a fast pan to a house door and a man leaving a restaurant. Barcelona is a test sequence with low motion objects, camera zoom-out and it is color saturated. Mobile is a critical video because of the different kinds of motions present in the same shot; a calendar and a little ball do not have constant speed.

The discontinuity caused by the frame dropping process was inserted in the reference video sequences by discarding  $n$  pictures from the temporal index  $p$  and repeating  $n$  times the picture corresponding to the temporal index  $p-1$  (Fig. 1). Thresholds for detection were evaluated over a range of impairment durations: 80, 120, 160, 200 and 240ms.

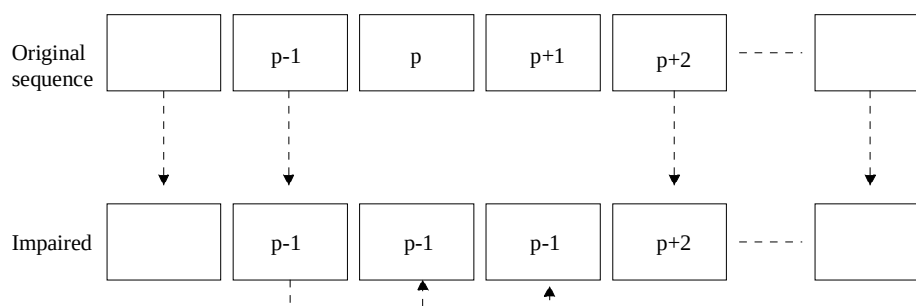


Fig. 1. Frame dropping process.

### 2.3. Observers and method

Results were obtained on 20 subjects ranged in age from 21 to 36 years. All had normal or corrected visual acuity. Participants were paid for their participation. The four-alternative forced choice method<sup>3</sup> was employed to evaluate the thresholds. The observers were asked to identify which of 4-alternative sequences contained the impairment. Each sequence was displayed in a random ordering. The masking test was organized in two sessions of about 10min each.

### 2.4. Visibility experiment results

The first goal of our study was to measure the visibility of a single burst of dropped frames. The probability of detection as a function of the discontinuity strength was calculated in order to characterize this visibility. A threshold for detection was defined as the burst duration such that the generated impairment was seen by 62.5% of the participants. The so defined threshold corresponds to the middle of the detection rate between 25 to 100%. It is assumed that the probability of detection tends to be 0.25 as the impairment strength tends to be zero because of the 4-alternative forced choice. When the impairment strength is not noticeable all of the four alternatives have the same probability to be selected as the right answer.

Fig. 2 shows the probability of detection as a function of burst duration by contents. The fluidity break in the news video was as visible as the discontinuity introduced in the motorcycle race even though the talking head exhibits lower motion. Barcelona, Advertisements clip and Mobile generated similar results even if the camera and object motion were different. We can see that the assessors detected temporal discontinuities with greater ease for fast human motion than for other contents. We could assume that detection rate is related to the apparent motion strength but the motorcycle race sequence does not confirm this. The difference in probability terms is more evident for a burst duration of 80ms. This apparent sensibility to the human motion on quality assessment was also found in the context of low bit rate video<sup>11</sup>. Psychology research in perception of biological motion could explain this phenomenon. There is evidence that people has robust mental representations of human motion<sup>12</sup>. For instance, one is able to anticipate human actions based on reduced information of starting or precedent posture. Some findings show that perception of biological dynamics is not the result of the human experience but an intrinsic capacity<sup>13</sup> of the visual system. The violation (miss match) of biological representations seems to be the origin of the sensibility to the human motion discontinuities.

Fig. 2 shows the percentage of correct detections and the corresponding curve fitting all contents. The results of the overall detection task were fitted using a logistic function:

$$y = ((p_{\max} - p_{\min}) / (1 + (b/x)^s)) + p_{\min} \cdot$$

Based on the above defined threshold, we can say that the discontinuity caused by one discarded-repeated picture is visible. The unequivocal detection rate (100%) is attained for a burst of dropped images equivalent to 200ms. The question now is how these discontinuity strengths, close to the detection threshold affect quality. We have therefore been studying the effect of sporadic frame dropping on the end-user quality perception.

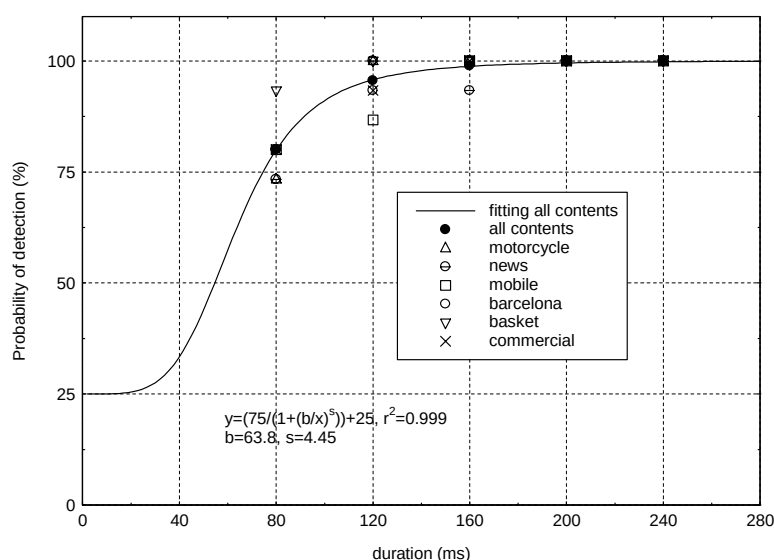


Fig. 2. Probability of detection as a function of burst duration: results by content and the fitting by logistical function.

### 3. QUALITY EXPERIMENTS

The main goal of our study was to characterize the effect of sporadically dropped frames on perceived quality under several controlled conditions. The psychovisual experiments were therefore focused on the effect of isolated temporal discontinuities and on the impact of several discontinuities distributed over time.

In our approach, we believe that for a perceptual characterization of this temporal degradation it is essential to reduce the number of perceptual variables. The quality study was therefore organized in three experiments. First, the impact of a single burst of discarded pictures was evaluated as a function of burst duration. Second, we have examined how quality is affected by a reduced number of discontinuities for constant burst duration. Finally, the effect of a reduced number of discontinuities combining two burst durations in the same profile was evaluated. Impairment conditions for every test are detailed in following sections.

#### 3.1. Apparatus

The test apparatus for the quality assessment experiment was the same as the material described in the visibility experiment. The SEOVQ<sup>8</sup> software tool was used to carry out the test. The evaluation and display interface is showed in Fig. 3. For each video content the observer is asked to assess explicit reference and N test conditions using a software slider placed on the right side of the evaluation interface. The buttons under the display zone are used to select one of the test conditions. The selection buttons are randomly related to test sequences in order to avoid ordering effect.

#### 3.2. Stimuli for quality tests

Four original video contents were selected: sport, entertainment, news, and the classic test sequence Mobile & Calendar. The image format was CIF, RGB 24 bits and 25fps. The sequences were all 10 sec long to avoid the forgiveness effect<sup>2</sup>. Artificially-controlled impairments covering a broad range of qualities were inserted.

The impaired sequences presented discarded pictures placed at different motion contexts. The duration of the impairments was greater than or equal to the detection thresholds and they were selected with enough perceptual distance. Spatial degradation was not introduced in order to avoid tradeoff between acuity and fluidity in the quality assessment. We have organized a quality experiment in three tests as described in the following subsections.

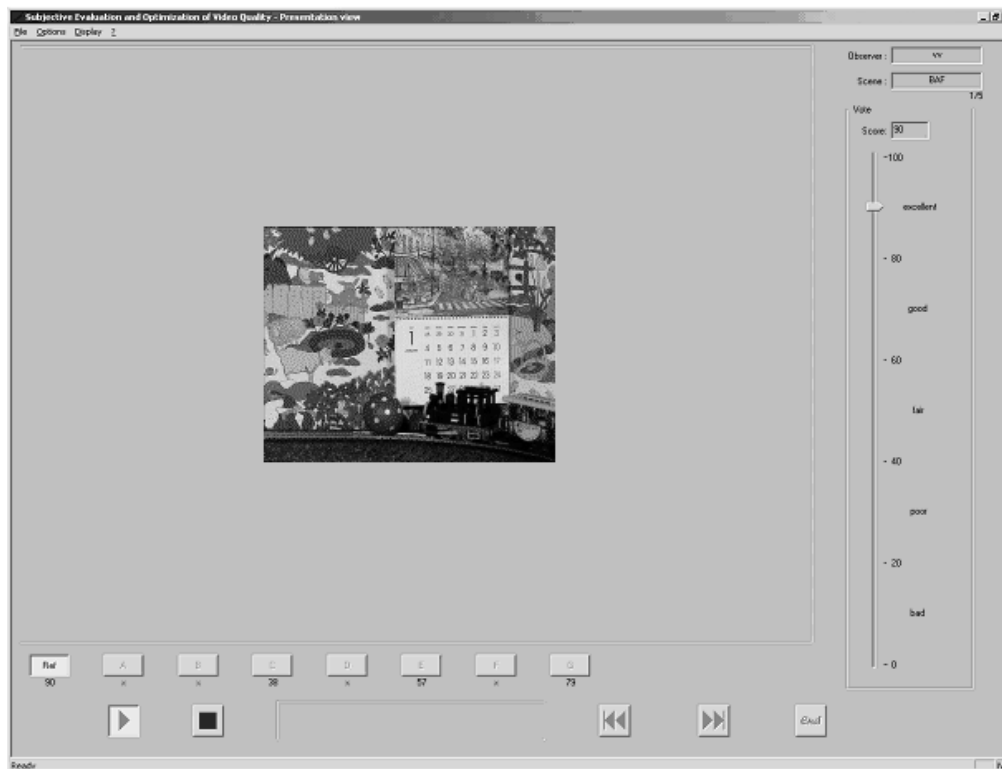


Fig. 3. Quality evaluation interface.

### 3.2.1. Experiment 1

The impact of a single burst of discarded pictures was evaluated as a function of burst duration. In order to cover a broad range of possible situations we have selected values from durations close to the detection threshold up to 50% of information loss: 80, 160, 280, 920, 1640, 2280 and 5040ms.

### 3.2.2. Experiment 2

A second experiment was conducted to quantify the effect of several impairments distributed over time. To do that, we have examined how quality is affected by a reduced number of discontinuities (1, 3, 5, 8) for constant burst duration. Two tests were carried out, the first for a burst duration of 160ms and the second for 280ms.

In both tests the variation on quality ratings as a function of the motion context was also observed. A single burst of dropped frames was placed at different temporal positions. The resulting temporal profiles are shown in Fig. 4.

### 3.2.3. Experiment 3

Finally, a third experiment was conducted to study how quality is impaired by a reduced number of discontinuities combining two burst durations in the same profile (Fig. 5). The number of generated discontinuities was 1, 2, 3, 5, 8 and 15. The employed strengths were 160ms and 280ms; these strengths were greater than the detection threshold and had enough perceptual distance. An intermediate anchor was added: 8 discarded burst of 280ms. This anchor can be also found in experiment 2.

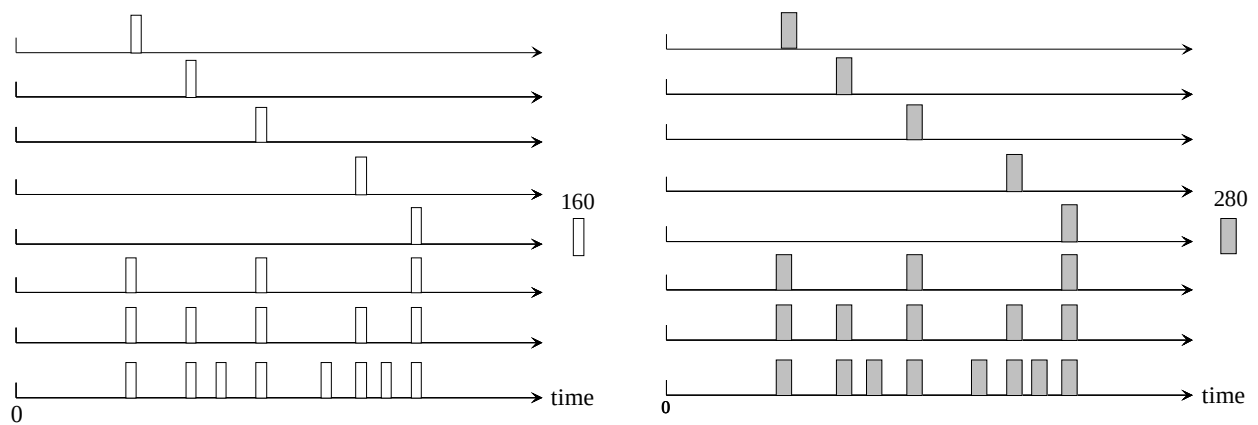


Fig. 4. Frame dropping profile for experiment 2. The bars in the profile represent consecutive dropped frames over time, in a) burst duration of 160 ms and b) 280ms.

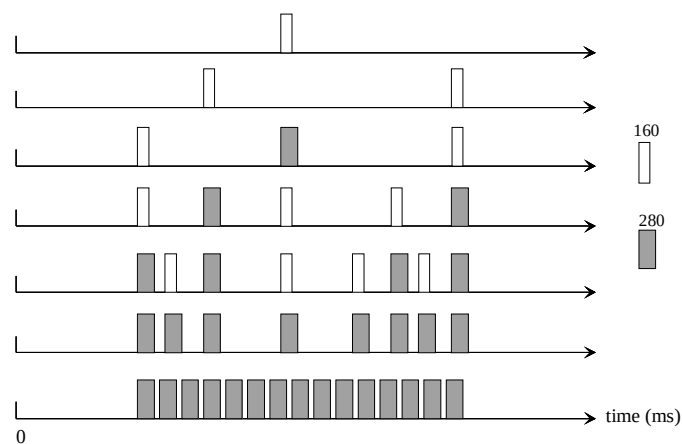


Fig. 5. Frame dropping profile for experiment 3. The bars in the profile represent consecutive dropped frames over time, two burst duration was used: 160 ms and 280ms.

### 3.3. Observers and method for quality tests

The quality ratings from 20 subjects were gathered using the new MSCQS<sup>4</sup> (Multi Stimuli using a Continuous Quality Scale) method developed by the EQS laboratory, France Telecom R&D. This method was submitted to the ITU-R study group 6Q (Quality assessment) and it was adopted by The European Broadcasting Union (EBU). Over the past 3 years, FTR&D has been involved in quality studies of mobile telecommunication and high speed internet services. The new method has provided reliable and stable results. The MSCQS also showed a high intra-inter laboratory repeatability and is well adapted to multimedia quality assessment.

The MSCQS was derived from the DSCQS<sup>5</sup> (Double Stimulus using a Continuous Quality Scale) method but it has the advantage of being more appropriate for the assessment of a wide range of qualities. A numerical continuous scale (0-100) for rating overall quality is used. This scale is related to five quality categories (bad, poor, fair, good, excellent) that are uniformly distributed. The method uses two high quality anchors: an explicit and a hidden reference (Fig. 6). The hidden reference is useful to assess the intrinsic quality of original sequence and it may be an indicator of the task effect.

In this method the assessor is able to switch between the N algorithms (in subjective test language, an algorithm means a signal degradation or image processing to be assessed); they are also able to change the current scores as many times as they want. This flexibility of access leads to an implicit comparison process increasing stability ratings. An instance of test organization may be seen in Fig. 6. In order to increase stability we have also added a low quality anchor that was common for most of the quality tests.

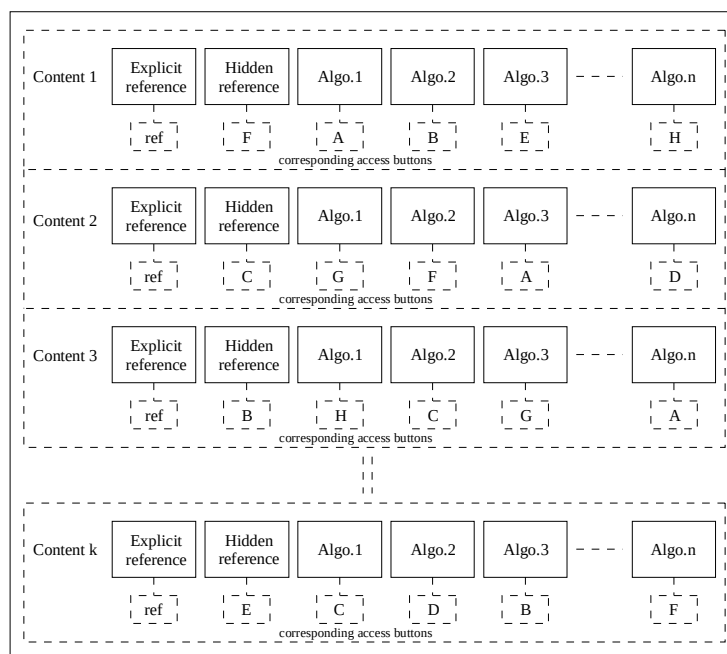


Fig. 6. Test organization in the MSCQS method.

### 3.4. Quality results

The data of the experiment consists of a mean opinion score (MOS) for each of the stimuli calculated after a pre-screening of participants. Confidence intervals are not plotted for clarity proposes.

#### 3.4.1. Quality vs discontinuity duration: experiment 1

With regard to the mean opinion scores over all contents (Fig. 7), subjects showed significant negative reaction for discontinuities of 160 and 280ms. Quality function exhibits a fall in ratings of 1 MOS category. After this range, quality function shows a slower decreasing rate. It is interesting to note that for the worst condition, 5040ms representing 50% of the presentation time, quality assessment is placed in the middle of the *poor* category.

After the visibility experiment, we were interested to quantify how quality is affected by a short dropping of 80ms. The results of assessment show a reduced impact on quality of 8 points. The perceived quality is still considered as excellent. In general, the results of assessment exhibit an inverted S-shape behavior.

Fig. 8 shows the assessment results by content. In the range of 80 to 920ms basketball and advertisements sequence were the most affected by discontinuities. They present fast human motion and fast camera panning respectively. The ratings difference is higher for 80ms. In general, quality variations as a function of contents are not greater than a half of a subjective category.

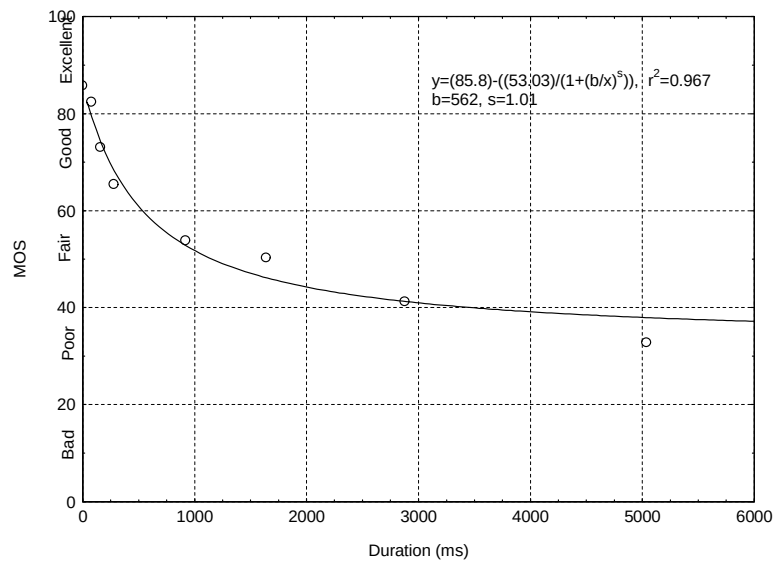


Fig. 7. MOS as a function of discontinuity duration computed over all contents.

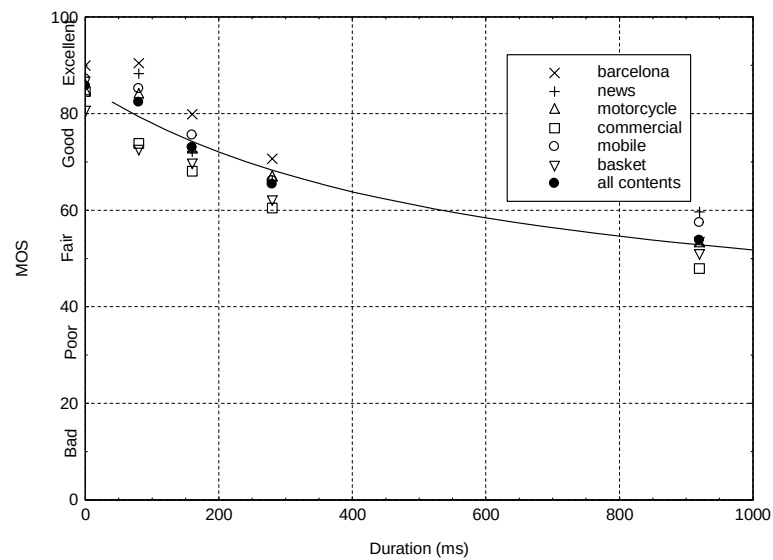


Fig. 8. MOS as a function of discontinuity duration computed by content. For clarity proposes data were plotted up to 920ms of burst duration.

### 3.4.2. Quality vs reduced number of discontinuities

#### 3.4.2.1. Experiment 2 results

Four test sequences were used in this experiment in order to reduce the overall experiment time. Based on the results of experiment 1, we have selected the contents that represent extreme testing conditions: basketball, mobile, motorcycle, and news.

The variation on quality ratings as a function of the different temporal (motion) context can be seen in Fig. 9. The semantic position of a noticeable burst of dropped frames has an influence on quality ratings. However, these variations are not greater than a half of an MOS category. This finding is valid for both 160 and 280ms burst durations. In addition, it is interesting to note that the mean opinion scores for a burst of 160ms and 280ms are consistent with those reported in experiment 1. In fact, the absolute differences respectively are 2.62 and 3.52 points.

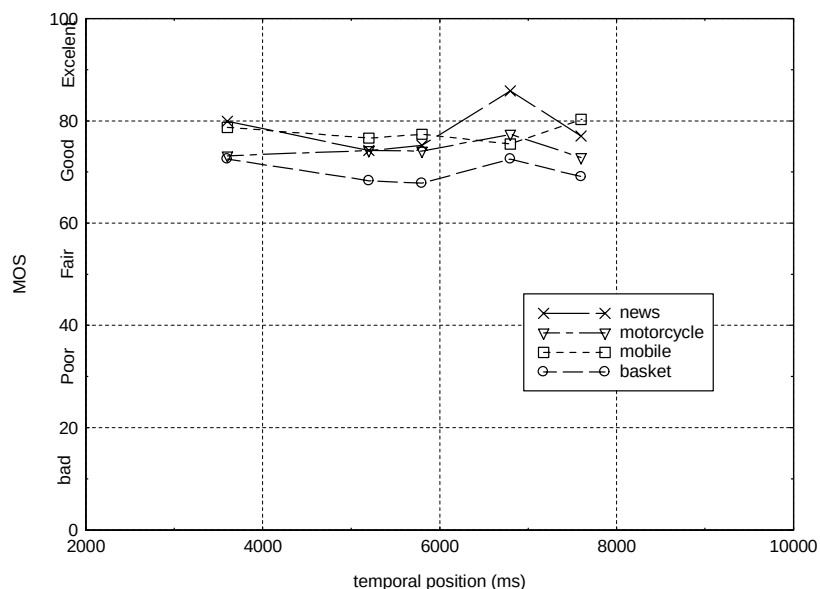


Fig. 9. Mean opinion scores as a function of different temporal positions.

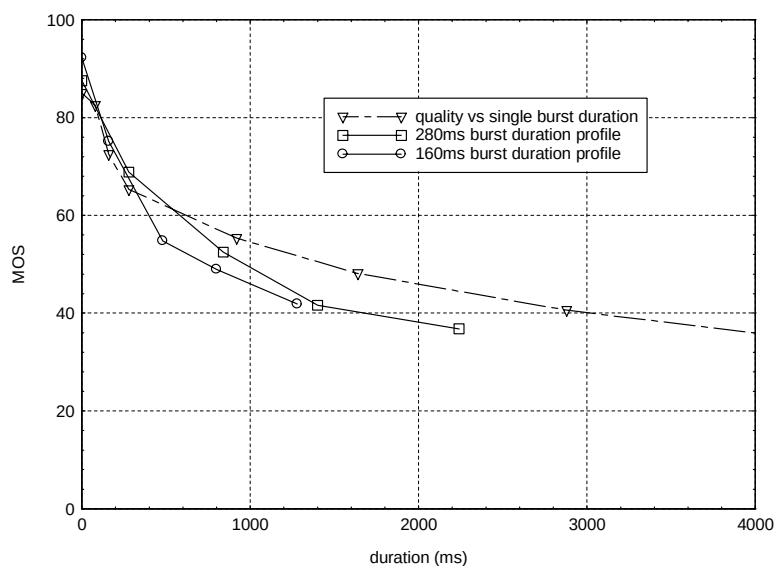


Fig. 10. Mean opinion scores vs cumulated discontinuities duration of the profiles in Fig. 4. For comparison purposes, quality vs discontinuity duration (experiment1) is also plotted.

Fig. 10. shows how quality is affected by a reduced number of discontinuities (1, 3, 5, 8) for a constant burst duration. Mean opinion scores are plotted as a function of the cumulated duration (addition) of N discontinuities. In the figure results of quality vs discontinuity duration (experiment1) are also plotted. Subjects seem to be more annoyed by several short discontinuities than by a un single burst of long duration. For instance, quality drops to 40 either with 8 bursts of 160ms or with a single burst of 2880ms.

### 3.4.2.2. Experiment 3 results

Finally, a third experiment was conducted to study how quality is impaired by a reduced number of discontinuities combining two burst durations in the same profile. The results are shown in Fig.11. The number of generated discontinuities was 1, 2, 3, 5, 8 and 15. The employed strengths were 160ms and 280ms.

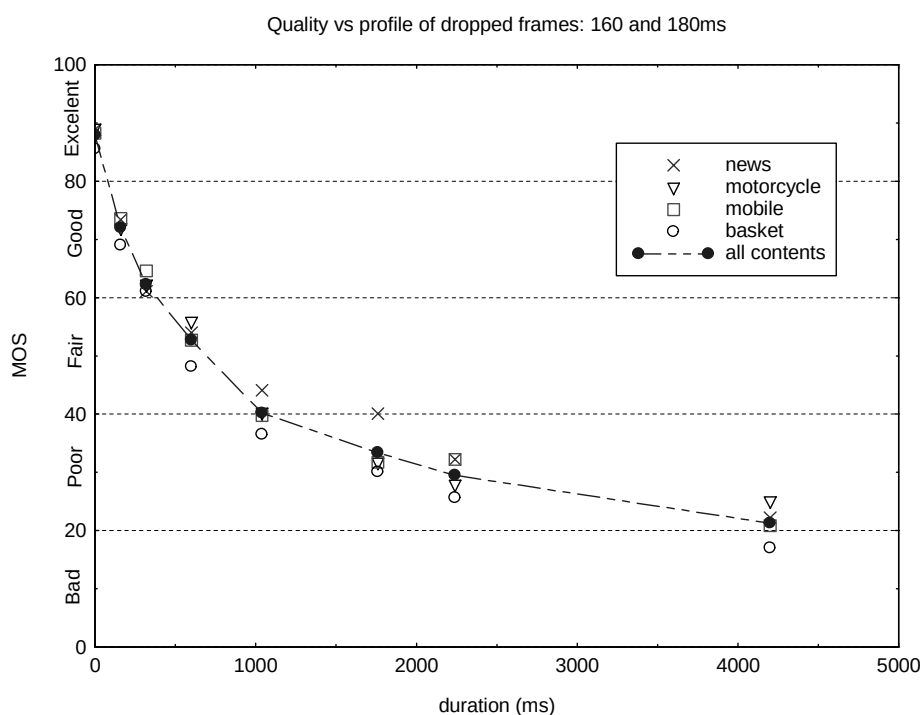


Fig. 11. Quality rating vs reduced number of discontinuities combining two burst durations in the same profile. The dotted line links the mean opinion score computed for all contents.

In general, sport content was systematically the sequence the most affected by frame dropping. The other sequences were impaired in a different manner depending on the artifact profile.

The talking head video was more impaired than the motorcycle sequence for certain profiles even though it has very low motion. This situation may be explained by the fact that some dropped frames were extracted at eye and lip movements. This is an instance of how annoyance is highly content dependent. It also reflects that a perceptual characterization has to be made by content type. In general, the variation on MOS ratings as a function of the content is not greater than a half of a perceptual category (10 points).

### 3.4.3. Comparing results

In Fig. 12 we can compare the results from three experiments. It shows that both duration and density of discontinuities have a non linear effect on user annoyance. For instance, a mean opinion score could be obtained by: a combination of 3 burst of 160ms and 2 burst of 280ms; a profile with 8 discontinuities of 160ms; or 5 bursts of 280ms.

Data shows that quality assessment follows a non linear decreasing function that depends on the cumulated duration of discontinuities.

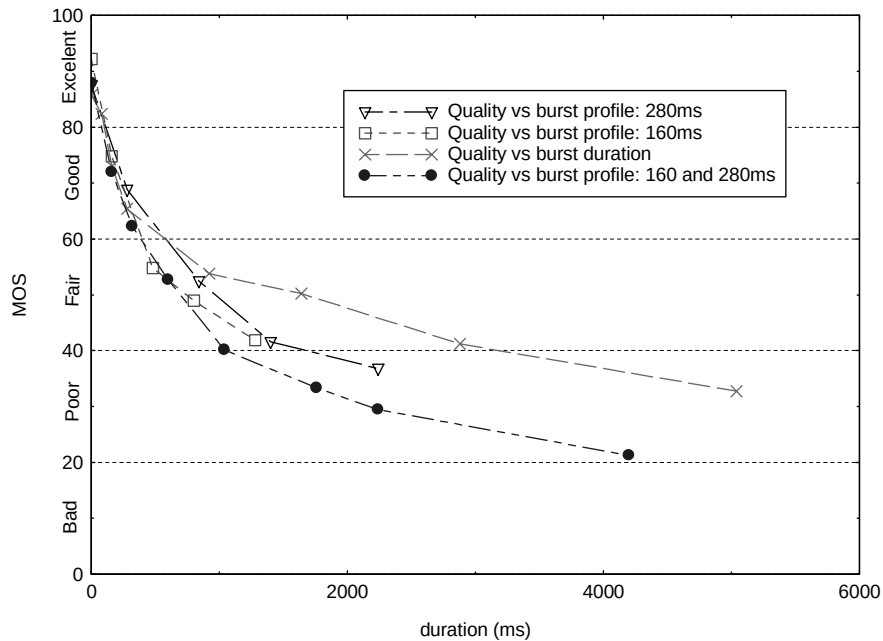


Fig. 12. Results of the three quality experiments: comparison.

## 4. CONCLUSIONS

First, the perceptual detection thresholds of generated temporal discontinuities were measured. We have found that the visibility of frame dropping is content, duration and motion dependent.

Second, the assessment results show how quality is impaired by a single burst of dropped frames in a 10 sec sequence. The effect of several bursts of discarded frames, irregularly distributed over the time was also discussed. Sporadic Frame dropping has a significant effect on quality perception that is mainly content dependent. The assessment results showed that duration could be a convenient descriptor to characterize the perceptual effect of sporadic dropping. Subjects seem to be more annoyed by several short discontinuities than by a single burst of long duration. However, several questions were left open in relation to the influence of these temporal discontinuities for high temporal densities and distributions. We will therefore study the effect of frame dropping at a constant rate either alone or combined with a burst of discarded pictures for different durations on perceived quality; the influence of different temporal distribution profiles will be explored.

Finally, this study has also outlined a first approach to characterize temporal discontinuities. The analysis of the results is useful for video quality metrics design, perceptual frame discard algorithms and for a better understanding of time-varying quality.

## ACKNOWLEDGEMENTS

We would like to thank Jean-Louis Blin and Emmanuel Wyckens at the EQS Lab, France Telecom R&D, involved in the design of the experiments and positive critics. We also thank the non remunerated viewers who participated in the set up of preliminary tests. Special thanks go to Sophie Lounds for her corrections.

## REFERENCES

1. J. M. Boyce, R. D. Gaglianella, "Packet Loss Effects on MPEG Video Sent Over the Public Internet," *ACM Multimedia 98*, pp. 181-190, ACM Press, Bristol, 1998.
2. R. Aldridge, J. Davidoff, M. Ghanbari, D. Hands, and D. Pearson, "Recency Effect in the Subjective Assessment of Digital Coded Television Pictures," *Image Processing and its Applications*, Vol. 410, pp. 336-339, IEE, Edinburgh, 1995.
3. B. Farell, and D. Pelli, "Psychophysical Methods, or How to Measure a Threshold and Why.," *A Practical Guide to Vision Research*, Eds. J. G. Robson and R. H. S. Carpenter, pp. 129-136, Oxford University Press, New York, 1998.
4. J. Blin, *Comparaison de Méthodes d'Evaluation de la Qualité des Images Streamées sur Réseau IP*, France Telecom R&D report FT.BD.FT&D/DIH/EQS/504, 2000.
5. ITU-R BT.500, *Methodology for the Subjective Assessment of the Quality of Television Pictures*, 1998.
6. ITU-R BT.814, *Specifications and Alignment Procedures for Setting of Brightness and Contrast of Displays*, 1994.
7. J. Blin, *Echantillonnage Temporel Préféré du Codage Vidéo pour les Formats CIF et QCIF Selon le Débit Choisi*, France Telecom R&D report FT.BD.FTR&D/DIH/EQS/502, 2000.
8. J. Blin, *SEOVQ Software Tool for Quality, Preference and Acceptability Evaluation of Multimedia Images*, France Telecom R&D report FT.BD.FTR&D/DIH/EQS/462/02/JLB, 2002.
9. W. Tam, L. Stelmach, L. Wang, D. Lauzon and P. Gray, "Visual Masking at Video Scene Cuts," *Human Vision, Visual Processing, and Digital Display VI*, Vol. 2411, pp. 111-119, SPIE, San Jose, 1995.
10. B. Girod, "The Information Theoretical Significance of Spatial and Temporal Masking in Video Signals," In *Proc. Human Vision, Visual Processing, and Digital Displays*, Vol. 1077, pp. 178-187, SPIE, Los Angeles, 1989.
11. M. Masry and S. Hemami, "An Analysis of Subjective Quality in Low Bit Rate Video," *Intl. Conf. on Image Processing*, Vol. 1, pp. 465-468, IEEE, Thessaloniki, 2001.
12. J. J. Freyd, "The Mental Representation of Movement When Static Stimuli are Viewed," *Perception and Psychophysics*, **Vol. 33** (6), pp. 575-581, 1983.
13. R. Fox, C. McDaniel. "The Perception of Biological Motion by Human Infants," *Science*, **Vol. 218**, pp. 486-487, 1982.