

UIT-R

Secteur des Radiocommunications de l'UIT

Recommandation UIT-R BT.1866
(03/2010)

**Techniques de mesure objective de la
qualité vidéo perceptuelle pour les
applications de radiodiffusion utilisant la
télévision basse définition en présence d'un
signal de référence complet**

Série BT
Service de radiodiffusion télévisuelle



Avant-propos

Le rôle du Secteur des radiocommunications est d'assurer l'utilisation rationnelle, équitable, efficace et économique du spectre radioélectrique par tous les services de radiocommunication, y compris les services par satellite, et de procéder à des études pour toutes les gammes de fréquences, à partir desquelles les Recommandations seront élaborées et adoptées.

Les fonctions réglementaires et politiques du Secteur des radiocommunications sont remplies par les Conférences mondiales et régionales des radiocommunications et par les Assemblées des radiocommunications assistées par les Commissions d'études.

Politique en matière de droits de propriété intellectuelle (IPR)

La politique de l'UIT-R en matière de droits de propriété intellectuelle est décrite dans la «Politique commune de l'UIT-T, l'UIT-R, l'ISO et la CEI en matière de brevets», dont il est question dans l'Annexe 1 de la Résolution UIT-R 1. Les formulaires que les titulaires de brevets doivent utiliser pour soumettre les déclarations de brevet et d'octroi de licence sont accessibles à l'adresse <http://www.itu.int/ITU-R/go/patents/fr>, où l'on trouvera également les Lignes directrices pour la mise en oeuvre de la politique commune en matière de brevets de l'UIT-T, l'UIT-R, l'ISO et la CEI et la base de données en matière de brevets de l'UIT-R.

Séries des Recommandations UIT-R

(Egalement disponible en ligne: <http://www.itu.int/publ/R-REC/fr>)

Séries	Titre
BO	Diffusion par satellite
BR	Enregistrement pour la production, l'archivage et la diffusion; films pour la télévision
BS	Service de radiodiffusion sonore
BT	Service de radiodiffusion télévisuelle
F	Service fixe
M	Services mobile, de radiorepérage et d'amateur y compris les services par satellite associés
P	Propagation des ondes radioélectriques
RA	Radio astronomie
RS	Systèmes de télédétection
S	Service fixe par satellite
SA	Applications spatiales et météorologie
SF	Partage des fréquences et coordination entre les systèmes du service fixe par satellite et du service fixe
SM	Gestion du spectre
SNG	Reportage d'actualités par satellite
TF	Emissions de fréquences étalon et de signaux horaires
V	Vocabulaire et sujets associés

Note: Cette Recommandation UIT-R a été approuvée en anglais aux termes de la procédure détaillée dans la Résolution UIT-R 1.

Publication électronique
Genève, 2010

© UIT 2010

Tous droits réservés. Aucune partie de cette publication ne peut être reproduite, par quelque procédé que ce soit, sans l'accord écrit préalable de l'UIT.

RECOMMANDATION UIT-R BT.1866

Techniques de mesure objective de la qualité vidéo perceptuelle pour les applications de radiodiffusion utilisant la télévision basse définition* en présence d'un signal de référence complet

(2010)

Domaine de compétence

La présente Recommandation définit des méthodes d'évaluation de la qualité vidéo perçue des applications de radiodiffusion utilisant la télévision basse définition (LDTV) lorsqu'un signal de référence complet est disponible.

L'Assemblée des radiocommunications de l'UIT,

Considérant

- a) qu'il n'est plus à démontrer depuis longtemps que la possibilité de mesurer automatiquement la qualité d'une séquence vidéo diffusée constitue un atout précieux pour l'industrie;
- a) que la Recommandation UIT-R BT.1683 décrit des méthodes objectives de mesure de la qualité vidéo perçue de la télédiffusion numérique à définition normale en présence d'une image de référence complète;
- c) que la Recommandation UIT-R BT.1833 décrit des systèmes multimédias pour la diffusion d'applications multimédias et d'applications de données destinées à la réception mobile au moyen de récepteurs portatifs;
- d) que la télévision basse définition (LDTV) est de plus en plus largement utilisée pour la diffusion d'applications multimédias et d'applications de données destinées à la réception mobile;
- e) que la Recommandation J.247¹ définit des techniques de mesure objective de la qualité vidéo perceptuelle destinées aux applications LDTV en présence d'une image de référence complète;
- f) que les mesures objectives de la qualité vidéo perçue de la LDTV peuvent compléter les méthodes d'évaluation subjective,

reconnaissant

- a) que la LDTV est destinée essentiellement à être utilisée pour le visionnage sur de petits écrans tels que ceux qu'offrent les récepteurs portatifs et mobiles,

recommande

- 1** d'utiliser les lignes directrices, les paramètres et les limites indiqués dans l'Annexe 1 pour appliquer les modèles de mesure objective de la qualité vidéo désignés au point 2 du *recommande*;

* La télévision basse définition (LDTV) est associée à des résolutions vidéo comportant un nombre plus faible d'éléments d'image que celles qui sont définies dans la Recommandation UIT-R BT.601. Une Recommandation pertinente de l'UIT-R sur la LDTV est à l'étude.

¹ La Recommandation UIT-T J.247 peut être consultée à l'adresse suivante: <<http://www.itu.int/rec/T-REC-J.247-200808-P/en>>.

2 d'utiliser les modèles de mesure objective de la qualité vidéo perceptuelle fournis dans la Recommandation J.247 de l'UIT-T pour les applications de radiodiffusion utilisant la télévision basse définition lorsqu'un signal de référence complet est disponible.

NOTE 1 – Des résumés des modèles de mesure sont présentés dans les Annexes 2 à 5, à titre d'information. Pour plus de précisions, voir la Recommandation J.247 de l'UIT-T.

Annexe 1

1 Introduction

La présente Recommandation définit les méthodes à appliquer pour évaluer la qualité vidéo perçue des applications de radiodiffusion utilisant la LDTV lorsqu'un signal de référence complet est disponible.

Les méthodes d'évaluation qui y sont décrites sont valables pour:

- l'évaluation du codec, ses spécifications et les essais d'homologation;
- le contrôle de qualité pendant le service, éventuellement en temps réel, à la source;
- le télécontrôle de qualité à la destination, lorsqu'une copie de la source est disponible;
- la mesure de qualité pour la surveillance d'un système d'archivage vidéo ou d'un système numérique vidéo qui utilise soit une application unique, soit une concaténation de techniques de compression et de décompression numériques;
- la mise à l'essai en laboratoire des systèmes vidéo.

La méthode de mesure avec image de référence complète peut être utilisée lorsque le signal vidéo de référence non dégradé est aisément disponible au point de mesure; tel peut être le cas des mesures effectuées en laboratoire ou dans un environnement fermé sur un seul équipement ou sur un enchaînement de processus. Les méthodes d'évaluation reposent sur le traitement de séquences vidéo à basse définition ayant une résolution VGA, CIF ou QCIF.

Les séquences soumises aux essais de validation contenaient à la fois de nombreuses dégradations dues au codage et diverses erreurs de transmission (erreurs sur les bits et perte de paquets, par ex.).

Si des distorsions du codage des signaux vidéo sont prises en considération, le codeur peut utiliser diverses méthodes de compression (dont, par ex., MPEG-2, H.264). Les modèles décrits dans la présente Recommandation peuvent être utilisés pour contrôler la qualité de réseaux déployés et s'assurer que ces réseaux sont prêts à être exploités. Les effets visuels des dégradations peuvent comprendre des dégradations aussi bien spatiales que temporelles (répétitions de trames, sauts d'image et réduction du débit de trame, par ex.). Les modèles décrits dans la présente Recommandation peuvent également être utilisés pour la mise à l'essai en laboratoire de systèmes vidéo.

La présente Recommandation est considérée comme convenant aux services assurés à 4 Mbit/s ou à des débits inférieurs et affichés sur des récepteurs mobiles. L'essai de validation a été effectué, pour chaque résolution, dans les conditions ci-après, qui se sont avérées satisfaisantes:

- QCIF (quart de format intermédiaire commun (176 x 144 pixels)): de 16 à 320 kbit/s.
- CIF (format intermédiaire commun (352 x 288 pixels)): de 64 kbit/s à 2 Mbit/s.
- VGA (récepteurs à écran graphique (640 x 480 pixels)): de 128 kbit/s à 4 Mbit/s.

TABLEAU 1
Facteurs utilisés dans l'évaluation des modèles

Facteurs d'essai
Erreurs de transmission avec perte de paquets
Résolution vidéo QCIF, CIF et VGA
Débits vidéo QCIF: de 16 à 320 kbit/s CIF: de 64 kbit/s à 2 Mbit/s VGA: de 128 kbit/s à 4 Mbit/s
Erreurs temporelles (arrêt dû au saut d'image) de 2 s maximum
Débits de trame vidéo: de 5 à 30 trames par seconde
Schémas de codage
H.264/AVC (MPEG-4 Partie 10), MPEG-4 Partie 2, et trois autres schémas de codage privés. (Voir la Note 1 ci-dessous.)

NOTE 1 – Les essais de validation des modèles comprenaient des séquences vidéo codées au moyen de 15 codecs vidéo différents. Les cinq codecs mentionnés dans le Tableau 1 sont ceux qui ont été le plus souvent utilisés pour coder les séquences d'essai et les modèles recommandés peuvent tous être considérés comme convenant à l'évaluation de ces codecs. Un nombre plus faible de séquences d'essai ont été créées à l'aide des codecs suivants: H.261, H.263, H.263⁺², JPEG-2000, MPEG-1, MPEG-2, H.264 SVC, et d'autres systèmes privés. On notera que certains de ces codecs ont été utilisés uniquement pour les résolutions CIF et QCIF car ils devraient, en principe, être employés principalement pour ces résolutions sur le terrain. Avant d'appliquer un modèle à des séquences codées au moyen de l'un de ces codecs, l'utilisateur devrait examiner avec soin sa performance en matière de prévision pour déterminer si cette performance est acceptable.

2 Application

Les applications des modèles d'évaluation décrits dans la présente Recommandation comprennent les suivantes, sans toutefois s'y limiter:

- 1 l'évaluation du codec, ses spécifications et les essais d'homologation;
- 2 le contrôle de qualité pendant le service, éventuellement en temps réel, à la source;
- 3 le télécontrôle de qualité à la destination, lorsqu'une copie de la source est disponible;
- 4 la mesure de qualité pour la surveillance d'un système d'archivage vidéo ou d'un système numérique vidéo qui utilise soit une application unique, soit une concaténation de techniques de compression et de décompression numériques; et
- 5 la mise à l'essai en laboratoire de systèmes vidéo.

² Le H.263+ est une configuration particulière du H.263 (1998).

3 Utilisation des modèles

La présente Recommandation porte notamment sur les modèles de calcul objectif indiqués au Tableau 2. Un aperçu général des performances de ces modèles est donné au Tableau 3. Des renseignements supplémentaires sont fournis à l'Appendice 1.

TABLEAU 2
Modèles de calcul objectif

Id. du modèle	Initiateur	Pays	Annexe
A	NTT	Japon	2
B	OPTICOM	Allemagne	3
C	Psytechnics	Royaume-Uni	4
D	Université de Yonsei	Corée (Rép. de)	5

Ces quatre modèles donnent tous une bien meilleure indication de la qualité que la valeur de crête du rapport signal sur bruit (PSNR).

Les modèles B et C donnent, dans l'ensemble, une meilleure indication que les modèles A et D pour certaines résolutions. Les modèles B et C donnent généralement des résultats statistiquement équivalents. Pour le format QCIF, le modèle A est souvent statistiquement équivalent aux modèles B et C. Pour le format VGA, le modèle D est généralement statistiquement équivalent aux modèles B et C. Les tableaux ci-dessous donnent une vue d'ensemble des performances des modèles.

Bien que les quatre modèles puissent être utilisés pour répondre de manière satisfaisante à différents besoins du secteur, il est fortement conseillé d'utiliser le modèle B, C ou D pour la résolution VGA si l'on souhaite obtenir une performance légèrement supérieure dans la plupart des cas. Pour cette même raison, il est fortement conseillé d'utiliser le modèle B ou C pour le format CIF et le modèle A, B ou C pour le format QCIF.

Le tableau montre que le modèle B offre la meilleure corrélation minimale globale. Les coefficients de corrélation minimale des modèles B, A, D et C sont respectivement 0,68, 0,60, 0,59 et 0,57.

Le modèle C est celui qui a figuré le plus souvent dans le groupe des modèles ayant donné les meilleurs résultats. Le nombre total de fois où ils ont figuré dans le groupe des modèles les plus performants est de 37 pour le modèle C, de 34 pour le modèle B, de 25 pour le modèle A et de 24 pour le modèle D.

TABLEAU 3

Aperçu général des performances des modèles

VGA	Modèle A	Modèle B	Modèle C	Modèle D	PSNR
Corrélation moyenne	0,786	0,825	0,822	0,805	0,713
Corrélation minimale	0,598	0,685	0,565	0,612	0,499
Nombre de fois premier	8	10	11	10	3
Analyse du classement	Deuxième	Meilleur	Meilleur	Meilleur	–

CIF	Modèle A	Modèle B	Modèle C	Modèle D	PSNR
Corrélation moyenne	0,777	0,808	0,836	0,785	0,656
Corrélation minimale	0,675	0,695	0,769	0,712	0,440
Nombre de fois premier	8	13	14	10	0
Analyse du classement	Deuxième	Meilleur	Meilleur	Deuxième	–

QCIF	Modèle A	Modèle B	Modèle C	Modèle D	PSNR
Corrélation moyenne	0,819	0,841	0,830	0,756	0,662
Corrélation minimale	0,711	0,724	0,664	0,587	0,540
Nombre de fois premier	9	11	12	4	1
Analyse du classement	Meilleur	Meilleur	Meilleur	Deuxième	–

4 Limites

Les modèles d'évaluation décrits dans la présente Recommandation ne peuvent se substituer entièrement aux essais subjectifs. Les valeurs de corrélation entre deux essais subjectifs conçus et exécutés avec soin (par exemple dans deux laboratoires différents) se situent normalement entre 0,95 et 0,98.

Les modèles de la présente Recommandation ont été validés en mesurant des séquences vidéo présentant des gels de trame d'une durée maximale de 2 s.

Les modèles de la présente Recommandation n'ont pas été validés pour la mesure de séquences vidéo affectées d'un retard qui s'accroît progressivement (par ex. une séquence vidéo qui ne rejette pas les trames manquantes après un gel d'image).

Il convient de noter que si de nouvelles techniques de codage et de transmission produisant des défauts (artéfacts) étaient utilisées, ce qui n'est pas le cas de la présente évaluation, les modèles objectifs pourraient donner des résultats erronés. En pareil cas, une évaluation subjective serait nécessaire.

Appendice 1 à l'Annexe 1

Résultats des travaux du Groupe d'experts en qualité vidéo

Les études de la mesure de la qualité vidéo perceptuelle sont menées au sein d'un groupe non officiel, désigné «Groupe d'experts en qualité vidéo (VQEG)», qui rend compte de ses travaux aux Commissions d'études 9 et 12 de l'UIT-T et à la Commission d'études 6 des Radiocommunications. Au cours du test de la Phase I (Multimédias), terminé récemment, le VQEG a évalué la performance d'algorithmes de mesure de la qualité vidéo perceptuelle en présence d'un signal de référence complet pour les formats QCIF, CIF et VGA.

Sur la base des éléments dont il dispose à l'heure actuelle, l'UIT-R est en mesure de recommander, pour l'instant, les quatre méthodes suivantes:

Modèle A (Annexe 2) – Initiateur du VQEG, NTT, Japon

Modèle B (Annexe 3) – Initiateur du VQEG, OPTICOM, Allemagne

Modèle C (Annexe 4) – Initiateur du VQEG, Psytechnics, Royaume-Uni

Modèle D (Annexe 5) – Initiateur du VQEG, Université de Yonsei, Corée (République de).

Une description technique de ces modèles est donnée aux Annexes 2 à 5 respectivement. On notera que l'ordre des annexes est purement arbitraire et ne donne aucune indication de la performance en matière de prévision de la qualité.

Le Tableau 4 fournit des précisions sur les performances des modèles au cours du test de la Phase I (Multimédias) effectué par le VQEG.

TABLEAU 4

**a) Résolution VGA: Performances des modèles au cours du test de la Phase I
(Multimédias) effectué par le VQEG –
Moyennes de 14 tests subjectifs**

Mesure	Modèle A	Modèle B	Modèle C	Modèle D	PSNR ⁽¹⁾
Annexe	2	3	4	5	
Corrélation de Pearson	0,786	0,825	0,822	0,805	0,713
Erreur quadratique moyenne	0,621	0,571	0,566	0,593	0,714
Proportion de points éloignés	0,523	0,502	0,524	0,542	0,615

**b) Résolution CIF: Performances des modèles au cours du test de la Phase I
(Multimédias) effectué par le VQEG –
Moyennes de 14 tests subjectifs**

Mesure	Modèle A	Modèle B	Modèle C	Modèle D	PSNR ⁽¹⁾
Annexe	2	3	4	5	
Corrélation de Pearson	0,777	0,808	0,836	0,785	0,656
Erreur quadratique moyenne	0,604	0,562	0,526	0,594	0,720
Proportion de points éloignés	0,538	0,513	0,507	0,522	0,632

**c) Résolution QCIF: Performances des modèles au cours du test de la Phase I
(Multimédias) effectué par le VQEG –
Moyennes de 14 tests subjectifs**

Mesure	Modèle A	Modèle B	Modèle C	Modèle D	PSNR ⁽¹⁾
Annexe	2	3	4	5	
Corrélation de Pearson	0,819	0,841	0,830	0,756	0,662
Erreur quadratique moyenne	0,551	0,516	0,517	0,617	0,721
Proportion de points éloignés	0,497	0,461	0,458	0,523	0,596

⁽¹⁾ Les valeurs du rapport PSNR indiquées dans le tableau sont extraites du rapport final du VQEG sur la Phase I (Multimédias) (voir: <http://www.its.bldrdoc.gov/vqeg/projects/multimedia/>). Ces valeurs ont été calculées par la NTIA/IITS.

D'après les mesures effectuées, pour la résolution VGA, chacun des modèles FR a figuré dans le groupe des modèles les plus performants le nombre de fois indiqué ci-dessous:

Statistique	Modèle A	Modèle B	Modèle C	Modèle D	PSNR
Corrélation	8	10	11	10	3
RMSE ⁽¹⁾	4	8	10	6	0
Proportion de points éloignés	9	11	12	8	4

⁽¹⁾ RMSE: erreur quadratique moyenne.

D'après les mesures effectuées, pour la résolution CIF, chacun des modèles FR a figuré dans le groupe des modèles les plus performants le nombre de fois indiqué ci-dessous:

Statistique	Modèle A	Modèle B	Modèle C	Modèle D	PSNR
Corrélation	8	13	14	10	0
RMSE ⁽¹⁾	6	10	13	9	0
Proportion de points éloignés	10	13	12	11	1

⁽¹⁾ RMSE: erreur quadratique moyenne.

D'après les mesures effectuées, pour la résolution QCIF, chacun des modèles FR a figuré dans le groupe des modèles les plus performants le nombre de fois indiqué ci-dessous:

Statistique	Modèle A	Modèle B	Modèle C	Modèle D	PSNR
Corrélation	9	11	12	4	1
RMSE ⁽¹⁾	7	10	11	2	1
Proportion de points éloignés	10	11	12	8	4

⁽¹⁾ RMSE: erreur quadratique moyenne.

NOTE 1 – A titre d'indication générale, les petites différences entre ces totaux ne dénotent pas de performances globalement différentes.

Analyse secondaire

Au cours de l'analyse secondaire, on établit la moyenne de toutes les séquences vidéo associées à chaque système (ou condition) vidéo et on peut déterminer ainsi le succès avec lequel le modèle mesure la performance moyenne du circuit fictif de référence (HRC). Les tableaux qui suivent indiquent les corrélations moyennes obtenues lors de l'analyse secondaire pour chaque modèle et chaque résolution.

Corrélation VGA

	Modèle A	Modèle B	Modèle C	Modèle D	PSNR
Moyenne	0,891	0,914	0,903	0,864	0,809

Corrélation CIF

	Modèle A	Modèle B	Modèle C	Modèle D	PSNR
Moyenne	0,915	0,919	0,913	0,892	0,817

Corrélation QCIF

	Modèle A	Modèle B	Modèle C	Modèle D	PSNR
Moyenne	0,942	0,937	0,920	0,893	0,882

Annexe 2

Modèle A

Le modèle A est constitué de trois modules logiciels distincts: un module d'alignement des signaux vidéo, un module de calcul des caractéristiques temporelles/spatiales, et un module d'évaluation subjective de la qualité vidéo (Fig. 1).

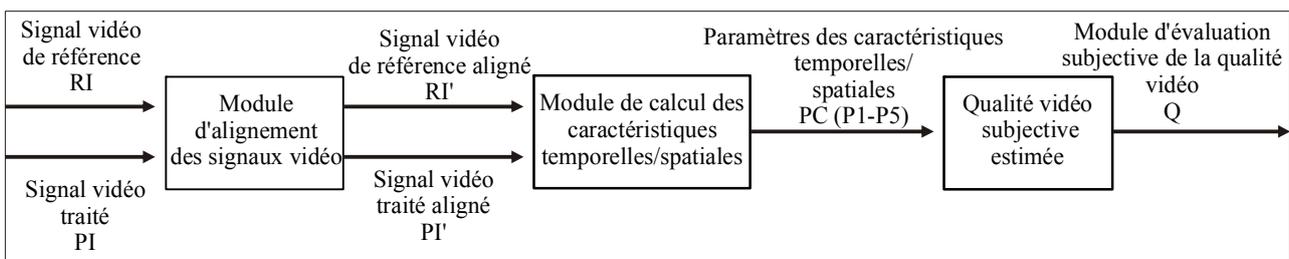
Le module d'alignement des signaux vidéo comprend un processus de macro alignement et un processus de micro alignement. Le processus de macro alignement consiste à établir une correspondance entre les pixels de signaux vidéo de référence (RI) et ceux de signaux vidéo traités (PI) dans l'espace et dans le temps et à filtrer les séquences vidéo pour rendre compte de l'influence de la saisie vidéo et du post-traitement du décodeur. Le processus de micro alignement consiste à établir une correspondance entre les trames de séquences vidéo de référence et celles de séquences de vidéo traitées pour rendre compte de l'influence du décalage et du gel d'images vidéo une fois le processus de macro alignement terminé.

Le module de calcul des caractéristiques temporelles/spatiales calcule un paramètre de dégradation spatiale et un paramètre de dégradation temporelle (PC) à partir d'un signal vidéo de référence aligné (RI') et d'un signal vidéo traité aligné (PI'). Le paramètre de dégradation spatiale repose sur quatre paramètres qui rendent compte, soit de la présence de bruit global, de contours perturbateurs ou de distorsion localisée du mouvement, soit d'une distorsion localisée spatialement. Le paramètre de dégradation temporelle, calculé en faisant la somme pondérée des durées des gels d'image, rend compte du gel d'image et de la variation du débit de trame.

Le module d'évaluation subjective de la qualité vidéo détermine la qualité (Q) vidéo objective en fonction des paramètres mentionnés précédemment.

FIGURE 1

Schéma fonctionnel du modèle A



BT.1866-01

Voir l'Annexe A de la Recommandation J.247 (08/2008) de l'UIT-T pour une description complète du modèle A.

Annexe 3

Modèle B

Le principe de base du modèle B est illustré à la Fig. 2.

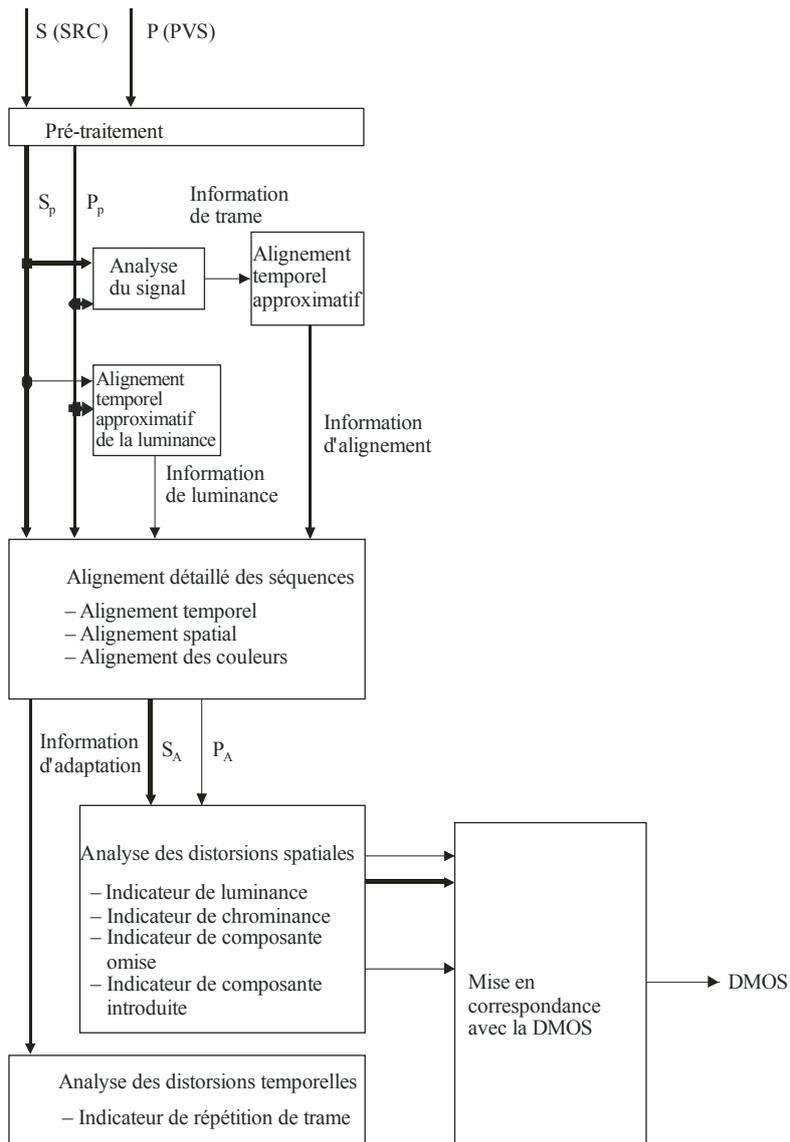
Au stade du pré-traitement, une région d'intérêt (ROI) spatiale est extraite du signal de référence et du signal d'essai. Tous les calculs ultérieurs sont effectués uniquement sur cette ROI, laquelle est représentée par les signaux recadrés S_p et P_p . Le pré-traitement est suivi d'un alignement approximatif des séquences d'entrée dans le domaine temporel et celui de la luminance. L'«information de luminance et d'alignement», obtenue grâce à ces modules est utilisée au cours du processus suivant d'«alignement détaillé des séquences», qui effectue l'alignement temporel, une image après l'autre, des deux séquences vidéo, et compense les décalages spatiaux et les différences de couleur et de brillance en fonction des évaluations de l'histogramme. L'«alignement détaillé des séquences» permet d'obtenir l'«information d'adaptation», qui est utilisée pour déterminer l'impact perceptuel des dégradations temporelles, ainsi que les séquences S_A et P_A recadrées et alignées.

Les distorsions spatiales sont analysées également au moyen du bloc «d'analyse des distorsions spatiales», lequel calcule les différences perceptuelles entre les séquences dans le domaine spatial et fournit quatre indicateurs de distorsion.

L'«information d'adaptation» fait ensuite l'objet d'une «analyse des distorsions temporelles» perceptuelles, qui permet d'obtenir un indicateur représentatif des répétitions de trames et d'autres distorsions temporelles.

Au cours de la dernière étape du modèle C, les cinq indicateurs calculés ci-dessus sont pondérés par des fonctions logistiques et conjugués pour fournir la note finale de la PEVQ (évaluation perceptuelle de la qualité vidéo), qui se rapproche de très près de la note d'opinion moyenne (MOS) obtenue à la suite de tests subjectifs.

FIGURE 2
Aperçu général du modèle B



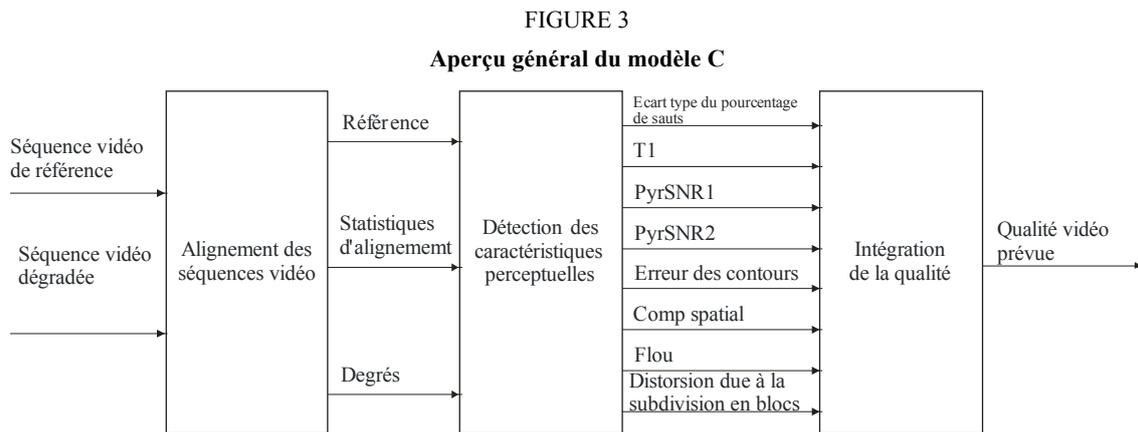
BT.1866-02

Voir l'Annexe B de la Recommandation J.247 (08/2008) de l'UIT-T pour une description complète du modèle B.

Annexe 4

Modèle C

Un aperçu général du modèle C est donné à la Fig. 3. Ce modèle comprend trois stades principaux: 1) Alignement de séquences vidéo, 2) détection des caractéristiques perceptuelles et 3) intégration de ces caractéristiques dans une note de prévision de la qualité globale.



BT.1866-03

Le premier stade consiste à aligner les séquences vidéo en associant temporairement chaque trame de la séquence vidéo dégradée à la trame de la séquence vidéo de référence dont elle se rapproche le plus. L'alignement des séquences vidéo est nécessaire pour mesurer correctement la qualité. A partir de la séquence vidéo de référence d'origine et de la séquence vidéo dégradée, on obtient une séquence vidéo de référence adaptée et une séquence vidéo dégradée adaptée, ainsi que des informations sur l'alignement. La séquence vidéo de référence adaptée contient toutes les trames de référence qui ont été adaptées aux trames de la séquence vidéo dégradée. La séquence vidéo dégradée adaptée contient, elle, toutes les trames dégradées pour lesquelles une trame de référence a été identifiée. L'algorithme d'alignement des séquences vidéo peut supporter des décalages temporels et spatiaux variables dans le temps entre la séquence vidéo de référence et la séquence vidéo dégradée.

Le deuxième stade consiste à analyser un ensemble de caractéristiques perceptuellement significatives extraites soit directement de la séquence vidéo dégradée, soit d'une comparaison entre la séquence vidéo de référence et la séquence vidéo dégradée. Ces caractéristiques perceptuelles ne reposent sur aucune hypothèse quant au type de contenu de la séquence vidéo ou à la manière dont cette séquence a été codée et transmise à l'utilisateur final. Les caractéristiques perceptuelles font office de couche sensorielle, qui élimine par filtrage les composantes de l'image auxquelles le spectateur humain n'est pas sensible. Ceci est dû au fait que les codecs vidéo réduisent la quantité de données à coder en appliquant des techniques de traitement des images pour éliminer les composantes du signal vidéo dont l'absence aurait le moins de chances d'être perçue par les spectateurs. L'incorporation d'un modèle du système visuel humain permet à ce stade de l'algorithme de déterminer le degré d'efficacité de ces processus et d'identifier les erreurs visibles. Si le schéma de codage ne permet pas d'atteindre cet objectif, l'utilisateur final percevra les dégradations, qui feront donc partie de la séquence de sortie de la couche sensorielle.

Le dernier stade consiste à conjuguer les paramètres perceptuels pour produire une prévision globale unique de la qualité vidéo. La forme optimale de la fonction d'intégration a été obtenue en

appliquant ce modèle à une longue série de tests subjectifs (base d'apprentissage) et en vérifiant ses performances au cours d'une série de tests non connus (base de validation).

Le format de la séquence vidéo de référence et de la séquence dégradée d'entrée utilisé avec le logiciel du modèle soumis à l'évaluation du VQEG au cours de la Phase I (Multimédias) était un format audio interactif non compressé avec espace chromatique UYVY (YUV 4:2:2), tel que défini dans le programme de tests multimédias du VQEG. Le modèle d'évaluation de la qualité est toutefois indépendant de ce format et peut donc être appliqué à d'autres formats (par ex. au format RGB24 AVI non compressé) à condition d'utiliser d'abord un filtre d'entrée adapté (par ex. un filtre de conversion de l'espace chromatique ou un lecteur de fichier).

Le format de la séquence vidéo de référence et celui de la séquence vidéo dégradée doivent être progressifs. Le débit de trame absolu de la séquence vidéo dégradée doit être identique à celui de la séquence vidéo de référence (par ex. 25 fps), le débit de trame absolu étant le nombre de trames (progressives) par seconde. Le débit de trame effectif de la séquence vidéo dégradée peut, par contre, être différent du débit de trame de la séquence vidéo de référence, le débit de trame effectif étant le nombre (moyen) de trames spécifiques par seconde. Le débit de trame effectif de la séquence vidéo dégradée peut également être variable dans le temps. L'exemple ci-après est fourni à titre d'illustration. La séquence vidéo de référence (A) a une fréquence de trame absolue de 25 fps et est codée avec une fréquence de trame (cible) effective de 12,5 fps (B). La séquence vidéo codée est lue (c'est-à-dire décodée) et saisie au rythme de 25 fps (C). Une trame sur deux de la séquence C est donc identique à la précédente. La séquence vidéo de référence et la séquence vidéo dégradée d'entrée appliquées au modèle d'évaluation de la qualité sont A et C respectivement.

Voir l'Annexe C de la Recommandation J.247 (08/2008) de l'UIT-T pour une description complète du modèle C.

Annexe 5

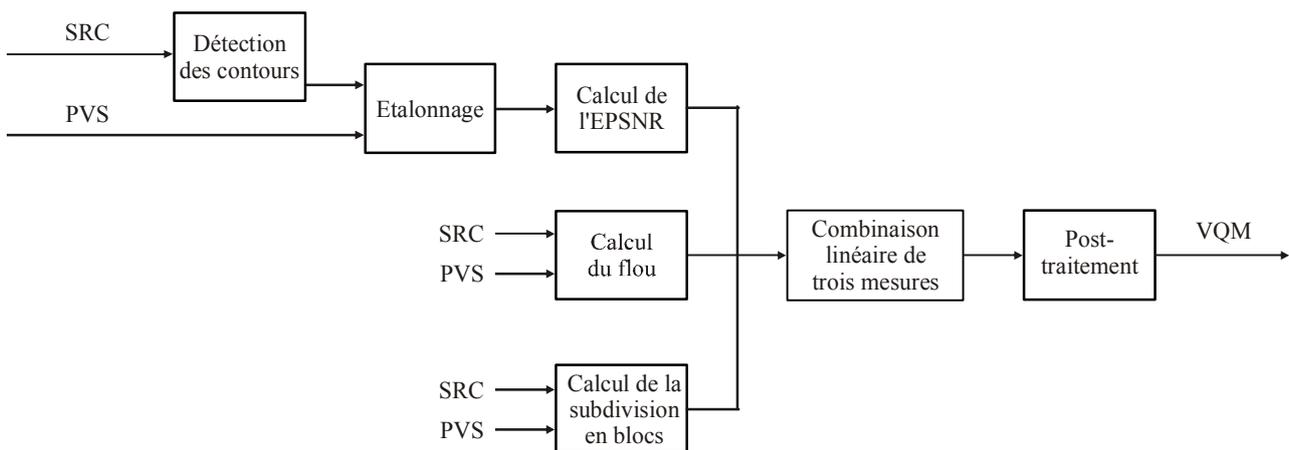
Modèle D

On observe que le système visuel humain est sensible aux dégradations autour des contours. On observe en outre que les algorithmes de compression vidéo ont tendance à produire davantage de défauts (artéfacts) autour des zones des contours. Sur la base de cette observation, le modèle fournit une méthode de mesure objective de la qualité vidéo qui permet de mesurer les dégradations autour des contours. Dans ce modèle, on applique tout d'abord un algorithme de détection des bords à la séquence vidéo source pour localiser les zones des bords. Ensuite, on mesure la dégradation de ces zones des bords en calculant l'erreur quadratique moyenne. A partir de cette erreur on calcule le rapport PSNR basé sur la dégradation des contours (EPSNR). Le modèle calcule, en plus, deux caractéristiques supplémentaires, qui sont conjuguées à l'EPSNR pour donner la mesure finale de la qualité vidéo (VQM).

La Fig. 4 présente le schéma fonctionnel du modèle D, basé sur la dégradation des contours d'un modèle à image de référence complète (FR) qui accepte deux entrées: une séquence vidéo source (SRC) et une séquence vidéo traitée (PVS).

FIGURE 4

Schéma fonctionnel du modèle D



BT.1866-04

Voir l'Annexe D de la Recommandation J.247 (08/2008) de l'UIT-T pour une description complète du modèle D.