

RECOMMANDATION UIT-R BT.1129-2

ÉVALUATION SUBJECTIVE DES SYSTÈMES DE TÉLÉVISION NUMÉRIQUE
À DÉFINITION NORMALE (TVDN)

(Question UIT-R 211/11)

(1994-1995-1998)

L'Assemblée des radiocommunications de l'UIT,

considérant

- a) qu'un certain nombre d'administrations et d'organisations dans le monde évaluent actuellement des systèmes numériques et que, dans de nombreuses parties du monde, la radiodiffusion numérique va sans doute devenir le premier média du siècle prochain;
- b) que les évaluations subjectives sont essentielles pour la conception et la comparaison des systèmes numériques;
- c) que la Recommandation UIT-R BT.500 recommande de préférer un certain nombre de méthodes d'évaluation subjective qui conviennent en grande partie à la télévision numérique;
- d) que, par leur nature et grâce aux techniques de transmission utilisées, les systèmes numériques permettent d'introduire dans le canal de radiodiffusion soit des trains de programmes multiples, soit des schémas de codage échelonnables ou hiérarchiques;
- e) que dans les systèmes de télévision numérique, les contenus des programmes peuvent influencer grandement sur la qualité d'image,

recommande

- 1 d'aligner les méthodes générales d'évaluation subjective des systèmes à définition standard numérique, y compris ceux qui permettent l'utilisation de trains de programmes multiples ou de schémas de codage échelonnables ou hiérarchiques, sur celles que décrit la Recommandation UIT-R BT.500;
- 2 d'aligner les procédures propres à l'évaluation subjective des systèmes numériques, y compris ceux qui permettent l'utilisation de trains de programmes multiples ou de schémas de codage échelonnables ou hiérarchiques, sur celles décrites à l'Annexe 1.

ANNEXE 1

1 Introduction

La présente annexe, qui est destinée à être utilisée conjointement avec la Recommandation UIT-R BT.500, fournit des précisions sur l'application des méthodes générales que décrit la Recommandation à l'évaluation subjective des systèmes numériques de qualité égale ou presque équivalente à celle des systèmes de télévision classiques. Les éléments de procédure présentés ici, ainsi que les informations générales correspondantes, concernent les essais de codecs (ou de systèmes) servant à transmettre des images produites conformément à la Recommandation UIT-R BT.601 dans les applications de contribution et de distribution ainsi que d'émission.

Pour les applications de distribution, la qualité peut être définie par rapport au jugement subjectif d'observateurs. Des codecs de ce type peuvent donc théoriquement être évalués subjectivement en fonction des spécifications de qualité ainsi établies. La qualité d'un codec destiné à des applications de contribution ne pourrait cependant pas en théorie, être définie en termes de caractéristiques subjectives, car le signal qu'il produit n'est pas destiné au visionnage immédiat, mais à la postproduction, à l'enregistrement et/ou au codage pour transmission ultérieure. Du fait qu'il est difficile de définir cette qualité pour une série d'opérations de postproduction différentes, le principe choisi a consisté à définir les performances d'une chaîne d'appareils, y compris pour la fonction de postproduction, considérée comme représentative

d'une application de contribution réelle. Cette chaîne pourrait normalement être constituée d'un codec, puis d'une fonction de postproduction studio (ou d'un autre codec dans le cas de l'évaluation de la qualité intrinsèque de la contribution), puis d'un autre codec avant présentation du signal à l'observateur. L'adoption de cette stratégie pour la spécification des codecs destinés aux applications de contribution a pour conséquence que les procédures de mesure décrites dans la présente Recommandation peuvent aussi être utilisées pour les évaluer.

Des travaux sont en cours, mais l'on ne dispose pas actuellement de résultats suffisants pour pouvoir préciser les méthodes d'évaluation objective de la qualité d'image des codecs. En matière d'évaluations subjectives, pour lesquelles on dispose d'une expérience considérable, on peut faire des recommandations sur les conditions expérimentales et sur la méthodologie. Il ne faut cependant pas oublier, lorsqu'on fixe des objectifs de qualité ou de dégradation, que les méthodes existantes peuvent donner non des notes subjectives absolues, mais des résultats partiellement affectés par les conditions de référence ou d'ancrage choisies. On peut adopter les mêmes méthodologies pour des codecs à longueur de mot fixe ou variable, inter ou intratrames, mais leur nature pourra influencer le choix des séquences d'essai.

Actuellement, la méthode la plus fiable pour le classement de codecs de haute qualité consiste à évaluer tous les systèmes en présence simultanément et dans des conditions identiques. Lorsqu'il n'existe que de très légères différences de qualité, les essais réalisés indépendamment ne peuvent donner qu'une indication et non une preuve indiscutable de supériorité.

Il pourrait être utile d'évaluer subjectivement les dégradations en fonction du taux d'erreur binaire sur la liaison entre le codeur et le décodeur. On ne possède actuellement qu'une connaissance expérimentale des statistiques d'erreurs réelles insuffisante pour pouvoir recommander des paramètres d'un modèle rendant compte du groupement des erreurs ou des salves d'erreurs. Tant que l'on ne disposera pas de données suffisantes, la loi de Poisson pourra être appliquée aux erreurs.

2 Conditions d'observation

Pour les évaluations subjectives, les conditions générales d'observation sont celles du § 2.1 de la Recommandation UIT-R BT.500. Les conditions d'observation propres aux évaluations subjectives des systèmes numériques sont indiquées dans les paragraphes qui suivent.

2.1 En laboratoire

Le laboratoire a pour objet de fournir les conditions critiques propices à l'examen des systèmes. Les conditions d'observation propres aux évaluations subjectives en laboratoire sont indiquées dans le Tableau 1 qui suit.

TABLEAU 1

**Conditions d'observation propres aux évaluations
subjectives des systèmes numériques en laboratoire**

Condition	Élément	Valeurs
a	Rapport de la distance d'observation à la hauteur de l'image	$4 H$ et $6 H$ ¹⁾
b	Luminance de crête de l'écran	70 cd/m^2
c	Angle d'observation sous-tendu par la zone d'arrière-plan qui répond aux spécifications	$\geq 43^\circ$ (hauteur) $\times 57^\circ$ (largeur)
d	Visualisation	Haute qualité, taille d'écran d'au moins 50 cm (20 pouces) ²⁾

¹⁾ $6 H$ est la distance préférée pour l'évaluation de systèmes de télévision à définition standard numérique; toutefois, on peut également recourir à des observateurs situés à une distance de $4 H$ à condition que les résultats soient consignés séparément.

²⁾ Il est prouvé que la taille d'écran peut influencer sur les résultats des évaluations subjectives: les expérimentateurs sont donc priés d'indiquer explicitement la taille d'écran ainsi que la marque et le modèle des écrans qu'ils utilisent dans leurs expériences.

2.2 A domicile

Cet environnement est censé offrir un moyen d'évaluer la qualité de la chaîne TV numérique du point de vue du consommateur. Les conditions d'observation propres aux évaluations subjectives des systèmes de télévision à définition standard à domicile sont indiquées dans le Tableau 2.

TABLEAU 2

Conditions d'observation propres aux évaluations subjectives des systèmes numériques à domicile

Condition	Élément	Valeurs
a	Rapport de la distance d'observation à la hauteur de l'image	6 <i>H</i>
b	Taille de l'écran pour un format 4/3	De 25 à 29 pouces ¹⁾
c	Taille de l'écran pour un format 16/9	De 32 à 36 pouces ¹⁾
d	Moniteur standard	TV à définition standard
e	Luminance de crête de l'écran	200 cd/m ²
f	Eclairage de l'environnement sur l'écran (la lumière incidente de l'environnement qui arrive sur l'écran doit être mesurée perpendiculairement sur l'écran)	200 Lux

¹⁾ Cet écran satisfait aux règles de la distance préférée d'observation; celle-ci est égale à 6 *H*.

3 Méthodes d'évaluation

3.1 Evaluation de la qualité intrinsèque de l'image

Lorsqu'on évalue un codec destiné à la distribution, cette qualité est celle d'images décodées après un seul passage dans une paire de codecs. Pour ceux affectés aux contributions, la qualité intrinsèque peut être évaluée à la sortie de plusieurs codecs en cascade, pour simuler des conditions d'exploitation types.

Lorsque la plage de qualité à évaluer est étroite, comme ce sera normalement le cas pour des codecs de télévision, la méthodologie à utiliser est celle de la variante II de la méthode double stimulus utilisant une échelle de qualité continue décrite dans la Recommandation UIT-R BT.500. La séquence de source initiale sera utilisée comme référence. La durée des séquences de présentation est encore à l'étude. Dans de récentes expériences sur des codecs destinés à la vidéo en composantes 4:2:2, il a été jugé avantageux de modifier la présentation par rapport à celle qui est décrite dans la Recommandation UIT-R BT.500. Des images composites ont été utilisées comme référence supplémentaire pour obtenir un niveau de qualité plus faible permettant d'évaluer la qualité du codec.

Il est recommandé d'utiliser dans le cadre de l'évaluation au moins six séquences d'images, plus une destinée à l'entraînement avant le début de l'expérience. Les séquences doivent être comprises entre "moyennement critique" et "critique" pour l'application à débit binaire réduit considérée.

Dans le cadre d'évaluations subjectives, on peut choisir des images fixes et des séquences mobiles parmi celles énumérées à l'Annexe 1 de la Recommandation UIT-R BT.1210. On notera, à ce propos, que les images et les séquences enregistrées sous forme numérique sont les signaux source les plus facilement reproductibles et qu'il convient donc de les préférer pour les évaluations. Les laboratoires peuvent s'échanger des éléments d'essai comme ceux que décrit la Recommandation UIT-R BT.1210 afin de donner plus de valeur aux comparaisons entre systèmes. Le format de bande D-1 4:2:2 décrit dans la Recommandation UIT-R BT.657 devrait servir à de tels échanges lorsque les magnétoscopes de ce type seront disponibles sur le marché, à des prix économiques. Les échanges sont aussi possibles en format de bande d'ordinateur.

D'un bout à l'autre de la présente annexe, on insiste sur l'importance qu'il y a de tester des codecs numériques au moyen de séquences d'images critiques pour la réduction de débit binaire en télévision. On peut donc raisonnablement se demander quel est le niveau critique d'une séquence donnée pour une application déterminée de la réduction du débit

inaire ou si une séquence est plus critique qu'une autre. Une réponse simple, mais pas particulièrement utile, est de dire que la notion de niveau critique revêt une signification très différente selon les codecs. C'est ainsi qu'une image fixe très détaillée pourrait bien être critique pour un codec intratrame, tandis que pour un système intertrames capable d'exploiter les similitudes entre images, la même scène ne poserait aucun problème. Certains types de séquences où figurent des textures mouvantes et des mouvements complexes seront critiques pour toutes les catégories de codecs et ce sont donc eux qu'il est le plus utile de produire ou de reconnaître. Les mouvements complexes peuvent être d'une forme prévisible par l'observateur, mais non par les algorithmes de codage; c'est le cas des déplacements tortueux périodiques.

Un examen de mesures statistiques possibles du niveau critique de l'image (obtenues par exemple au moyen de méthodes corrélatives, spectrales, entropiques conditionnelles, etc.) a permis de découvrir une mesure simple, mais utile basée sur une mesure de l'entropie adaptative intertrames-intratrame. Cette méthode a été utilisée pour "calibrer" les séquences d'images que l'UIT-R propose d'employer dans les essais de codecs à 34, 45 et 140 Mbit/s et s'est révélée utile dans le choix de celles à adopter. La manière la plus facile de réaliser ces mesures sur les séquences est de transférer celles-ci à des ordinateurs de traitement de l'image, puis de les soumettre à un logiciel d'analyse.

Lorsqu'il n'est pas possible de recourir à ces techniques, on pourra utiliser les indications générales qui suivent sur la manière de choisir des images critiques.

a) *Codecs intratrame à mots de longueur fixe*

S'il est possible et justifié d'évaluer ces codecs sur des images fixes, il est recommandé d'employer des séquences mobiles car le résultat du bruit de codage est plus facile à observer et cette solution est plus représentative de la télévision réelle. Si l'on emploie des images fixes dans la simulation de codecs sur ordinateur, le traitement doit être effectué sur la totalité de la séquence d'évaluation afin de conserver l'aspect temporel de tout bruit à la source, par exemple. Les scènes choisies doivent contenir le plus possible des détails suivants: régions texturées fixes et mobiles (certaines colorées), objets mobiles et immobiles avec des arêtes aiguës fortement contrastées (quelques-unes en couleur) dans diverses orientations, zones uniformes fixes d'un gris moyen. Au moins une des séquences de l'ensemble doit présenter un bruit de source à peine perceptible et une, au minimum, doit être artificielle, c'est-à-dire produite par ordinateur pour s'affranchir des imperfections des caméras, comme celles dues à l'ouverture d'analyse et au traînage.

b) *Codecs intertrames à mots de longueur fixe*

Les scènes choisies doivent toutes comporter des mouvements et le plus possible des détails suivants: régions texturées mobiles (certaines en couleur), objets comportant des arêtes et fortement contrastés (certains colorés) se déplaçant dans diverses orientations, perpendiculaires aux arêtes. Une séquence au moins dans l'ensemble doit comporter un bruit à la source à peine perceptible et une autre (au minimum) doit être artificielle.

c) *Codecs intratrame à mots de longueur variable*

Il est recommandé d'essayer ces codecs avec des séquences d'images mobiles pour les mêmes raisons que dans le cas des codecs à longueur de mot fixe. Il est à noter que, par leur codage à longueur de mot variable et leur mémoire tampon, ces codecs peuvent répartir dans toute l'image la capacité binaire disponible pour le codage. C'est ainsi que si la moitié d'une image est constituée d'un ciel monotone, que l'on peut coder avec une faible capacité, on en économise pour le reste de l'image qui peut alors être reproduite avec une grande qualité, même si elle est critique. La conclusion importante à tirer est que, si une séquence d'images doit être critique pour un codec de ce type, toutes les parties de l'écran doivent comporter des détails avec des textures fixes et en mouvement et autant de couleurs qu'il est possible; il doit aussi y figurer des objets à contours fins et très contrastés. Au moins une séquence de l'ensemble doit présenter un bruit à la source à peine perceptible et une (au minimum) doit être artificielle.

d) *Codecs intertrames à mots de longueur variable*

Il s'agit là de la catégorie de codecs la plus élaborée et du genre qui exige les images les plus difficiles pour les pousser à leurs limites. Il faut non seulement que toutes les parties de l'image comportent des détails comme pour les codecs intratrame à mots de longueur variable, mais aussi que ces détails soient en mouvement. De plus, comme de nombreux codecs utilisent des méthodes de compensation des mouvements, les déplacements pendant la séquence doivent être complexes. On peut citer les exemples suivants: scènes avec changements de cadrage et de focales simultanés, scènes ayant comme fond un rideau structuré ou comportant des détails, agité par le vent, scènes avec des objets tournant dans les trois dimensions de l'espace, scènes où des objets détaillés accélèrent leurs mouvements de traversée de l'écran. Toutes les scènes doivent comporter d'importants mouvements d'objets à des vitesses différentes, des textures et des arêtes très contrastées, ainsi que des couleurs variées. Au moins une séquence de l'ensemble doit présenter un bruit de source à peine perceptible et une autre (au minimum) doit montrer des mouvements de caméra complexes synthétisés sur ordinateur à partir d'une image fixe naturelle, de manière qu'elle soit exempte de défauts dus au bruit et au traînage de la caméra. Enfin, une séquence au moins doit être entièrement produite sur ordinateur.

3.2 Evaluation de la qualité de l'image après post-traitement

Le but de cette évaluation est de juger si un codec destiné à des applications de contribution est bien adapté à cette fonction du point de vue d'opérations de post-traitement particulières telles qu'incrustations, ralenti ou recadrage électronique. Pour cette évaluation, le matériel employé doit au minimum assurer un passage dans le codec étudié, suivi par le post-traitement considéré, puis par l'observation. Il peut toutefois être plus représentatif des applications de contribution d'insérer d'autres codecs après le post-traitement.

La méthodologie à utiliser est celle de la variante II de la méthode à double stimulus utilisant une échelle de qualité continue. Dans ce cas, la référence sera cependant la séquence à la source soumise au même post-traitement que les images décodées. Si l'on juge avantageux de prévoir une référence de moindre qualité, cette dernière doit aussi être soumise au même post-traitement.

Les séquences d'essai nécessaires à l'évaluation du post-traitement sont soumises à des critères de criticité exactement identiques à ceux des séquences des autres applications numériques. Il peut cependant être difficile de les respecter dans les scènes d'avant-plan pour les incrustations, car elles comportent généralement d'importantes parties d'un bleu uniforme.

Du fait des contraintes matérielles que comporte la possibilité de devoir évaluer un codec avec plusieurs opérations de post-traitement, le nombre des séquences d'images utilisées peut être au minimum de trois, avec une supplémentaire pour la démonstration. La nature des séquences dépendra de l'opération de post-traitement étudiée, mais devra être comprise entre "modérément critique" et "critique" pour la télévision à débit binaire réduit et pour le procédé en question. Pour évaluer le ralenti, une vitesse de reproduction égale au 1/10 de celle de la source peut convenir.

3.3 Evaluation du comportement en présence de défauts

Quand on évalue subjectivement les défauts des images des codecs imputables à des imperfections dans le canal de transmission ou d'émission, il faut sélectionner au moins cinq taux d'erreur binaires ou situations de transmission ou d'émission données, mais de préférence davantage, répartis de façon à peu près logarithmique et couvrant bien toute la gamme où se produisent les dégradations dues au codec et comprise entre "imperceptible" et "très gênant".

Il est possible que l'on doive évaluer des codecs à des taux d'erreur binaires de transmission provoquant des distorsions passagères visibles si rares qu'il risque de ne pas s'en produire pendant la durée d'une séquence de 10 s. Il est alors évident que le rythme de présentation proposé ici ne convient pas.

Si l'on enregistre l'image sortant d'un codec avec un taux d'erreur binaire assez faible (donnant un petit nombre de distorsions passagères visibles pendant une période de 10 s) en vue d'un montage pour constituer une séquence d'évaluation subjective, il faut veiller à ce que l'enregistrement utilisé soit représentatif de l'image sortant du codec observé pendant une période de temps plus longue.

Comme il faut étudier la qualité du codec pour une série de taux d'erreur binaires de transmission, trois séquences, plus une pour la démonstration, suffiront probablement compte tenu des contraintes matérielles. La durée d'une séquence doit être d'environ 10 s, mais il convient de noter que les observateurs préféreront peut-être une durée de 15 à 30 s. Cette séquence doit être comprise entre "moyennement critique" et "critique" pour la télévision à débit binaire réduit.

Comme les essais couvriront toute la plage des dégradations, la méthode à double stimulus utilisant une échelle de dégradation est applicable et doit être utilisée.

3.4 Caractéristiques de défaillance fonction du contenu de l'image

Le concept général de caractéristiques de défaillance fonction du contenu de l'image est donné dans l'Appendice 1 de l'Annexe 1 de la Recommandation UIT-R BT.500. Pour appliquer ce concept aux systèmes de télévision numérique, il faut procéder de la manière suivante.

3.4.1 Définition de la criticité

Il faut définir une certaine mesure appelée "criticité", qui représente les caractéristiques du système de télévision numérique sous test et qui est déterminée par une mesure objective. Le système de télévision numérique servant d'exemple est le système MPEG-2 MP@ML et on applique la méthode - à quantificateur fixe - de détermination de la criticité basée sur l'entropie, qui est décrite dans la Recommandation UIT-R BT.1210.

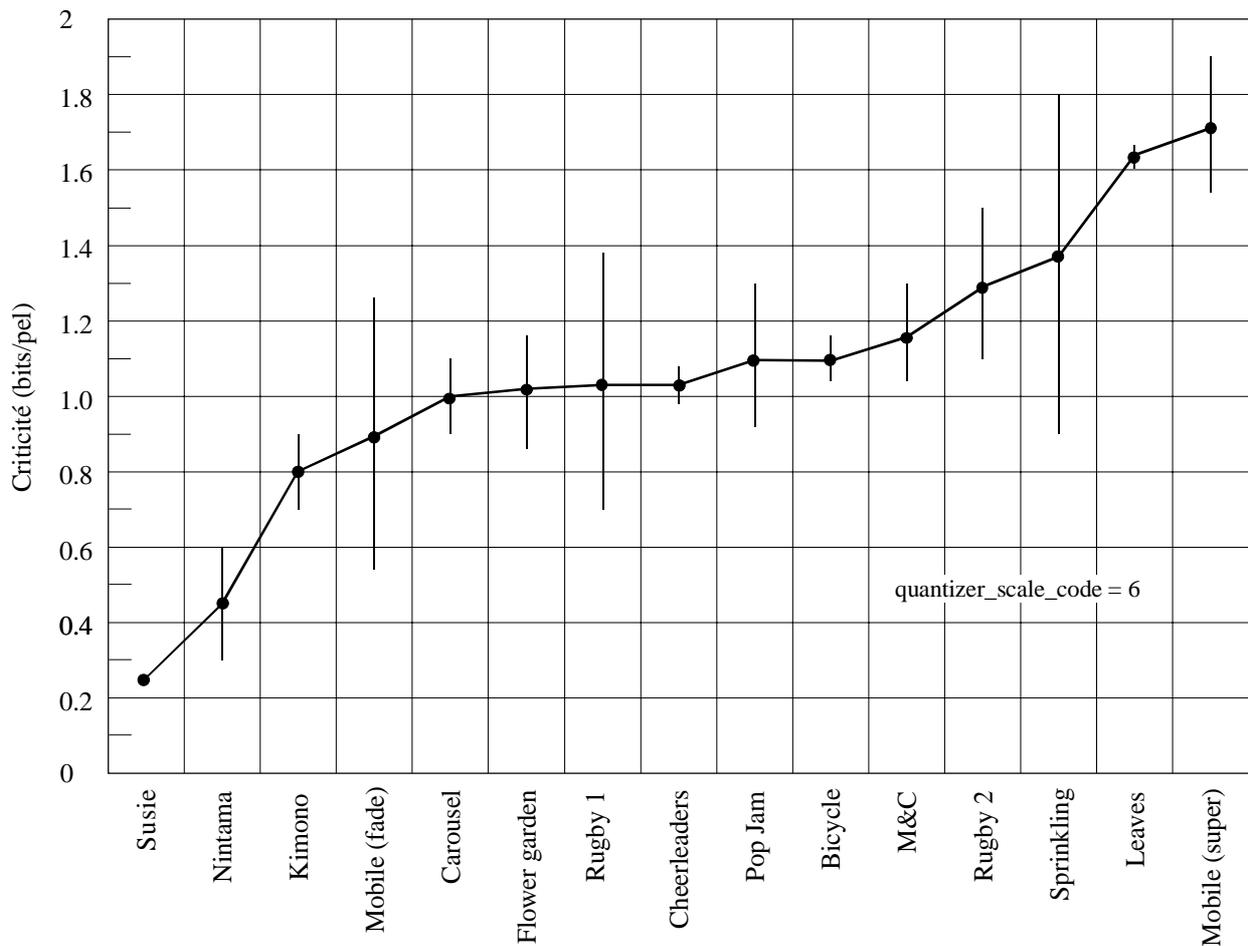
3.4.2 Procédure d'établissement des caractéristiques de défaillance fonction du contenu de l'image

- *Etape 1:* Mesurer la criticité des séquences d'essai utilisées pour l'évaluation subjective.

On mesure la criticité des séquences d'essai utilisées pour l'évaluation subjective décrite à l'étape 3 ci-dessous. La Figure 1 représente la moyenne et l'écart type de chaque séquence pour le système utilisé comme exemple. La criticité de la plupart des séquences est comprise entre 0,8 et 1,4 bits/pixel. L'écart type de certaines séquences est grand car le contenu de l'image varie beaucoup pendant la séquence.

FIGURE 1

Moyennes et écarts types de la criticité associée aux séquences d'essai



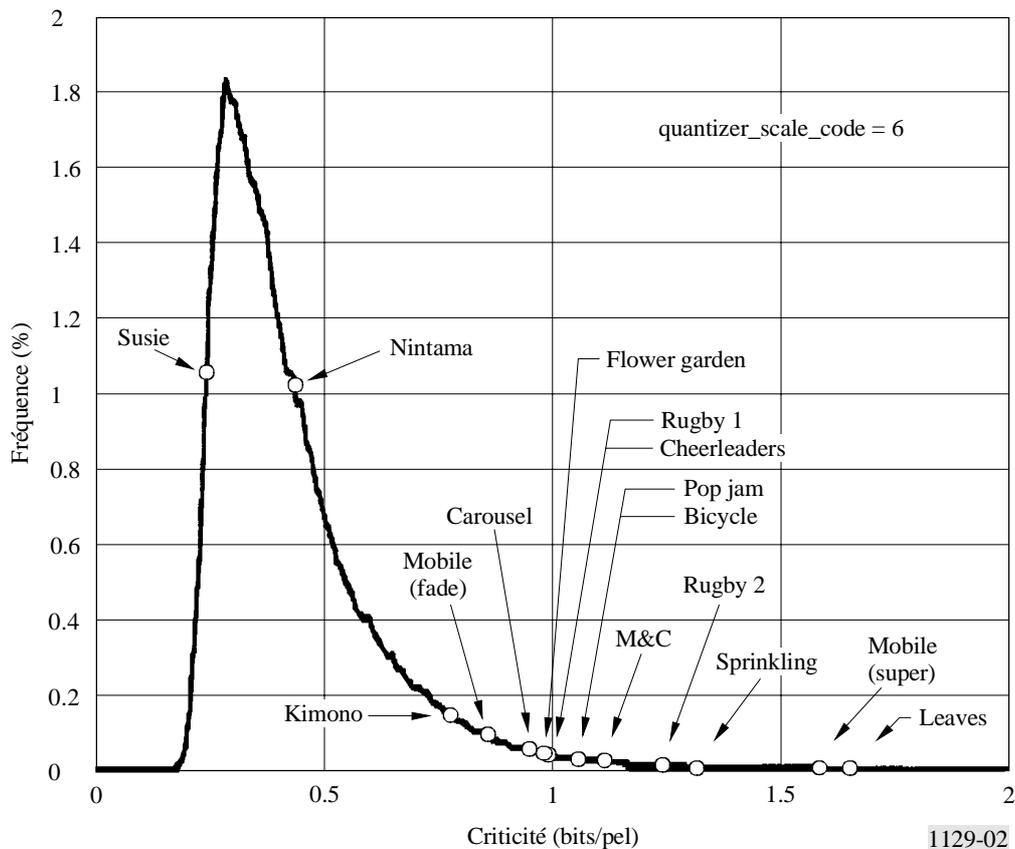
1129-01

- *Etape 2:* Mesurer, sur une longue période temporelle, la distribution de la criticité associée à des programmes de radiodiffusion.

La distribution de la criticité associée à des programmes de radiodiffusion télévisuelle est mesurée sur une période temporelle suffisamment longue, par exemple une semaine. La Figure 2 montre un exemple de distribution mesurée sur une semaine, soit un total de 130 h pour des signaux de radiodiffusion NTSC, qui ont été convertis en signaux à composantes de luminance et de chrominance (Y/C) pour la mesure. La fréquence d'occurrence de la criticité pour les programmes de radiodiffusion télévisuelle a été calculée tous les 5×10^{-3} bits/pixel. Cette figure montre aussi la criticité des séquences d'essai utilisées pour l'évaluation subjective.

FIGURE 2

**Distribution de la criticité associée à des programmes de radiodiffusion
et criticité des séquences d'essai**



- *Etape 3:* Faire une évaluation subjective de la qualité d'image du système sous test et en déduire une relation entre la criticité et la qualité d'image subjective.

La qualité d'image du système de télévision numérique est évaluée à l'aide de la méthode à double stimulus utilisant une échelle de qualité continue (DSCQS). La combinaison du résultat de l'évaluation subjective et de la criticité obtenue à l'étape 1 permet d'établir la relation entre la criticité et les notes de l'évaluation. La Figure 3 montre la qualité d'image du système utilisé comme exemple pour les débits binaires suivants: 4, 6, 9 et 15 Mbit/s. La différence de qualité (DSCQS %) sur la figure représente la dégradation par rapport à la référence: une séquence en composantes 4:2:2 d'origine. La Figure 4 montre la relation existant entre la criticité et la différence de qualité. Dans cet exemple, on a supposé une relation linéaire entre la criticité et la qualité d'image; les droites de régression ont été établies à l'aide de la méthode des moindres carrés. La droite de régression correspondant à chaque débit binaire est représentée sur la figure. En général, on peut appliquer une relation non linéaire en fonction des résultats de l'évaluation.

- *Etape 4:* Établir les caractéristiques de défaillance en fonction du contenu de l'image (qualité en fonction de la fréquence d'occurrence) en combinant les résultats de l'étape 3 (criticité en fonction de la qualité) et de l'étape 2 (criticité en fonction de la fréquence d'occurrence).

La combinaison des résultats obtenus aux étapes 2 et 3 permet d'établir les caractéristiques de défaillance en fonction du contenu de l'image, c'est-à-dire la distribution de la qualité d'image des programmes de télévision codés numériquement. La dégradation d'image dans les programmes de radiodiffusion télévisuelle est convertie en fréquence d'occurrence cumulative. La Figure 5 représente les caractéristiques de défaillance en fonction du contenu pour le système utilisé comme exemple.

FIGURE 3

Résultat de l'évaluation subjective (système MP@ML à 6H)

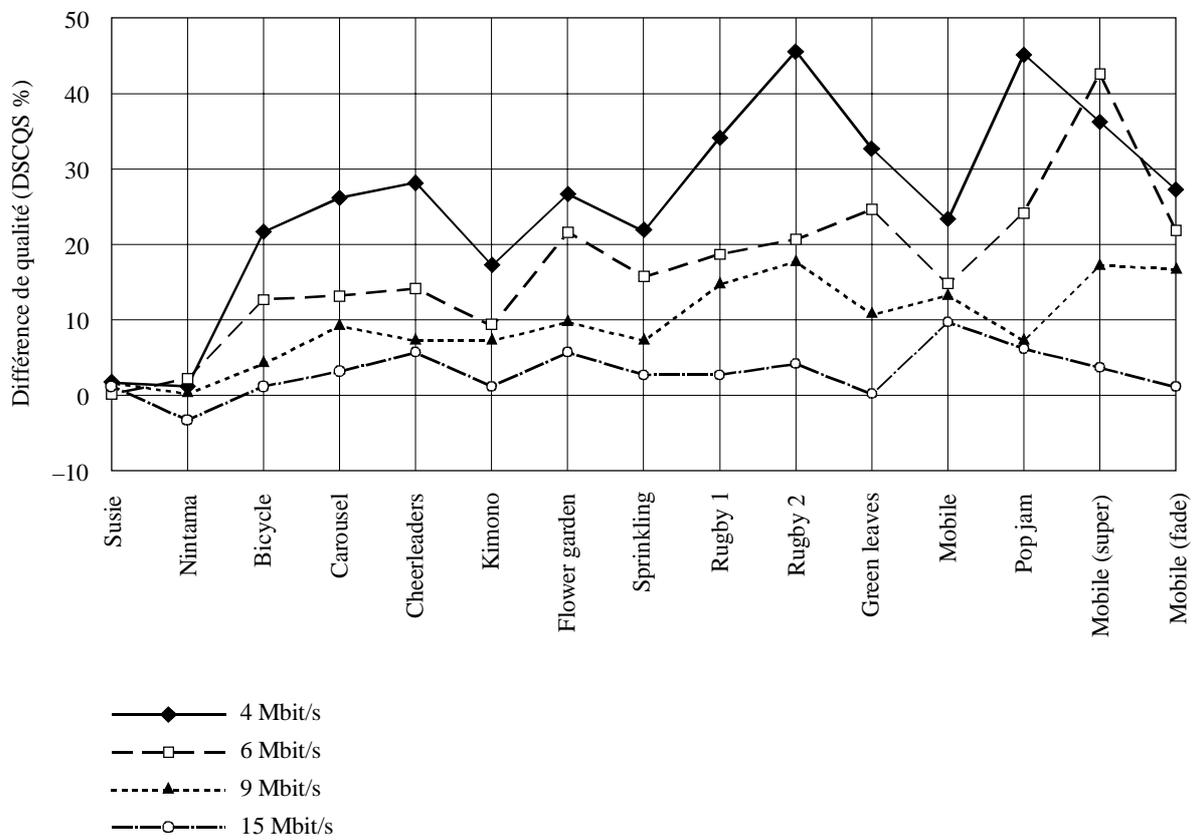
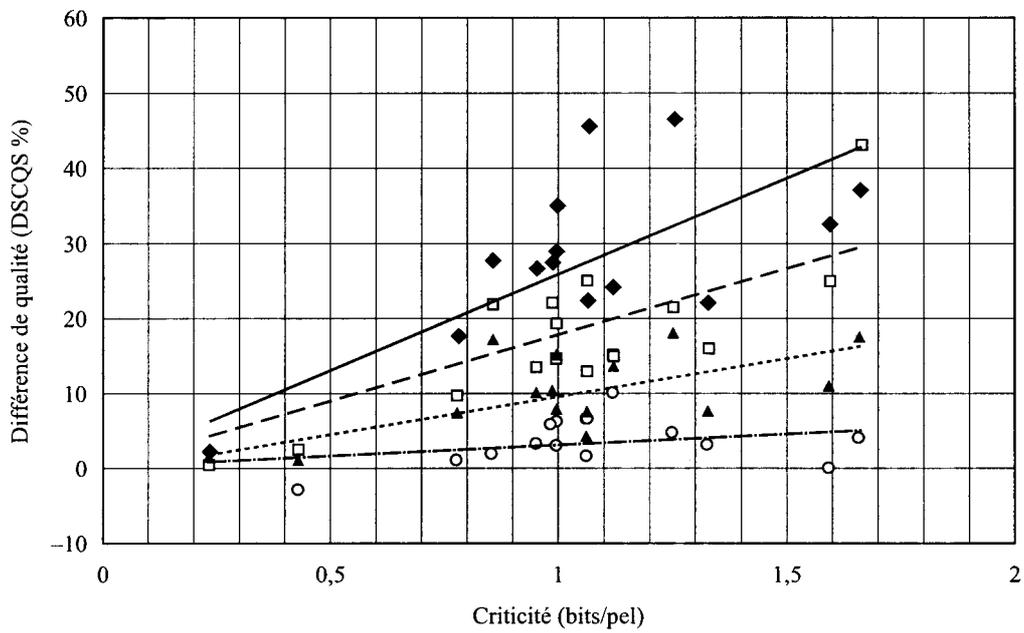


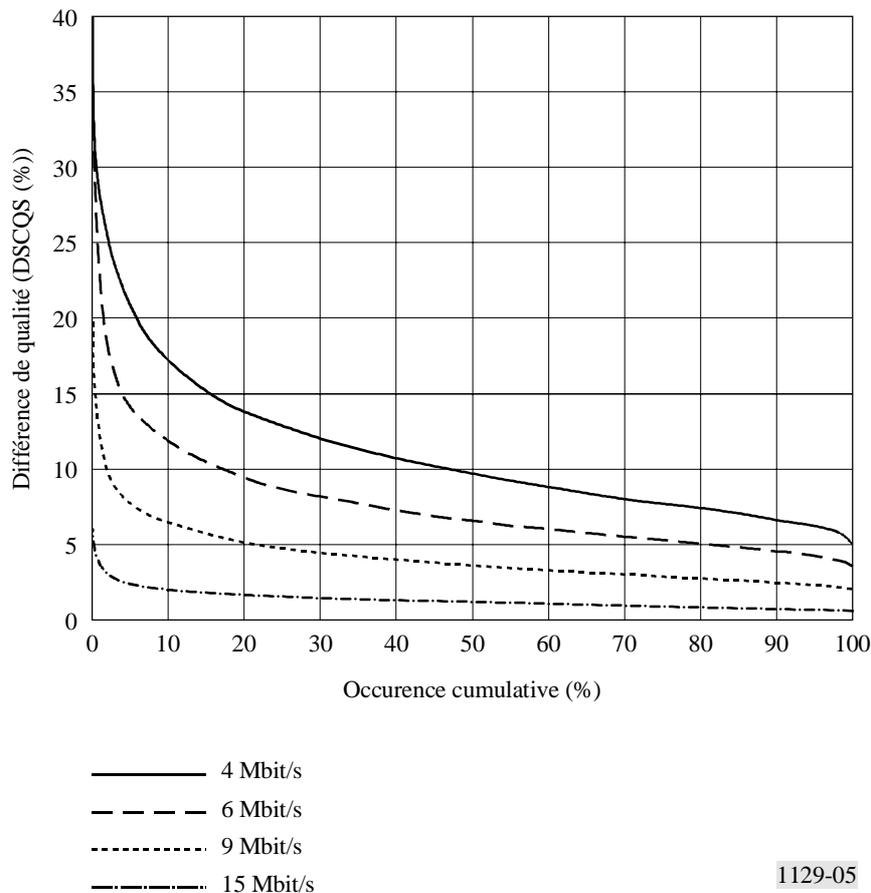
FIGURE 4
 Relation entre la criticité et la note d'évaluation (MP@ML at 6H)



- ◆ 4 Mbit/s
- 6 Mbit/s
- ▲ 9 Mbit/s
- 15 Mbit/s

FIGURE 5

Fréquence d'occurrence cumulative de la dégradation d'image
(système MP@ML à 6 H)



4 Notes concernant les applications

Lorsqu'il est inutile de juger la qualité ou la dégradation absolue d'un codec et qu'on ne demande que de classer les systèmes ou lorsqu'on souhaite confirmer un classement obtenu par la méthode du double stimulus, il faut utiliser une technique de comparaison entre couples de stimulus.

Telle qu'elle est décrite dans la Recommandation UIT-R BT.500, la méthode permet une comparaison précise et donne le moyen de déterminer la relation entre couples de systèmes. Il est possible d'étendre cette méthode au classement de la qualité ou de la dégradation pour plus de deux systèmes. Selon ce principe, on établit un classement global à partir de ceux donnés par les observateurs pour tous les couples possibles de séquences d'images.

L'analyse est compliquée par le fait qu'un observateur peut juger, par exemple, que l'image A est meilleure que l'image B, elle-même meilleure que l'image C, tout en préférant C à A. On parle alors de "triade intransitive".

Un problème posé par cette méthode tient au fait que le nombre de présentations nécessaires augmente avec le carré du nombre de séquences d'essai et de codecs, si bien que la mise en œuvre de cette méthode peut se révéler impossible.

Si le canal de radiodiffusion est utilisé pour acheminer soit des trains de programmes multiples, soit des schémas de codage échelonnables ou hiérarchiques, il peut se révéler nécessaire d'adapter la méthode d'évaluation pour tenir compte de ce qui suit:

- pour savoir si un service est acceptable, le critère n'est pas forcément la transparence du codage à la source; il peut s'agir en fait de la faculté du système, pour une attribution de débit binaire donnée, d'offrir une solution de rechange convenable au service classique. Dans les essais de qualité, il convient donc peut-être de prendre pour référence des images acheminées par un système classique dans des conditions de réception normales plutôt que des images sous forme numérique non comprimée. Il convient peut-être aussi d'utiliser des éléments d'essai choisis pour être représentatifs d'une gamme de contenus de programme actuels ou futurs (voir l'Appendice 1 de l'Annexe 1 de la

Recommandation UIT-R BT.500). Au cours des essais, les conditions d'observation doivent être celles que décrivent la Recommandation UIT-R BT.500 et le § 1 de la présente Annexe, alors que la méthode d'essai générale sera la méthode à double stimulus utilisant une échelle de qualité continue (§ 5 de la Recommandation UIT-R BT.500); et

- une question est de savoir si le système est à même de conserver complètement les trains de programmes en cas de pleine charge du canal et de défauts de transmission. Il convient peut-être donc, au cours des essais sur les dégradations, de maintenir le canal à pleine charge et d'avoir une gamme de niveaux de dégradation représentant celle qui se produit en général dans des conditions de réception normales (voir l'Appendice 2 de l'Annexe 1 de la Recommandation UIT-R BT.500). Au cours des essais, les conditions d'observation doivent être celles que décrivent la Recommandation UIT-R BT.500 et le § 1 de la présente Annexe, alors que la méthode d'essai générale sera la méthode à double stimulus utilisant une échelle de dégradation (§ 4 de la Recommandation UIT-R BT.500).

NOTE 1 – Lorsqu'on évalue des systèmes analogiques et numériques dans un même contexte, il importe de choisir un ensemble d'éléments d'essai présentant des difficultés proportionnées aux deux types de système. Dans ce cas, il peut être utile d'appliquer, en vue d'une analyse supplémentaire, la procédure d'échelonnement multidimensionnel mentionnée au Tableau 2 de l'Annexe 1 de la Recommandation UIT-R BT.500.
