

# Métrie Perceptuelle des Rupture de Fluidité Vidéo sans Référence

R. Pastrana-Vidal<sup>1,2</sup> C. Colomes<sup>1</sup> J. Gicquel<sup>1</sup> H. Cherifi<sup>2</sup>

<sup>1</sup> France Telecom R&D, DIH/EQS/M@I

4, rue de Clos Courtel  
35512 Cesson Sévigné -France

{ricardo.pastrana, catherine.colomes, jeancharles.gicquel}@francetelecom.com

<sup>2</sup> LIRSIA

Université de Bourgogne, Faculté de Science Mirande  
BP 400, 21011 Dijon cedex-France

cherifi@u.bourgogne.fr

## Résumé

*La discontinuité de mouvement apparent provoquée par la perte d'images est une dégradation temporelle souvent présente dans les services de vidéo en temps réel sur Internet (streaming vidéo). L'utilisateur final perçoit une rupture de la fluidité de l'information visuelle ayant un effet sur son jugement de qualité de la séquence. Une métrique permettant d'évaluer l'impact de la rupture de fluidité sur la qualité perçue par un groupe d'observateurs est présentée. La mesure est basée sur la détection de perte d'images, d'un seuillage perceptuel, d'une fonction de qualité obtenue dans nos études psychovisuelles précédentes et d'une fonction de cumul temporel modélisant le mécanisme de jugement des observateurs. La métrique n'a pas besoin du signal original pour estimer la qualité. Les estimations obtenues par la métrique et les votes des observateurs montrent une corrélation significative.*

## Mots clefs

Métrique, Qualité, Vidéo streaming, Perte d'Image, Gel, Saccades, Fluidité, Perception, Modélisation.

## 1 Introduction

L'intérêt pour les nouveaux services de vidéo en temps réel sur Internet a augmenté de façon importante dans les dernières années. Mis à part le flou, les artefacts de bloc et la distorsion de la couleur, la discontinuité de mouvement apparent est une dégradation temporelle souvent présente dans le *streaming* vidéo. L'observateur final perçoit une **rupture de la fluidité** de l'information visuelle. Le mouvement apparaît comme saccadé lorsque les

discontinuités ont une distribution temporelle régulière. Une rupture instantanée de la fluidité est perçue lorsque la perte de continuité est sporadique. Les techniques de réduction du débit, les pertes de paquets dans le réseau et la stratégie de *bufferisation* adoptée par le décodeur peuvent être la cause de ce mouvement saccadé. Dans le codage de source, le sous-échantillonnage temporel est une des techniques les plus employées pour la réduction de débit: une opération de suppression périodique d'images est appliquée sur la séquence originale. Ainsi, la suppression d'images affecte l'information de mouvement. D'autre part, la perte de paquets, le délai et la variation du délai peuvent entraîner une perte irrégulière d'images au moment du décodage due aux limitations du *buffer* de réception [1]. En effet, la dernière image reçue est alors affichée jusqu'à ce qu'une nouvelle image soit reconstruite. L'utilisateur final perçoit donc un gel d'image suivi d'un déplacement abrupt des objets dans la scène. Tout au long de l'article nous utiliserons les termes *discontinuité temporelle* et *gel* comme synonymes de la *rupture de fluidité*.

La qualité d'un service multimedia est souvent estimée en fonction de l'ensemble des paramètres techniques intervenant dans la chaîne de communication (codage, débit, pertes de paquets, délais) [2] [3] ou à partir de mesures de fidélité (PSNR, MSE). Néanmoins cette approche n'est pas suffisamment adaptée [4] aux services de diffusion pour lesquels la *perception de l'utilisateur* revêt un intérêt central. Il est donc indispensable de définir des métriques de qualité en adéquation avec la perception et le jugement humain.

Aujourd'hui, pour quantifier l'impact des dégradations visuelles sur le jugement humain, nous utilisons des

méthodes d'évaluation subjectives. Celles-ci sont les seules solutions unanimement reconnues pour évaluer et caractériser de façon fiable la qualité perçue. Cependant, la mise en place d'un test subjectif demande un effort considérable en terme de conception et de réalisation. Il faut de plus disposer d'un environnement contrôlé (laboratoire de test). Ceci justifie la nécessité de développer des outils automatiques d'évaluation de la qualité basés sur des métriques corrélées à la perception humaine.

Les métriques peuvent être classifiées en fonction de leur dépendance vis à vis du signal original. Les métriques dites *avec référence* estiment la qualité à partir des différences entre le signal original et le signal dégradé. Si la mesure est basée sur la différence entre quelques primitives extraites de chacun des signaux il s'agit alors d'une technique *avec référence réduite*. Lorsque la mesure de qualité est basée sur la détection aveugle de dégradations (à partir d'un modèle) on parle de *métrique sans référence*. La recherche de métriques sans référence est motivée par le fait que la qualité peut se mesurer même si la séquence originale est inaccessible ou inexistante. Des métriques de flou [5] et d'effet de bloc [6] sans référence ont été proposées récemment.

Dans la littérature nous n'avons pas trouvé pour l'instant des métriques pour quantifier les effets liés à la perte d'images sur la qualité perçue. A ce sujet, une série de tests subjectifs a été menée au sein de notre équipe de recherche pour caractériser les effets de la perte d'images sur le jugement de la qualité. Les résultats ont permis de définir une métrique fournissant une note de qualité fortement corrélée à celle d'un groupe d'observateurs.

## 2 Effet de la perte d'images sur la qualité perçue

La qualité perçue par l'utilisateur est estimée grâce à la note moyenne de jugement (MOS<sup>1</sup>) provenant de tests subjectifs. Un ensemble d'observateurs évalue la qualité en utilisant une échelle d'appréciation subjective (excellente, bonne, assez bonne, mauvaise, médiocre) liée à une échelle de valeurs (0 à 100). Les votes des observateurs sont obtenus en utilisant la méthode SAMVIQ (Subjective Assessment Methodology for Video Quality) [7] développée par le laboratoire EQS de France Telecom R&D. Cette méthode a été récemment standardisée par la EBU (European Broadcasting Union).

Nous présentons brièvement ci-dessous un aperçu de nos études sur la visibilité et les effets de la perte d'images sur la qualité [8].

Tout d'abord, nous avons remarqué qu'une discontinuité temporelle provoquée par un gel isolé de 80ms est visible par 80% des observateurs. Le taux de détection franche (100%) est atteint pour une discontinuité de 200ms. En ce qui concerne les effets sur le jugement de la qualité, un gel isolé de 80 ms n'a pas un impact significatif puisque la qualité est encore considérée comme excellente. Lorsque le gel a une durée entre 160 et 280ms la note MOS décroît d'un item de qualité. Au delà de cette valeur les notes décroissent plus lentement (Figure 1).

Ensuite, lorsqu'un nombre réduit de gels dégrade la vidéo, les résultats des tests montrent que les observateurs sont plus gênés par quelques discontinuités de courte durée que par un gel isolé de longue durée.

Enfin, en faisant varier la densité (nombre) de gel, sa durée et sa distribution temporelle, nous avons constaté que les discontinuités périodiques sont moins gênantes que la perte irrégulière des images [9].

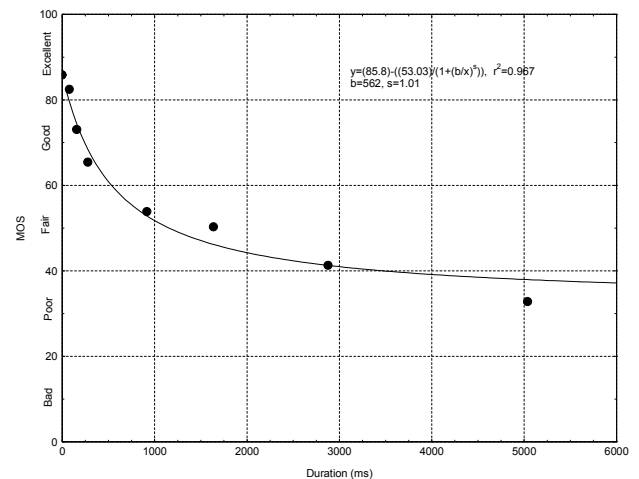


Figure 1 – *Qualité perçue en fonction de la durée d'un gel moyennée sur un ensemble de contenus vidéos variés.*

## 3 Modèle de prédiction

Les résultats de nos travaux sur l'effet de la perte d'images sur la qualité perçue ont montré que la dégradation de la qualité dépend, de façon non linéaire, de la durée du gel et de sa densité. Nous avons donc défini un modèle de prédiction combinant la fonction de qualité pour un gel isolé, la densité des gels en fonction de leur durée (histogramme) et une fonction de pondération dépendante de la densité :

$$\hat{Q} = Q_{ref} - d_{total}$$

<sup>1</sup> Mean Opinion Score

$$d_{total} = \left[ \sum_{t_g = T_{min}}^{T_{max}} d_{t_g} \right]^{1/2},$$

$$d_{t_g} = n(t_g) \times [\hat{e}(t_g)]^{p(n(t_g))},$$

$$\hat{e}(t_g) = Q_{ref} - \hat{q}(t_g),$$

$$\hat{q}(t_g) = m_{max} - \left( (m_{max} - m_{min}) / \left( 1 + (b/t_g)^f \right) \right),$$

$$p(n(t_g)) = p_{max} - \left( (p_{max} - p_{min}) / \left( 1 + (c/n(t_g))^f \right) \right).$$

$Q_{ref}$  est une constante correspondant à la note MOS pour l'ensemble de séquences originales (non dégradées),  $d_{total}$  est la dégradation globale calculée dans la séquence de 10s,  $t_g$  est la variable correspondant à la durée du gel,  $T_{min}$  est la durée minimale d'un gel ( $T_{min} >$  seuil de détection).  $T_{max}$  est égale au temps d'analyse globale qui est de 10s. Le terme  $d_i$  correspond à la dégradation provenant de toutes les discontinuités ayant la même durée  $t_g$ . Le terme  $n(t_g)$  est l'historgramme des durées des discontinuités. L'expression  $q(t_g)$  est la fonction de qualité perçue associée à une seule discontinuité (Figure 1). Les constantes  $m_{max}$  et  $m_{min}$  correspondent aux valeurs extrêmes de qualité trouvées dans le test psychovisuel. La contribution de chaque discontinuité est pondérée en fonction de sa densité grâce à la fonction de pondération  $p(n(t_g))$ , la contribution de chaque discontinuité est donc moins significative lorsque le nombre de gels de la même durée est élevé. Cette fonction prend en compte le fait que les observateurs sont plus sensibles à quelques ruptures de fluidité qu'aux ruptures périodiques.

## 4 Métrique

La métrique (Figure 2) est basée sur une détection de perte d'images, un seuillage perceptuel [10], des fonctions psychovisuelles de qualité obtenues dans nos travaux précédents [8] [9] et sur une fonction d'intégration temporelle des défauts modélisant le mécanisme de jugement des observateurs. La métrique produit une note entre 0 et 100. L'échelle est annotée par des critères de qualité, allant de mauvais à excellent, linéairement répartis le long de l'échelle.

Nous avons souligné auparavant que la rupture de fluidité perçue, provoquée par la perte d'images, est caractérisée par la présence d'un gel d'image suivi d'une discontinuité de mouvement. La perte d'image est donc liée à la présence du gel (du point de vue de la restitution du signal

vidéo). Le gel d'image est détecté en calculant si la dérivée de luminances par rapport au temps est nulle :

$$I(x, y, t) - I(x, y, t + 1) = 0.$$

Un gel détecté est considéré comme perçu si sa durée est supérieure au seuil de détection ( $\tau_d$ ). Après la détection perceptuelle, un histogramme des durées est réalisé. Finalement, la note MOS est calculée grâce au modèle de jugement réalisant le cumul temporel des dégradations (Sec. 3).

Séquence Dégradée

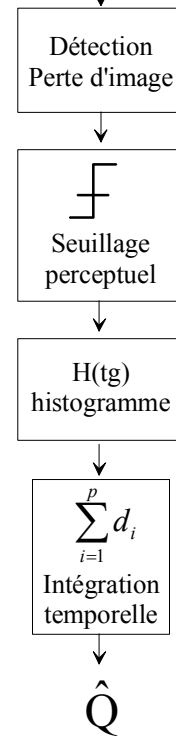


Figure 2 – Métrique de rupture de fluidité.

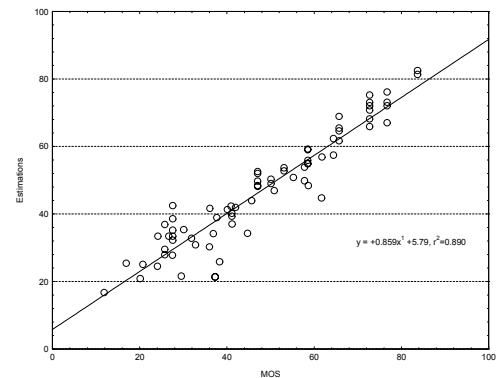


Figure 3 – Corrélation entre les notes estimés par la métrique et celles des observateurs.

La métrique est fortement corrélée aux votes provenant d'un groupe d'observateurs (corrélation linéaire de 0.94). La Figure 3 montre la dispersion entre les calculs de la métrique et les résultats de jugement de la qualité, pour 74 conditions de test. Les paramètres de la métrique ont été optimisés en utilisant 21 distributions différentes de perte d'images.

## 5 Limitations et perspectives

D'une part, il est important de remarquer que pour la plupart des métriques de qualité sans référence, il existe des problèmes de fausses détections. Dans notre cas, la métrique ne distingue pas un gel dû au contenu d'un gel résultant de la perte d'images. Cependant, la limitation peut être surmontée en introduisant des informations sur l'état du canal (par exemple la perte de paquets). Ce problème est d'ailleurs minime, car la probabilité de trouver un gel dû au contenu est assez réduite parce que l'information dans la vidéo porte sur la notion de mouvement.

D'autre part, la mesure est valable pour une fenêtre d'analyse temporelle de 10s car notre objectif principal a été d'éviter les effets d'oubli inhérents à la mémoire à court terme humaine. Dans nos perspectives nous souhaitons étudier la rupture de fluidité pour des séquences de plus longue durée.

## 6 Conclusions

Une métrique fournissant une note de qualité fortement corrélée à celle d'un groupe d'observateurs lorsque la vidéo est dégradée par une rupture de fluidité (perte d'images) a été présentée. Notre métrique prend en compte différentes distributions temporelles de pertes d'images : régulières, irrégulières, et en rafales.

La métrique s'intègre dans un processus global d'évaluation automatique de la qualité, mais elle peut également intéresser d'autres domaines tels que le codage vidéo, par exemple pour le développement d'un algorithme de sous échantillonnage temporel basé sur la perception humaine. Cette étude montre ainsi une méthodologie prometteuse permettant la caractérisation et la modélisation perceptuelle des discontinuités temporelles. Ceci participe à une meilleure compréhension des mécanismes d'évaluation humains de la qualité fluctuante.

## Références

[1] M. Kalman, E. Steinbach, B. Girod, "R-D optimized media streaming enhanced with adaptive media playout," *ICME*, Vol. 1, IEEE, Lausanne, pp. 869 - 872, Août 2002.

[2] Z. Wang, S. Banerjee, and S. Jamin, "Studying Streaming Video Quality: From An Application Point of View," *ACM Multimedia*, ACM Press, Berkeley CA, pp. 327 - 330. November 2003.

[3] S. Mohamed, and G. Rubino, "A study of real-time packet video quality using random neural networks," *Circuits and Systems for Video Technology*, Vol. 12, IEEE, pp. 1071 - 1083, December 2002.

[4] P. Pruthi, I. Dragos, and A. Popescu, "Application Level Performance of Multimedia Services," *SPIE-Internet II: Quality of Services and Future Directions*, Vol. 3842, SPIE, Boston MA, pp. 154-164, September 1999.

[5] P. Marziliano, F. Dufaux, S. Winkler, T. Ebrahimi, "A no-reference perceptual blur metric." *ICIP*, Vol. 3, IEEE, Rochester NY, pp. 57-60, September 2002.

[6] S. Suthaharan, "A Perceptually Significant Block-Edge Impairment Metric for Digital Video Coding," *ICASSP*, Vol. 3, IEEE, Hong Kong, pp. 681-684, April 2003.

[7] *SAMVIQ - Subjective Assessment Methodology for Video Quality*, EBU Project Group B/VIM Video In Multimedia, BPN 056 report, Mai 2003.

[8] R.R. Pastrana-Vidal, J.C. Gicquel, C. Colomes, and H. Cherifi, "Sporadic Frame Dropping Impact on Quality Perception," *Human Vision and Electronic Imaging IX*, SPIE, vol. 5292, San José CA, Janvier 2004.

[9] R.R. Pastrana-Vidal, J.C. Gicquel, C. Colomes, and H. Cherifi, "Frame Dropping Effects on User Quality Perception," Dans *Actes de WIAMIS 04*, Lisbonne, Avril 2004.

[10] 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*, issue 410, IEE, Edinburgh, pp. 336-339, Juillet 1995.

[11] B. Farell, and D. Pelli, *Psychophysical methods, or how to measure a threshold and why*. In J. G. Robson and R. H. S. Carpenter (Eds.), *A Practical Guide to Vision Research*, Oxford University Press, New York, 1998.