UNION INTERNATIONALE DES TÉLÉCOMMUNICATIONS

# UIT-T P.930

(08/96)

SECTEUR DE LA NORMALISATION
DES TÉLÉCOMMUNICATIONS
DE L'UIT

SÉRIE P: QUALITÉ DE TRANSMISSION TÉLÉPHONIQUE

Qualité audiovisuelle dans les services multimédias

# Principes d'un système de dégradation de référence pour signaux vidéo

Recommandation UIT-T P.930

(Antérieurement Recommandation du CCITT)

RECOMMANDATIONS UIT-T DE LA SÉRIE P

**QUALITÉ DE TRANSMISSION TÉLÉPHONIQUE**

| | | |
|---|---|---|
| Vocabulaire et effets des paramètres de transmission sur l'opinion des usagers | Série | P.10 |
| Lignes et postes d'abonnés | Série | P.30 |
| | | P.300 |
| Normes de transmission | Série | P.40 |
| Appareils de mesures objectives | Série | P.50 |
| | | P.500 |
| Mesures électroacoustiques objectives | Série | P.60 |
| Mesures de la sonie vocale | Série | P.70 |
| Méthodes d'évaluation objective et subjective de la qualité | Série | P.80 |
| | | P.800 |
| **Qualité audiovisuelle dans les services multimédias** | **Série** | **P.900** |

*Pour plus de détails, voir la Liste des Recommandations de l'UIT-T.*

**RECOMMANDATION UIT-T P.930**

# PRINCIPES D'UN SYSTEME DE DEGRADATION DE REFERENCE POUR SIGNAUX VIDEO


## Résumé

La présente Recommandation décrit les principes d'un système de référence vidéo réglable que l'on peut utiliser pour produire les conditions de référence nécessaires afin de caractériser la qualité subjective d'images vidéo issues de systèmes vidéo numériques à compression. Un système de dégradation de référence pour signaux vidéo (RISV, *reference impairment system for video*) peut être utilisé afin de simuler les dégradations résultant de la compression de séquences vidéo, indépendamment de l'algorithme de compression. Les méthodes d'évaluation subjective sont décrites dans la Recommandation P.910.

L'Appendice I décrit le système VIRIS (Video Reference Impairment System élaboré par Bellcore), réalisation spécifique d'un système de dégradation de référence réglable pour signaux vidéo. Bien que les études effectuées par Bellcore aient été faites avec le MPEG-1, le système VIRIS peut aussi être utilisé avec d'autres schémas de compression comme celui de la Recommandation H.261.

Il est recommandé qu'un système RISV puisse produire les catégories ci-après de distorsion, séparément ou en combinaison, avec réglage indépendant de chaque niveau de dégradation:

a) Effets de dégradation dus aux conversions entre formats analogiques et numériques (par exemple, bruit et flou).

b) Effets de dégradation dus au codage et à la compression (par exemple, sautillement, gigue de contour et distorsion en mosaïque).

c) Effets de dégradation dus aux erreurs dans la voie de transmission (par exemple, blocs erronés).

Dans la présente Recommandation, cinq types de dégradation (distorsion en mosaïque, flou, gigue de contour, bruit et sautillement) sont définis et des méthodes générales d'application de ces dégradations sont fournies. L'Appendice I décrit une réalisation précise de ces dégradations dans le système VIRIS. D'autres dégradations feront l'objet d'un complément d'étude.

Du point de vue du visualisateur, les dégradations produites par le système RISV doivent être une bonne approximation des dégradations produites par les systèmes numériques de codage vidéo et de transmission.

Les trois applications possibles d'un système RISV sont les suivantes:

1) création de conditions de référence dans les essais subjectifs de systèmes vidéo numériques pour faire en sorte que la qualité des scènes présentées aux spectateurs couvre la gamme totale de qualité d'image;

2) définition de niveaux normaux de dégradation vidéo qui pourront être utilisés pour comparer des résultats d'essai subjectif;

3) quantification de la qualité perçue par l'utilisateur d'un système vidéo, par rapport à une référence connue.

Bien que les caractéristiques d'un système RISV soient décrites dans la présente Recommandation, il est nécessaire d'effectuer des essais de validation avant qu'une réalisation puisse être recommandée.



## Source

La Recommandation UIT-T P.930, élaborée par la Commission d'études 12 (1993-1996) de l'UIT-T, a été approuvée le 30 août 1996 selon la procédure définie dans la Résolution n° 1 de la CMNT.

## AVANT-PROPOS

L'UIT (Union internationale des télécommunications) est une institution spécialisée des Nations Unies dans le domaine des télécommunications. L'UIT-T (Secteur de la normalisation des télécommunications) est un organe permanent de l'UIT. Il est chargé de l'étude des questions techniques, d'exploitation et de tarification, et émet à ce sujet des Recommandations en vue de la normalisation des télécommunications à l'échelle mondiale.

La Conférence mondiale de normalisation des télécommunications (CMNT), qui se réunit tous les quatre ans, détermine les thèmes d'études à traiter par les Commissions d'études de l'UIT-T lesquelles élaborent en retour des Recommandations sur ces thèmes.

L'approbation des Recommandations par les Membres de l'UIT-T s'effectue selon la procédure définie dans la Résolution n° 1 de la CMNT (Helsinki, 1er-12 mars 1993).

Dans certains secteurs de la technologie de l'information qui correspondent à la sphère de compétence de l'UIT-T, les normes nécessaires se préparent en collaboration avec l'ISO et la CEI.

NOTE

Dans la présente Recommandation, l'expression «Administration» est utilisée pour désigner de façon abrégée aussi bien une administration de télécommunications qu'une exploitation reconnue.

## TABLE DES MATIÈRES


| | | Page |
|---|---|---|
| 1 | Domaine d'application | 1 |
| 2 | Références | 1 |
| 3 | Abréviations et définitions | 1 |
| 4 | Attributs d'un système de dégradation de référence pour la vidéo (RISV) | 2 |
| 4.1 | Gamme réglable de niveaux de dégradation vidéo | 3 |
| 4.2 | Applications possibles d'un système RISV | 3 |
| 5 | Dégradations produites par le système RISV | 4 |
| 5.1 | Distorsion en mosaïque | 4 |
| 5.2 | Flou | 5 |
| 5.3 | Gigue de contour | 5 |
| 5.4 | Bruit | 5 |
| 5.5 | Sautillement | 6 |
| 5.6 | Combinaison de dégradations | 6 |
| 6 | Conclusion | 7 |
| Appendice I | | 8 |
| VIRIS, mise en oeuvre spécifique d'un système RISV | | 8 |
| I.1 | Introduction | 8 |
| I.2 | Simulation des dégradations | 8 |
| I.2.1 | Distorsion en mosaïque | 9 |
| I.2.2 | Flou | 10 |
| I.2.3 | Gigue de contour | 11 |
| I.2.4 | Bruit | 13 |
| I.2.5 | Sautillement | 14 |
| I.3 | Calcul de la valeur de crête du rapport signal sur bruit (PSNR) | 15 |
| I.4 | Programme VIRIS | 15 |
| I.5 | Plan d'essai subjectif avec le système VIRIS | 17 |
| I.5.1 | Séquences d'images traitées par VIRIS | 17 |
| I.5.2 | Conditions d'essai VIRIS | 18 |
| I.5.3 | Méthode d'essai VIRIS | 19 |
| I.5.4 | Conditions de visualisation VIRIS | 20 |
| I.5.5 | Procédures d'essai VIRIS | 21 |
| I.5.6 | Résultats des essais subjectifs VIRIS | 21 |

**Page**

I.6 Plan d'essai subjectif de gigue de contour ..... 25
I.6.1 Séquences d'images avec gigue de contour ..... 26
I.6.2 Conditions d'essai VIRIS et préparation du matériel d'essai ..... 26
I.6.3 Méthode d'essai subjectif par gigue de contour ..... 27
I.6.4 Méthode d'essai subjectif pour la gigue de contour ..... 27
I.6.5 Résultats d'essai subjectif de gigue de contour ..... 28
I.7 Conclusions ..... 30

Bibliographie ..... 30

## Introduction

Les conditions vidéo de référence sont souvent utilisées dans des évaluations subjectives de la qualité vidéo parce que les jugements par évaluation selon des catégories peuvent varier d'un essai à l'autre ou d'un laboratoire à un autre, selon les particularités expérimentales (c'est-à-dire effets de dispersion et de fréquence, population des sujets, etc.). Les conditions de référence permettent d'effectuer des comparaisons entre les conditions vidéo de référence et les conditions de référence étudiées. Elles permettent également d'indiquer la qualité subjective en termes de mesure objective de la référence. Le fait d'inclure des conditions vidéo de référence dans un essai permet de présenter une large gamme de conditions d'essai tout en offrant une mesure de qualité vidéo qui peut être reproduite de façon plus fiable lors d'un autre essai ou dans un autre laboratoire .

Si l'on considère un système capable de produire un tel ensemble de conditions de référence vidéo réglables, plusieurs éléments doivent être examinés:

1) il faut déterminer le type de dégradations vidéo ou de procédés de compression numérique qu'il y a lieu de simuler;

2) la perception des dégradations vidéo simulées doit être analogue à celle des dégradations réelles;

3) le moyen de produire les dégradations doit être défini précisément, de manière qu'elles puissent être reproduites dans tout laboratoire;

4) l'échelle des niveaux de dégradation doit couvrir une large gamme de qualités d'image.

Les algorithmes de compression vidéo peuvent introduire de nombreux effets de dégradation dans l'image. Cinq des types les plus courants sont les suivants: la distorsion en mosaïque, le flou, la gigue de contour, le sautillement et le bruit. Ces effets de dégradation sont créés pendant la phase de compression et sont généralement dus à la perte d'informations significatives par quantification, codage ou autres techniques de compression. Ils sont généralement associés au contenu cinétique des images.

Un système de référence vidéo réglable doit être assez simple à mettre en oeuvre et doit simuler les cinq dégradations ci-dessus de façon qu'elles soient semblables à celles qui sont produites par les codeurs vidéonumériques et qu'elles soient indépendantes du système de compression. La gamme des dégradations simulées doit couvrir une large étendue de qualités d'image, permettant d'évaluer la qualité vidéo des codeurs vidéonumériques. Cela implique que des essais subjectifs soient effectués afin de déterminer la gamme de qualité d'image couverte par les dégradations simulées.

La présente Recommandation décrit certaines des dégradations qui doivent être produites par un système de dégradation vidéo de référence réglable ainsi que les méthodes générales utilisées pour les produire. L'Appendice I contient une description du système VIRIS, qui est une mise en oeuvre spécifique du système de dégradation de référence pour signaux vidéo, permettant l'évaluation des codeurs vidéonumériques à la norme MPEG-1. On notera que des essais de validation sont requis avant que toutes les Administrations puissent accepter le système VIRIS ou toute autre mise en oeuvre d'un système RISV.

La présente Recommandation reflète l'état d'avancement actuel de la recherche en matière d'évaluations interactives de la qualité audiovisuelle. Au fur et à mesure des progrès accomplis dans ce domaine, il est certain que la compréhension de ces méthodes d'essai interactives ira en s'améliorant. La présente Recommandation sera révisée lorsque de nouvelles connaissances auront été obtenues.

**Recommandation P.930**

# PRINCIPES D'UN SYSTEME DE DEGRADATION DE REFERENCE POUR SIGNAUX VIDEO

*(Genève, 1996)*

## 1 Domaine d'application

La présente recommandation décrit un système de référence vidéo réglable que l'on peut utiliser pour produire les conditions de référence nécessaires afin de caractériser la qualité subjective d'images vidéo issues de systèmes vidéonumériques à compression. Un système de dégradation de référence pour signaux vidéo (RISV, *Reference impairment system for video*) peut être utilisé afin de simuler les dégradations résultant de la compression de séquences vidéo, indépendamment de l'algorithme de compression. Les méthodes d'évaluation subjective ne sont pas traitées dans la présente Recommandation mais sont décrites dans la Recommandation P.910.

## 2 Références

Les Recommandations et autres références suivantes contiennent des dispositions qui, par suite de la référence qui y est faite, constituent des dispositions valables pour la présente Recommandation. Au moment de la publication, les éditions indiquées étaient en vigueur. Toute Recommandation ou autre référence est sujette à révision; tous les utilisateurs de la présente Recommandation sont donc invités à rechercher la possibilité d'appliquer les éditions les plus récentes des Recommandations et autres références indiquées ci-après. Une liste des Recommandations ITU-T en vigueur est publiée régulièrement.

– Recommandation UIT-R BT.500-6 (1994), *Méthodes d'évaluation subjective de la qualité des images de télévision.*

– Recommandation UIT-R BT.601-4 (1994), *Paramètres de codage de télévision numérique pour studios*.

– Recommandation UIT-R BT.802 (1992), *Images et séquences d'essai pour l'évaluation subjective des codecs numériques véhiculant des signaux produits conformément à la Recommandation 601*.

– Recommandation UIT-T P.920 (1996), *Méthodes d'essai interactives pour communications audiovisuelles.*

– Recommandation UIT-T P.910 (1996), *Méthodes subjectives d'évaluation de la qualité vidéographique pour les applications multimédias.*

– Recommandation UIT-T H.261 (1993), *Codec vidéo pour services audiovisuels à* $p \times 64$ *kbit/s.*

## 3 Abréviations et définitions

La présente Recommandation définit les termes suivants et utilise les abréviations suivantes.

**3.1** **pixel**: élément d'image qui décrit la brillance ou la couleur d'un point discret dans une image.

**3.2** **bloc**: groupe de pixels. Par exemple, un bloc de $8 \times 8$ pixels est le plus petit bloc de codage utilisé dans les algorithmes MPEG-1. Une image au format SIF contient 1320 de ces blocs

élémentaires: 44 dans la direction horizontale (352 pixels/8) et 30 dans la direction verticale (240 lignes/8).

**3.3 distorsion en mosaïque; pavage**: distorsion de l'image caractérisée par l'apparition d'une structure sous-jacente de codage par blocs.

**3.4 flou**: distorsion globale de l'image entière, caractérisée par une diminution de la netteté des contours et du détail.

**3.5 erreurs chromatiques**: distorsion de tout ou partie de l'image finale, caractérisée par l'apparition de teintes ou de niveaux de saturation peu naturels ou inattendus, qui n'étaient pas présents dans l'image d'origine.

**3.6 gigue de contour**: distorsion concentrée sur les contours des objets et caractérisée plus en détail par ses coordonnées temporelles et spatiales.

**3.7 blocs erronés**: forme de distorsion en mosaïque dans laquelle un ou plusieurs blocs de l'image n'ont aucune ressemblance avec la scène courante ou précédente et font souvent un net contraste par rapport aux blocs adjacents.

**3.8 sautillement**: mouvement qui était à l'origine coulé mais qui est perçu comme une suite de saccades ("instantanés") distinctes.

**3.9 papillotement**: forme de distorsion par gigue de contour parfois associée à un mouvement, caractérisée par des effets mobiles et/ou par des tachetures parasites qui se superposent aux objets (donnant un peu l'impression d'un moustique voletant autour de la tête et des épaules d'une personne).

**3.10 bruit de quantification**: effet de "neige" ou de "sel et poivre" semblable à un processus de bruit aléatoire mais non uniformément réparti sur l'image.

**3.11 format intermédiaire commun (CIF, *common intermediate format*)**: (352 pixels de luminance × 288 lignes) utilisé par les codeurs H.261.

**3.12 quart de format CIF (QCIF, *quarter CIF*)**: (176 pixels de luminance × 144 lignes).

**3.13 format d'entrée de source (SIF, *source input format*)**: utilisé par les codeurs MPEG comme format progressif non entrelacé (352 pixels de luminance × 480 lignes × 29,97 Hz ou 352 pixels de luminance × 288 lignes × 25 Hz).

**3.14 format CCIR**: norme vidéonumérique de l'UIT-R (ex-CCIR) utilisant des formats entrelacés de 720 pixels de luminance × 480 lignes × 30 Hz ou de 720 pixels de luminance × 576 lignes × 25 Hz.

**3.15 groupe d'experts en images animées (MPEG, *moving pictures experts group*)**: groupe de travail organisé par l'ISO et chargé de mettre au point des normes sur les systèmes multimédias.

## 4 Attributs d'un système de dégradation de référence pour la vidéo (RISV)

Il est recommandé qu'un système RISV soit capable de produire des effets de dégradation (soit isolément ou en combinaisons) dus aux opérations suivantes:

a) conversions entre formats analogiques et numériques (par exemple bruit et flou);

b) codage et compression (par exemple sautillement, gigue de contour et distorsion en mosaïque);

c) erreurs dans la voie de transmission (par exemple blocs erronés).

Du point de vue du visualisateur, les dégradations produites par le système RISV doivent être une bonne approximation des dégradations produites par les systèmes numériques de codage vidéo et de transmission. Cinq de ces dégradations seront décrites plus en détail ci-dessous. L'Appendice I décrit des algorithmes permettant de produire ces types de dégradation.

### 4.1 Gamme réglable de niveaux de dégradation vidéo

La grandeur de la distorsion due à chaque dégradation doit être réglable de manière indépendante. Les étendues de réglage doivent être suffisantes pour couvrir la gamme totale de qualité pour les systèmes numériques de codage vidéo et de transmission. Il doit également être possible de combiner au moins deux dégradations et d'en régler le niveau combiné sur une étendue qui soit utile pour les essais de ces systèmes. Ces deux sujets font l'objet d'une étude permanente et le paragraphe 5.6 donne des directives supplémentaires à ce propos.

### 4.2 Applications possibles d'un système RISV

Trois applications possibles d'un système RISV sont les suivantes:

1) création de conditions de référence dans les essais subjectifs de systèmes vidéonumériques pour faire en sorte que la qualité des scènes présentées aux spectateurs couvre la gamme totale de qualité d'image;
2) définition de niveaux normaux de dégradation vidéo qui pourront être utilisés pour comparer des résultats d'essai subjectif;
3) quantification de la qualité perçue par l'utilisateur d'un système vidéo, par rapport à une référence connue.

La normalisation du système RISV est suffisante pour la première application. Dans un essai subjectif, une condition de référence est constituée par la scène vidéo qui est traitée par le système RISV afin d'y ajouter des dégradations d'un niveau spécifié. Habituellement, plusieurs conditions de référence sont ajoutées à la liste des scènes d'essai subjectif. Les niveaux de dégradation apportés aux conditions de référence sont choisis de façon à couvrir toute la gamme de qualité - depuis EXCELLENT jusqu'à MAUVAIS. Cela permet aux participants à un essai subjectif de porter un jugement plus fiable sur la qualité.

Pour les deuxième et troisième applications, la normalisation du système RISV seul n'est pas suffisante. Pour ces deux applications, les scènes vidéo utilisées avec le système RISV doivent aussi être normalisées parce que le niveau perçu des dégradations vidéonumériques dépend étroitement du contenu de la scène vidéo (c'est-à-dire des informations spatiales et temporelles de la scène).

Avec la deuxième application, l'utilisation d'un système RISV devient nécessaire si plusieurs laboratoires différents effectuent indépendamment des essais subjectifs sur le même système vidéonumérique. Les laboratoires peuvent souhaiter comparer les résultats d'essai, déterminer les différences entre eux et combiner les résultats dans une analyse globale. Si différents laboratoires souhaitent comparer leurs résultats d'essai indépendants pour une analyse globale et calculer les différences entre laboratoires, ceux-ci doivent utiliser les mêmes scènes d'essai et les mêmes conditions de référence. Cela est possible si l'on utilise un système RISV pour produire les niveaux de dégradation associés à un ensemble spécifique de scènes d'essai. En général, le fait de ne rendre compte que des réglages de dégradation du système RISV n'est pas suffisant. La scène vidéo utilisée pour déterminer les réglages des niveaux de dégradation RISV doit également être décrite (cette scène devant faire partie des scènes d'essai normalisées et courantes, comme celles qui se trouvent dans la Recommandation UIT-R BT.802.

La Figure 1 illustre la troisième application potentielle d'un système RISV. Celui-ci peut être utilisé pour quantifier la qualité perçue par l'utilisateur d'une séquence vidéo soumise à un système de

compression. Par comparaison de la qualité du signal vidéo à la sortie d'un système de compression avec celle du signal produit par un système RISV, on peut obtenir une mesure objective de la qualité, en termes de niveau de dégradation RISV. Sur la base de la Figure 1, le fonctionnement du système RISV est le suivant. Une scène vidéo est injectée dans le système vidéo en essai et dans le système RISV. Ensuite, les signaux vidéo issus du système vidéo en essai et du système RISV sont alternativement affichés sur un moniteur. Les niveaux de dégradation RISV sont réglés au moyen d'un ensemble de "boutons" de réglage (qui sont en fait des paramètres logiciels) jusqu'à ce qu'un sujet visualisateur de l'essai juge que les deux qualités vidéo sont égales.

L'utilisation du système RISV de cette manière permet de spécifier objectivement la qualité vidéo, en termes de niveaux de dégradation RISV. Au fur et à mesure du développement de nouvelles réalisations du système RISV, l'algorithme spécifié pour évaluer la qualité vidéo, décrit pour la troisième application, pourra être raffiné. Une telle utilisation d'un système RISV est tout à fait préliminaire et fera l'objet d'un complément d'étude.

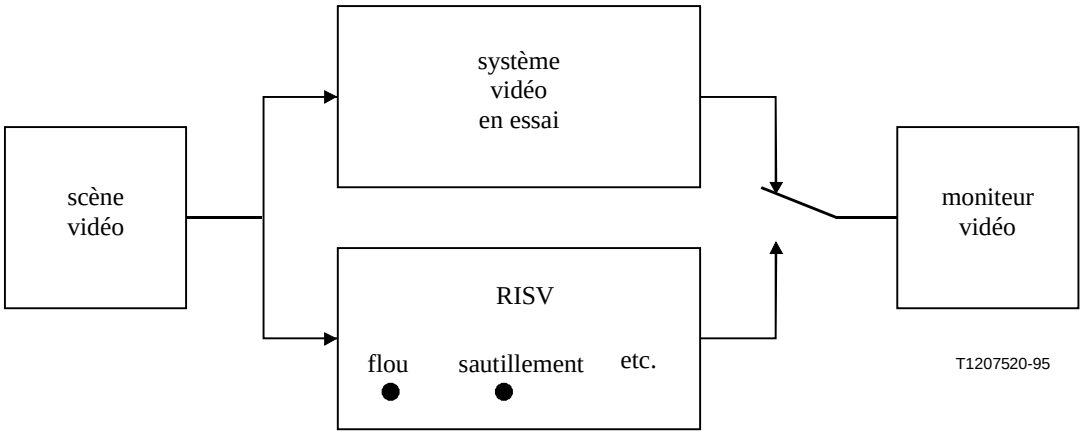


FIGURE 1/P.930

**Exemple de système RISV utilisé pour quantifier la performance d'un système vidéonumérique**

NOTE – Selon la réalisation, le système RISV peut ne pas fonctionner en temps réel car il nécessite une forte puissance de traitement du signal. Les scènes produites devront peut-être être mises en mémoire pour reproduction ultérieure lors du processus d'adaptation.

## 5 Dégradations produites par le système RISV

Il existe de nombreuses dégradations que le système RISV doit simuler. Actuellement, les cinq dégradations qu'un système RISV peut simuler par construction sont les suivantes: distorsion en mosaïque, flou, gigue de contour, bruit et sautillement. D'autres dégradations à simuler, telles que les blocs erronés et la distorsion chromatique, feront l'objet d'une future étude.

### 5.1 Distorsion en mosaïque

La distorsion en mosaïque (ou en blocs) est souvent causée par une quantification grossière des composantes de fréquence spatiale au cours du processus de codage. Pour un niveau de quantification donné, la distorsion en mosaïque est habituellement plus visible dans les zones les plus homogènes de l'image. Un moyen de simuler la distorsion en mosaïque consiste à modifier les valeurs de luminance dans un bloc de pixels de manière à rendre ce bloc visible par rapport aux blocs qui l'entourent.

Dans de nombreux codeurs utilisant la transformation discrète en cosinus (DCT, *discrete cosine transform*), les pixels sont par exemple groupés en blocs de 8 × 8 pixels. Dans le codeur MPEG-1 par exemple, il y a 1320 blocs dans une image de format SIF dérivé du format à 525 lignes CCIR-601, soit 44 blocs dans la direction horizontale (352 pixels/8) et 30 blocs dans la direction verticale (240 lignes/8). Etant donné que la distorsion en mosaïque est surtout perçue dans les zones homogènes à fort contenu cinétique, une simulation peut consister à dégrader un sous-ensemble des blocs trouvés dans ces zones en modifiant la valeur de luminance des blocs sélectionnés, selon un algorithme numérique. La position et le nombre des blocs dégradés, ainsi que l'intensité du changement de luminance déterminent la perceptibilité et le niveau de dégradation.

D'autres méthodes de simulation de la distorsion en mosaïque feront l'objet d'un complément d'étude.

### 5.2 Flou

Le flou est la diminution de netteté des contours et du détail spatial. Il est souvent provoqué, dans les algorithmes de compression, par le compromis consenti sur le nombre de bits pour coder la définition et le mouvement. Les formats d'image tels que SIF, CIF et QCIF ont une définition plus faible que le format d'image CCIR 601 et, comparés à celui-ci, sont effectivement flous même avant la compression.

On peut simuler la dégradation par flou en appliquant à la matrice de signaux vidéonumériques un filtre à réponse impulsionnelle de durée finie (FIR, *finite duration impulse response*) passe-bas, symétrique et à deux dimensions. Plusieurs filtres, ayant différentes fréquences de coupure, peuvent être mis en oeuvre afin de régler le degré de flou. D'autres méthodes de simulation de la dégradation par flou sont en cours d'étude.

### 5.3 Gigue de contour

La gigue de contour est souvent provoquée par l'utilisation de niveaux de quantification trop espacés au cours du processus de codage dans un bloc contenant à la fois des zones homogènes délimitant des contours et des pixels de niveau moyen nettement différent. Le résultat en est l'apparition d'une distorsion concentrée sur les contours des objets, caractérisée par une netteté variant dans le temps ou un bruit variant dans l'espace.

La gigue de contour peut être simulée par adjonction à l'image numérique d'échos périphériques positifs ou négatifs (en avance ou en retard) dans les directions verticale et/ou horizontale afin de créer un halo autour des contours des objets. Les déphasages des échos peuvent être modifiés afin de commander le degré de gigue de contour perçue. On peut provoquer un miroitement du halo en modifiant le déphasage de l'écho toutes les quelques trames.

Les échos affectant l'image numérique sont le résultat d'ondulations de la réponse en amplitude dans la bande passante de la matrice d'image. La fréquence de ces ondulations détermine le déphasage de l'écho; leur amplitude détermine celle de l'écho et leur phase détermine la polarité de l'écho (positive ou négative). Les échos peuvent donc être produits par filtrage de l'image au moyen d'un filtre multibande faisant onduler l'amplitude dans la bande passante.

### 5.4 Bruit

Il existe de nombreux types de dégradations par le bruit couramment produit par les algorithmes de compression. Deux dégradations communes sont le bruit de papillotement et le bruit de quantification. Le bruit de papillotement est une forme de gigue de contour caractérisée par des effets mobiles et/ou par des taches parasites se superposant aux objets. L'algorithme permettant de simuler le bruit de papillotement doit faire l'objet d'un complément d'étude. Le bruit de quantification peut être apporté à l'échelon de quantification au cours du processus de codage de compression. Ce bruit

ressemble à un processus de bruit aléatoire (neige). Il peut être gris ou coloré mais sa répartition n'est pas uniforme sur l'image.

Dans le cas de procédés de codage par blocs, la dégradation par bruit de quantification peut être simulée par remplacement de la valeur de luminance de pixels situés à des endroits aléatoires par une valeur aléatoire conditionnée. L'emplacement aléatoire à dégrader peut être déterminé par sélection de deux nombres aléatoires, le premier indiquant la position du pixel sur la ligne et le second indiquant la position du pixel dans la colonne. Une fois la position d'un pixel déterminée, la valeur de luminance non dégradée du pixel choisi est remplacée par une valeur aléatoire extraite de la gamme spécifiée par les niveaux de luminance le plus faible et le plus élevé (ou gamme étendue) constatés dans le bloc entourant la position du pixel à dégrader.

D'autres positions de pixel sont sélectionnées jusqu'à obtenir le nombre souhaité de pixels dégradés, ce nombre étant une indication du niveau de dégradation par bruit de quantification.

### 5.5 Sautillement

Le sautillement, ou mouvement saccadé, affecte un mouvement originellement coulé et continu, qui est perçu comme une série de saccades ("instantanés") distinctes. Cette dégradation est couramment observée sur des visiophones et d'autres systèmes vidéo à faible débit. Elle est le résultat du saut de trames vidéo afin de diminuer la quantité d'informations vidéo que le système est appelé à transmettre ou à traiter par unité de temps. La fréquence de trame résultante indique le niveau de cette dégradation.

Le sautillement peut être simulé par répétition de trames vidéo. Par exemple, lorsqu'on utilise un système à 30 trames par secondes, on crée une fréquence de 10 trames par seconde en prenant une trame sur trois afin d'éliminer les deux trames suivantes dans une séquence vidéo. D'autres méthodes de simulation du sautillement sont à l'étude afin de mettre en corrélation le contenu cinétique d'une séquence avec la sévérité de la dégradation par sautillement.

### 5.6 Combinaison de dégradations

Un système RISV doit être capable de simuler au moins les cinq types suivants de dégradation: distorsion en mosaïque, flou, gigue de contour, bruit et sautillement. Le système RISV peut aussi simuler d'autres dégradations, comme les erreurs chromatiques, le papillotement et les blocs erronés. Mais ces dégradations doivent faire l'objet d'un complément d'étude. Pour chacune des cinq dégradations définies, le système RISV doit être capable de régler le niveau de la dégradation dans une plage allant de l'imperceptibilité de la dégradation visuelle à un niveau très gênant de dégradation visuelle. La sélection des niveaux maximaux et minimaux de dégradation est habituellement fondée sur une expérimentation. Elle repose sur l'observation visuelle et sur le jugement subjectif. Les dégradations à étudier et le nombre d'échelons de dégradation correspondant à la gamme de qualité souhaitée sont laissés aux soins de l'expérimentateur. Habituellement, le processus consiste à sélectionner expérimentalement le niveau et la plage appropriés de chaque dégradation, en fonction du contenu de la scène et du codeur en essai.

Un système RISV doit être capable d'introduire plus d'une des cinq dégradations dans une scène vidéo. Dans le cas d'une injection de dégradations multiples, l'ordre de leur introduction devient important. Lorsque l'on ajoute deux ou plus de deux dégradations à une scène, le système RISV peut fonctionner de façon séquentielle. Les trames de la scène sont traitées pour la première dégradation choisie puis les trames sont retraitées pour chaque dégradation choisie. Les points suivants doivent être pris en compte lors de la détermination de la séquence des dégradations.

1) Du point de vue du temps de traitement, il est très efficace de supprimer d'abord des trames pour simuler le sautillement, avant les autres processus de dégradation. Il importe également de préparer une séquence de trames identique à celle que le codeur de compression simulé

aura à traiter lors (de la sélection d'un quantificateur et) de la production de distorsion en mosaïque. Dans la plupart des cas, il est souhaitable que le processus de détection des zones en mouvement utilise cette séquence à nombre de trames réduit. Par ailleurs, la durée de traitement pour la détection des contours bénéficiera également du fonctionnement avec moins de trames.

2) Toutes les opérations de détection devront être exécutées sur cette séquence réduite en trames et avant d'appliquer les autres dégradations (conformément au point 1). Par exemple, la détection précise des contours doit être effectuée avant d'ajouter le flou et les zones cinétiques doivent être repérées avant d'ajouter le sautillement.

3) Le flou doit être ajouté avant les autres dégradations localisées, afin d'éviter de modifier le niveau de celles-ci.

4) De nouveau, l'efficacité du traitement conduit à effectuer la répétition de trames en dernier lieu (afin d'effectuer la simulation du sautillement). Cela est également nécessaire pour représenter correctement d'autres dégradations, car celles-ci ne changeront pas lors de la répétition des trames correspondantes.

D'autres considérations sont recherchées dans le cadre de l'étude en cours à ce sujet.

La Figure 2 représente les dégradations dans le même ordre que celui du système VIRIS (réalisation particulière d'un système RISV, voir Appendice I). Chacun des niveaux de dégradation peut être réglé indépendamment dans la gamme appropriée. Sinon, le système RISV peut comporter une commande principale permettant de régler simultanément les niveaux de toutes les dégradations. L'utilisation d'une commande principale n'est cependant pas encore bien cernée et fera donc l'objet d'un complément d'étude.

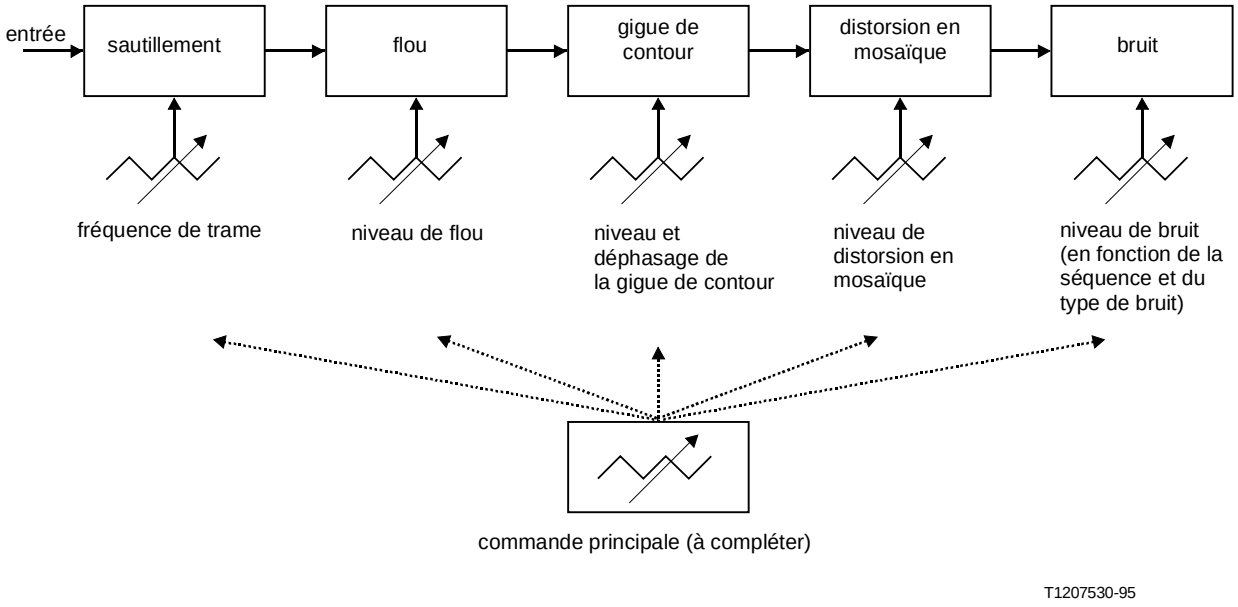


FIGURE 2/P.930

**Schéma du système RISV montrant l'ordre des dégradations et les commandes de niveau**

## 6 Conclusion

La présente Recommandation décrit les principes d'un système réglable de référence vidéo qui peut être utilisé pour produire les conditions de référence nécessaires afin de caractériser la qualité

subjective d'images vidéo issues de systèmes vidéonumériques à compression. L'utilisation de conditions de référence au cours d'un essai subjectif permet d'effectuer des comparaisons entre conditions vidéo de référence et conditions vidéo étudiées, avec donc une mesure plus fiable de la qualité vidéo. La présente Recommandation décrit trois applications d'un système de référence réglable, ainsi que les dégradations qu'un tel système doit produire. Bien que les caractéristiques d'un système RISV y soient décrites, il est nécessaire d'effectuer des essais de validation avant qu'une réalisation puisse être recommandée.

## Appendice I

## VIRIS, mise en oeuvre spécifique d'un système RISV

### I.1 Introduction

L'objet du présent appendice est de rendre compte d'efforts visant à concevoir un système de dégradation vidéo de référence (VIRIS, *video reference impairment system*) qui simule les dégradations par distorsion en mosaïque, flou, gigue de contour, sautillement et bruit. Ce système est destiné à servir d'outil général de laboratoire vidéo pour évaluer les performances de codeurs vidéonumériques. Le système VIRIS (parfois appelé VIRIS1, version plus récente du système de dégradation vidéo de référence) est un progiciel écrit en langage C qui permet de manipuler des fichiers vidéonumériques afin d'introduire dans une image vidéo des dégradations de codage simulées. Il est conçu pour fonctionner sur des images en format SIF mais la méthode peut s'appliquer à d'autres formats d'image comme CIF, QCIF et CCIR 601. Le système VIRIS en est au stade préliminaire et n'est utile que comme outil de laboratoire à usage général.

Le paragraphe I.2 analyse et définit les méthodes utilisées pour simuler les dégradations dans un système VIRIS. Le paragraphe I.3 décrit le calcul de la valeur de crête du rapport signal sur bruit (PSNR, *peak signal-to-noise ratio*). Le paragraphe I.4 décrit les caractéristiques de la plus récente réalisation du système VIRIS. Les paragraphes I.5 et I.6 décrivent:

1) le plan d'essais subjectifs permettant de déterminer les plages de dégradation pour produire des plages assez larges de performance en termes de qualité d'image;

2) les résultats des essais subjectifs effectués pour caractériser la relation entre le niveau de dégradation et la qualité objective. Les conclusions sont données dans le I.7.

### I.2 Simulation des dégradations

Les images d'entrée pour les algorithmes de compression MPEG-1 sont au format SIF, qui est à la fréquence de 30 Hz, progressif et non entrelacé, avec 240 pixels $\times$ 352 pixels de composante *Y* (signal de luminance) et 120 pixels $\times$ 176 pixels des composantes *U* et *V* (signaux de chrominance). Le format SIF est converti à partir du format CCIR 601 à 30 Hz, entrelacé, de 480 $\times$ 720 pixels *Y* et 480 $\times$ 360 pixels *U* et *V*. La résolution du format SIF est la moitié de celle du format CCIR 601 dans les directions verticale et horizontale. Le SIF a été choisi comme format d'entrée pour le système VIRIS parce que celui-ci peut être utilisé pour évaluer des codecs vidéo à 1,5 Mbit/s utilisés dans le cadre de services vidéo à demande par tonalité téléphonique. Les méthodes de simulation de dégradation décrites dans ce paragraphe peuvent cependant s'appliquer à d'autres formats d'image, au seul prix de légères modifications pour tenir compte des différences entre formats d'image.

Le programme en langage C, VIRIS, fonctionne sur une seule trame vidéo au format SIF à la fois. Seuls les pixels de luminance sont manipulés pour créer les dégradations simulées. Les pixels chromatiques ne sont pas modifiés.

Les cinq paragraphes suivants détaillent les méthodes utilisées dans le système VIRIS pour simuler la distorsion en mosaïque, le flou, la gigue de contour, le bruit et le sautillement dans des images vidéo au format SIF.

### I.2.1 Distorsion en mosaïque

La distorsion en mosaïque est souvent causée par une quantification trop grossière au cours du processus de compression, ce qui produit une distorsion ou une perte de composantes à haute fréquence. Le résultat d'une telle distorsion est la visibilité de la structure de codage en mosaïque sous-jacente. Pour un niveau de quantification donné, la distorsion en mosaïque est habituellement plus visible dans les zones homogènes de l'image.

Un bloc de $8 \times 8$ pixels est le plus petit bloc de codage utilisé dans les algorithmes MPEG-1. Il y a 1320 de ces blocs dans une image au format SIF, soit 44 dans la direction horizontale (352 pixels de luminance/8) et 30 dans la direction verticale (240 lignes/8).

Les emplacements des blocs à dégrader sont choisis par repérage des zones de l'image où l'on s'attend à voir apparaître une distorsion en mosaïque. Il s'agit normalement de zones homogènes avec contenu cinétique. Le processus utilisé par le programme VIRIS pour simuler les effets de distorsion en mosaïque est le suivant [1]:

a) les contours contenus dans la trame courante sont repérés par filtrage de la trame à bourrage de zéros[1] au moyen de la technique de Sobel [8, 10]. Etant donné que l'on voit généralement une distorsion en mosaïque dans les zones homogènes de l'image, l'identification des contours de celle-ci permet automatiquement de localiser les zones homogènes dans lesquelles la distorsion en mosaïque due à la compression est visible;

b) la distorsion en mosaïque est introduite dans les zones homogènes de l'image qui ont un contenu cinétique.

La façon dont on détermine le mouvement ou le changement d'information de perception temporelle (*TI*, en abrégé, *temporal perceptual information*) consiste à calculer la différence d'intensité de luminance pour un pixel de coordonnées ($x$,$y$), d'une trame à l'autre.

$$TI(x,y,t+1) = Y(x,y,t+1) - Y(x,y,t) \qquad \text{(I.2-1)}$$

où le terme $Y(x,y,t+1)$ est le pixel de la trame de luminance d'image à bourrage de zéros à l'instant $t+1$ et où le terme $Y(x,y,t)$ est le pixel de la trame de luminance d'image à bourrage de zéros à l'instant $t$ et où le terme $TI(x,y,t+1)$ est l'information de perception temporelle à l'instant $t+1$ au pixel ($x$,$y$). Etant donné que la distorsion en mosaïque est généralement constatée dans les zones à contenu cinétique, on détermine l'emplacement du bloc de pixels à dégrader en accumulant les informations *TI* pour la trame courante et pour la trame antérieure, à chaque bloc de la scène en cours:

$$\sum_{i=-7}^{0} \sum_{J=-7}^{0} abs(TI(8r+i, 8s+j, t+1)) \qquad \text{(I.2-2)}$$

[Si le bloc $(8r+i,8s+j)=(x,y)$ est sur un contour au temps $t$ ou $t+1$, l'information $TI(.)=0$ et les coordonnées $r$=1 à 44, $s$=1 à 30 définissent la position du bloc sous-échantillonné de l'image au format SIF]

Plus grande est cette somme, plus grand est le contenu cinétique du bloc: une recherche est donc effectuée dans toute la trame pour localiser le bloc ($r$,$s$) ayant le plus grand contenu cinétique (ou, ce

---

[1] Le bourrage de zéros consiste à ajouter une bande de zéros de part et d'autre de la trame. Cette bande de zéros aura une largeur supérieure ou égale à ($N$-1)/2 pixels, où $N$ est le nombre de pôles utilisés dans le plus grand filtre du système VIRIS.

qui revient au même, le bloc ayant la plus grande information *TI*). Si ce bloc ne contient par ailleurs pas plus de 5 pixels ayant été classifiés comme des contours[2], ce bloc est dégradé et sa position est éliminée lors des déterminations suivantes d'emplacement de distorsion en mosaïque pour cette trame. En fonction du nombre souhaité de blocs à dégrader, ce processus est répété afin de localiser dans une zone homogène le prochain bloc possédant le plus grand contenu cinétique de la trame.

Une fois que les blocs appropriés à dégrader ont été sélectionnés, on peut simuler la distorsion en mosaïque en modifiant les valeurs de luminance dans les blocs de pixels de façon à rendre ces blocs de pixels visibles par rapport à leurs voisins. Le processus de dégradation est le suivant: on détermine d'abord la valeur moyenne de luminance des 64 pixels du bloc, $P_{ave}$. Chaque valeur de pixel du bloc est ensuite réglée à la somme des deux grandeurs suivantes:

1) la moyenne entre la valeur courante de luminance du pixel et la luminance du bloc, $P_{ave}$;

2) un nombre choisi au hasard entre −2 et 2. En introduisant un blocage dans la séquence d'images, on modifie l'emplacement des blocs toutes les 15 trames afin de maintenir constante la position du bloc dégradé, de manière que l'oeil puisse percevoir la dégradation.

### I.2.2 Flou

Le flou est la diminution de netteté des contours et du détail spatial dans une image. Il est souvent provoqué, dans les algorithmes de compression, par le compromis consenti sur le nombre de bits pour coder la définition et le mouvement. Les formats d'image tels que SIF, CIF et QCIF ont par construction une définition plus faible que le format d'image CCIR 601 et, comparés à celui-ci, paraissent flous même avant la compression. Dans le système VIRIS, on peut simuler la dégradation par flou en appliquant à la matrice de signaux vidéonumériques SIF, dans la direction horizontale, un filtre FIR (à réponse impulsionnelle de durée finie) passe-bas, symétrique, à 15 pôles et à une seule dimension. Six filtres, ayant différentes fréquences de coupure, ont été mis au point au moyen d'un progiciel de construction de filtres [6]. La sélection des fréquences de coupure selon les divers modèles a été fondée sur une procédure empirique (essais et erreurs). Quelques trames d'une image vidéo en format SIF ont été traitées avec différents filtres passe-bas. Puis on a effectué des essais de comparaison non formels entre images filtrées et images non filtrées. Des filtres spécifiques ont alors été sélectionnés de façon à couvrir une large gamme de qualités d'image. Ces filtres ont des fréquences de coupure de 1,5, 1,0, 0,75, 0,5, 0,375 et 0,25 MHz, correspondant respectivement à des niveaux de dégradation par flou de 1 à 6. Les valeurs des coefficients pour les 15 pôles de chaque filtre sont indiquées dans le Tableau I.1.

---

[2] On a déterminé expérimentalement que si la grandeur obtenue par application du filtre de Sobel à des blocs de $3 \times 3$ pixels est supérieure à 500, le pixel est classifié comme appartenant à un contour dans le cadre de la dégradation par distorsion en mosaïque.

TABLEAU I.1/P.930

**Valeurs des coefficients pour les filtres passe-bas**

| | Coefficients des pôles du filtre | | | | | |
|---|---|---|---|---|---|---|
| Pôle du filtre | Niveau de dégradation par flou (fréquences de coupure du filtre passe-bas, MHz) | | | | | |
| | 1 (1,5) | 2 (1,0) | 3 (0,75) | 4 (0,5) | 5 (0,375) | 6 (0,25) |
| [–7], [7] | –1 | –2 | –3 | 0 | 2 | –2 |
| [–6], [6] | 1 | -1 | –3 | –3 | 1 | 2 |
| [–5], [5] | 3 | 1 | –1 | –5 | –4 | 4 |
| [–4], [4] | 6 | 5 | 3 | –3 | -6 | –3 |
| [–3], [3] | 10 | 9 | 8 | 5 | -1 | –9 |
| [–2], [2] | 13 | 14 | 15 | 15 | 13 | 3 |
| [–1], [1] | 15 | 17 | 20 | 24 | 28 | 31 |
| [0] | 16 | 19 | 22 | 28 | 34 | 47 |

Le système VIRIS effectue le filtrage en appliquant le filtre passe-bas à la matrice de signaux de luminance en format SIF, sur chacune des 240 lignes. Le signal de sortie filtré, $y_n$, est donné par la relation suivante:

$$y_n = \frac{1}{S} \sum_{i=-7}^{7} h_i \, x_{n-i} \qquad \text{(I.2-3)}$$

où:

$x_n$ signal d'entrée de la matrice d'image au format SIF

$h_i$ coefficients du filtre (voir le Tableau I.1)

$$S = \sum_{i=-7}^{7} h_i$$

$0 \le n < 352$

Dans l'équation (I.2-3), le signal d'entrée délimitant les contours de la trame verticale, $x_n$, est augmenté de 7 échantillons au début et à la fin de chaque ligne afin de permettre le calcul du signal de sortie rendu flou près des contours de trame. La valeur attribuée à ces échantillons d'extension est égale à celle du pixel de luminance délimitant le contour.

### I.2.3 Gigue de contour

Deux algorithmes différents ont été mis au point pour simuler les dégradations par gigue de contour:

1) la dégradation par gigue de contour n'est appliquée qu'aux bordures verticales des objets;

2) la dégradation par gigue de contour est appliquée aux bordures verticales comme aux bordures horizontales.

L'algorithme correspondant à la simulation d'une dégradation par gigue de contour dans une seule direction de bordure n'est plus recommandé. La plus récente simulation d'une dégradation par gigue de contour par système VIRIS est décrite ci-dessous.

### I.2.3.1 Mise en oeuvre de la gigue de contour par le système VIRIS

La distorsion par gigue de contour est provoquée par un niveau de quantification trop élevé dans un bloc contenant à la fois une zone homogène et quelques pixels de niveau moyen nettement différent. Le système VIRIS simule la gigue de contour en ajoutant des échos périphériques négatifs ou positifs (en avance ou en retard) dans les directions verticale et horizontale afin de créer un halo autour des contours des objets. Ces échos ont par définition des déphasages de 0,375, 0,5 et 0,75 µs et le miroitement du halo est créé par modification du déphasage d'écho toutes les cinq trames.

Etant donné que le système VIRIS a été initialement mis au point pour évaluer la performance de codeurs MPEG-1, la simulation de dégradation par gigue de contour a été conçue de façon à simuler au mieux ce type de dégradation dans une image vidéo MPEG-1. L'observation expérimentale a permis de constater que l'effet de gigue de contour se produisait à la fois sur les côtés des objets vidéo MPEG-1 (c'est-à-dire sur les bordures verticales) et aux extrémités verticales des objets (c'est-à-dire sur les bordures horizontales). L'algorithme de simulation utilise donc un algorithme de filtrage à deux dimensions. Par ailleurs, bien que la norme MPEG-1 introduise un effet de flou, le filtrage de l'image pour simuler l'effet de halo a permis d'éviter d'ajouter un effet de flou supplémentaire à l'image pour simuler l'effet de flou MPEG-1. En effet, les échos utilisés pour simuler la dégradation par gigue de contour ont également tendance à rendre flous les bordures dans l'image.

### I.2.3.2 Méthode de simulation de la gigue de contour par le système VIRIS

Comme on y a fait allusion ci-dessus, le système VIRIS simule la dégradation par gigue de contour en utilisant un filtre introduisant des ondulations dans la courbe d'amplitude de la bande passante afin de créer une dégradation par écho. La fréquence d'ondulation détermine le déphasage de l'écho. L'amplitude d'ondulation détermine l'amplitude de l'écho et la phase d'ondulation détermine la polarité de l'écho (positif ou négatif). Les ondulations dans la réponse en fréquence sont produites par application de filtres multibandes numériques à 13 pôles [6] sur la matrice d'image au format SIF, d'abord dans la direction horizontale et ensuite dans la direction verticale. Le Tableau I.2 montre les valeurs des coefficients pour les trois ensembles de filtres multibandes qui produisent les trois déphasages d'écho. Le déphasage d'écho spécifiquement utilisé sur une trame quelconque est déterminé par les données introduites dans le programme VIRIS par l'utilisateur.

La valeur du coefficient du pôle 0 de chacun des filtres est rapportée à une valeur égale à 175. A l'exception d'une seule paire de coefficients de pôle pour chacun des trois filtres de déphasage d'écho, tous les autres coefficients des pôles de filtres sont 0. La position du pôle de filtre ayant un coefficient de valeur non nulle, par rapport à la position du pôle médian, détermine le déphasage de l'écho. La valeur du coefficient détermine l'amplitude de l'écho. La valeur réelle du coefficient est introduite par l'utilisateur dans le programme VIRIS et peut être comprise entre -1 et -30, l'échelonnement étant en nombres entiers. La même valeur de coefficient est appliquée à chacun des trois filtres.

Dans le système VIRIS, on met en oeuvre le filtrage en appliquant les filtres multibandes numériques à la matrice de luminance de l'image au format SIF, d'abord le long de chacune des 240 lignes puis de chacune des 352 colonnes. Le signal de sortie filtré, $z_n$, est donné dans l'un et l'autre cas par les équations suivantes:

$$z_n = \frac{1}{S} \sum_{i=-6}^{6} h_i \, x_{n-i} \qquad \text{(I.2-4)}$$

où:

$x_n$ signal d'entrée de la matrice d'image au format SIF

$h_i$ coefficients du filtre (voir le Tableau I.2)

$$S = \sum_{i=-6}^{6} h_i$$

$0 \le n < 352$ (horizontal)

$0 \le n < 240$ (vertical)

Dans l'équation (I.2-4), la valeur de luminance du signal d'entrée délimitant les contours verticaux et horizontaux de la trame, $x_n$ pour $n$=0, $n$=352 (lignes) et $n$=0, $n$=240 (colonnes), est augmentée de 6 échantillons au-delà des bordures gauche, droite, supérieure et inférieure de la trame, afin de permettre le calcul du signal de sortie $z_n$.

TABLEAU I.2/P.930

**Valeur des coefficients pour filtres multibandes à 13 pôles**

| | Coefficients des pôles du filtre | | |
|---|---|---|---|
| Pôle du filtre | Déphasage de l'écho (µs) | | |
| | 0,375 | 0,50 | 0,75 |
| [−6], [6] | 0 | 0 | −1 à −30 |
| [−5], [5] | 0 | 0 | 0 |
| [−4], [4] | 0 | −1 à −30 | 0 |
| [−3], [3] | −1 à −30 | 0 | 0 |
| [−2], [2] | 0 | 0 | 0 |
| [−1], [1] | 0 | 0 | 0 |
| [0] | 175 | 175 | 175 |

La méthode utilisée pour réaliser la simulation ci-dessus a été empirique. Elle a mis en oeuvre un processus d'essais et erreurs pour aborder divers moyens d'ajouter des échos puis de comparer les résultats avec des images MPEG-1. Des modifications ont été apportées à la méthode de simulation pour essayer d'améliorer la simulation mais ces modifications n'ont pas pu être faites en temps réel car, pour les obtenir, il a été nécessaire de recompiler le programme VIRIS. On estime cependant qu'une simulation assez fidèle des dégradations MPEG-1 a été obtenue par la méthode ci-dessus. Evidemment, la perceptibilité des dégradations MPEG-1 réelles est corrélée au contenu cinétique de l'image alors que, dans le programme VIRIS, la plupart des dégradations simulées n'ont pas cette corrélation. Le degré auquel les dégradations simulées ressemblent aux dégradations réelles dépend donc, dans une grande mesure, de la séquence vidéo particulière qui est utilisée. De toute façon, une des principales préoccupations lors de la conception d'un système réglable de dégradation vidéo de référence est que celui-ci soit simple à mettre en oeuvre et qu'il soit indépendant de tout procédé de codage. Cela exclut, actuellement, la corrélation du contenu cinétique de l'image avec la dégradation simulée.

### I.2.4 Bruit

Le système VIRIS produit deux types de bruit: le bruit de quantification et le bruit corrélé au signal.

#### I.2.4.1 Bruit de quantification

Le bruit de quantification est une perturbation qui apparaît quelquefois dans le pas de quantification du processus de compression. Ce bruit ressemble à un phénomène parasite aléatoire mais n'est pas uniformément réparti sur l'image. Il est simulé dans le système VIRIS par remplacement de la

valeur de luminance de certains pixels, de position aléatoire, par une valeur aléatoire comprise entre 16 (noir) et 255 (blanc saturé). La position aléatoire est déterminée par extraction de deux nombres aléatoires, le premier étant compris entre 1 et 240 pour la position du pixel dans la ligne et le deuxième étant compris entre 1 et 352 pour la position du pixel dans la colonne. Une fois une position de pixel déterminée, la valeur de luminance de cet élément est remplacée par une valeur aléatoirement choisie entre 16 et 255. Cette position de pixel est ensuite éliminée lors des déterminations suivantes des positions de pixel à bruit aléatoire pour cette trame. Le nombre de valeurs de pixels de bruit pouvant être modifiées dans une trame donnée varie de 1 à 84 480 ($352 \times 240$ pixels). Ce nombre est déterminé par la valeur introduite dans le système VIRIS par l'utilisateur. A chaque trame successive, le germe du générateur de nombres aléatoires est modifié dans le cadre des données introduites dans le système VIRIS, de manière que les emplacements des pixels de bruit changent d'une trame à l'autre.

### I.2.4.2 Bruit corrélé au signal

Le bruit corrélé au signal est un terme utilisé pour décrire l'effet visible résultant de la combinaison des dégradations par gigue de contour et par papillotement, comme on les constate en vidéo comprimée. Le bruit corrélé au signal est une distorsion constatée couramment autour des objets en mouvement. Il est caractérisé par un halo parasite de fine texture qui se superpose à proximité et sur les contours des objets en mouvement. Une simulation de la combinaison de ces dégradations nécessite d'extraire les contours dans une scène, de déterminer s'ils ont un contenu cinétique et enfin d'injecter un bruit de dégradation [1].

On localise les contours dans une scène au moyen de la technique du filtre de Sobel. Si la valeur obtenue par application du filtre de Sobel sur les 9 pixels adjacents est supérieure à 50, la position du pixel est classifiée comme appartenant à un contour dans le cadre de la dégradation par bruit corrélé au signal. Si la position du pixel possède également un contenu cinétique (déterminé par calcul de la différence de luminance à l'emplacement du pixel sur deux trames successives et par vérification que l'information *TI* résultante est supérieure à 2), cette position de pixel est étiquetée comme étant à dégrader.

Pour dégrader la position du pixel, on ajoute un nombre aléatoire compris entre $-\beta$ et $\beta$ à la valeur de luminance du pixel non dégradé.

$$Y(x, y, t) = Y(x, y, t) + random\,(-\beta, +\beta) \qquad \text{(I.2-5)}$$

où $Y(x,y,t)$ est la valeur de luminance de la position $(x,y)$ au temps $t$ du pixel identifié comme appartenant à un contour en mouvement.

Pour chaque position de pixel répondant aux conditions d'appartenance à un contour en mouvement, la valeur de luminance de l'image non dégradée est modifiée d'une quantité comprise entre $-\beta$ et $\beta$. Le niveau de la dégradation, spécifié par $\beta$, est celui de l'intensité de la dégradation. En augmentant $\beta$, on augmente la visibilité de la dégradation. Afin de corréler plus efficacement la dégradation à la variation temporelle, les pixels à dégrader sont en fait l'ensemble des pixels identifiés comme appartenant à des contours dans la trame précédente et dans la trame actuelle.

### I.2.5 Sautillement

Le sautillement est défini comme un mouvement, initialement régulier et continu, qui est perçu comme une série d'instantanés distincts. On l'observe couramment sur les visiophones et sur certains systèmes de téléconférence vidéo à faible débit. Il est le résultat de la répétition de trames vidéo pour réduire la quantité d'informations vidéo qu'un système est appelé à acheminer ou à traiter. Le système VIRIS utilise un facteur de répétition de trame (FRF, *frame repetition factor*) afin de régler le niveau de sautillement visible. Un facteur FRF égal par exemple à 3 se traduira par le fait qu'une trame sur trois remplacera les deux trames suivantes dans la séquence vidéo. La fréquence de trame effective,

EFR, est égale au rapport 30/FRF. Pour un facteur FRF égal à 3 dans cet exemple, la fréquence EFR sera égale à 10 trames par seconde.

## I.3 Calcul de la valeur de crête du rapport signal sur bruit (PSNR)

Actuellement, chacune des dégradations simulées dans le système VIRIS est objectivement caractérisée par le calcul de la valeur de crête du rapport signal sur bruit (PSNR) dans chaque trame traitée et par le calcul du rapport PSNR moyen sur toutes les trames d'une séquence vidéo traitée. Le rapport PSNR non pondéré est une mesure permettant d'évaluer la distorsion de la séquence traitée. Pour chaque trame traitée, *k*, la valeur efficace du bruit, $N_{rms_k}$, est calculée comme suit:

$$N_{rms_k} = \sqrt{\sum_{i=1}^{N} \sum_{j=1}^{M} \frac{\left[U_{ij} - I_{ij}\right]^2}{N \cdot M}} \qquad \text{(I.3-I)}$$

où:

$U_{ij}$ valeur de luminance du pixel non dégradé dans la ligne *i* et la colonne *j*

$I_{ij}$ valeur de luminance du pixel dégradé dans la ligne *i* et la colonne *j*

$N$ 240 pour une image au format SIF

$M$ 352 pour une image au format SIF

$k$ $k^{\text{ième}}$ trame.

Pour calculer le rapport PSNR sur une séquence de *K* trames, on détermine d'abord le bruit moyen par trame, $N_{rms_k}$, selon l'équation (I.3-1). On calcule ensuite le bruit moyen, $N_{\text{rms}}$, sur la séquence de *K* trames selon l'équation (I.3-2) suivante:

$$N_{rms} = \frac{1}{K} \sum_{k=1}^{K} N_{rms_k} \qquad \text{(I.3-2)}$$

Finalement, on calcule le rapport PSNR au moyen de l'équation (I.3-3) ci-après:

$$PSNR = 20 * \log_{10} \frac{S_p}{N_{rms}} \qquad \text{(I.3-3)}$$

où $S_p$ est égal au nombre de niveaux auxquels l'intensité de luminance est quantifiée. Le système traité par le système VIRIS quantifie les pixels de luminance sur 8 éléments binaires, soit $S_p = 255$.

Etant donné que le rapport PSNR non pondéré est, comme décrit ci-dessus, faiblement corrélé avec le système visuel humain, une mesure objective plus précise est parfois requise. Ce point fera l'objet d'un complément d'étude.

## I.4 Programme VIRIS

La Figure I.1 montre le processus de dégradation d'une trame vidéo au format SIF par le système VIRIS à un instant donné. Le système VIRIS injecte des dégradations dans une image SIF convertie à partir d'une image au format CCIR 601 (par le programme *preproc*) enregistrée sur un vidéodisque. Les niveaux VIRIS étant enregistrés dans des fichiers de données, les trames sont traitées et la séquence dégradée est reconvertie au format CCIR 601 (par le programme *postproc*) puis transférée sur un vidéodisque numérique pour visualisation.

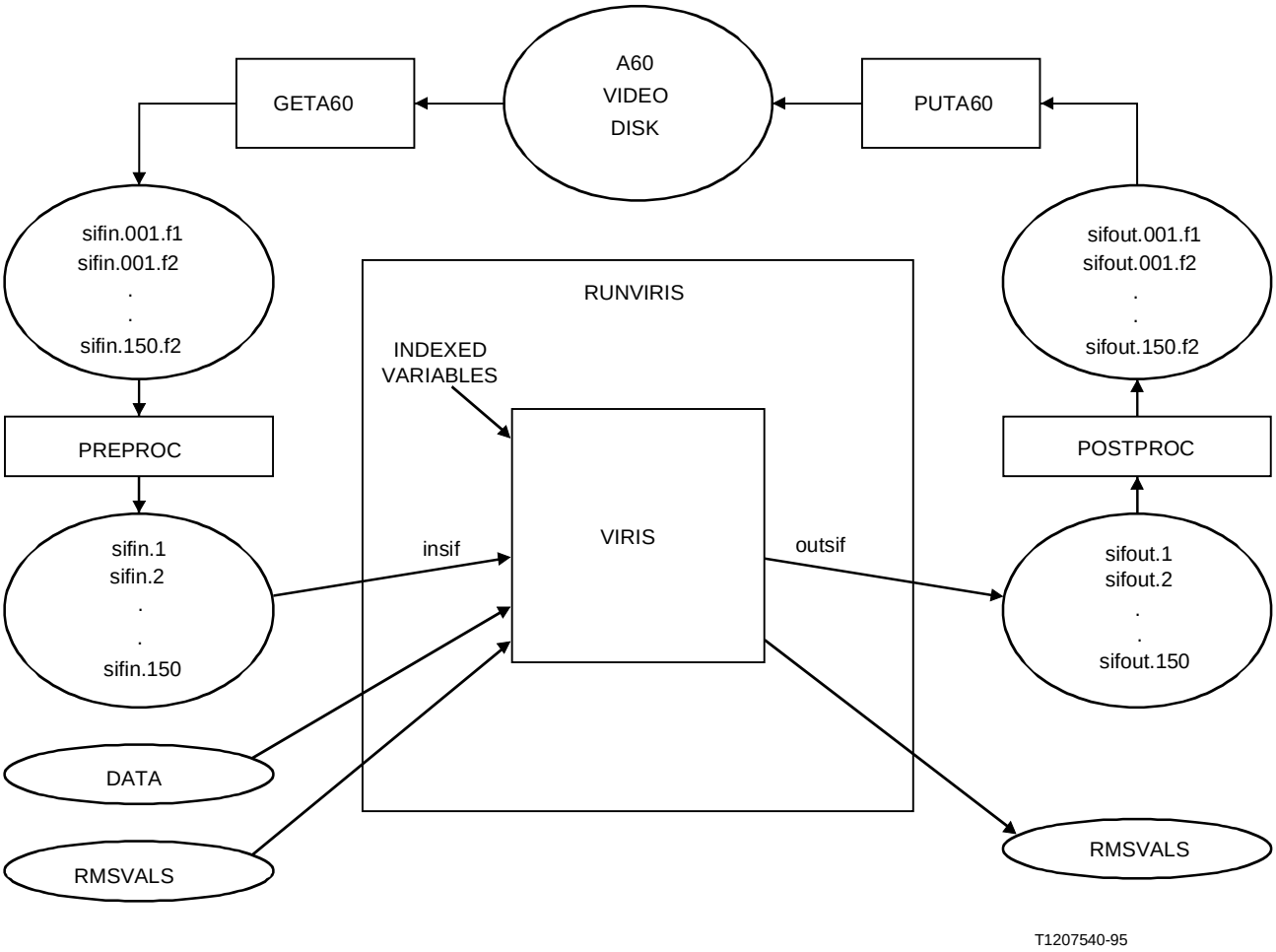


FIGURE I.1/P.930

**Fonctionnement du système VIRIS**

Le fichier de données règle le niveau des dégradations ajoutées à une séquence vidéo. Pour exploiter le système VIRIS, il y a six principaux paramètres de commande de niveau de dégradation:

a) le niveau de distorsion en mosaïque, introduit sous la forme d'un nombre entier et représentant 0,1 % de l'ensemble des 1320 blocs à modifier (le résultat est arrondi au plus proche entier). Par exemple, un niveau de distorsion en mosaïque de 10 se traduit par la dégradation de 13 blocs ($10 \times 0{,}001 \times 1320 = 13{,}2$, arrondi à 13);

b) le niveau de bruit de quantification, introduit sous la forme d'un nombre entier et représentant 0,001 % de l'ensemble des 84 480 valeurs des pixels de luminance (arrondi au plus proche entier). Par exemple, un niveau de bruit de 10 se traduit par la dégradation de la valeur de luminance de 8 pixels ($10 \times 0{,}00001 \times 84\,480 = 8{,}4$, arrondi à 8);

c) le niveau de bruit corrélé au signal, introduit sous la forme d'un nombre entier et représentant l'ensemble des valeurs de luminance qui peuvent être modifiées pour un pixel. Par exemple, un niveau de 10 indique que la valeur de luminance initiale de pixels classifiés comme appartenant à des contours peut être modifiée aléatoirement dans une étendue de –10 à +10 niveaux de luminance;

d) le niveau de flou, introduit sous la forme de nombres entiers compris entre 0 et 6 avec la valeur 0 indiquant l'absence de flou. Les nombres 1 à 6 sélectionnent les filtres passe-bas ayant respectivement des fréquences de coupure de 1,5, 1,0, 0,75, 0,5, 0,375 et 0,25 MHz;

e) le déphasage d'écho pour la gigue de contour, qui est représenté par un nombre entier compris entre 0 et 3, la valeur 0 sélectionnant l'absence de simulation de gigue de contour et les valeurs 1, 2 ou 3 correspondant respectivement à des déphasages de 0,5, 0,75 et 0,375 ms;

f) le niveau d'amplitude de l'écho pour la gigue de x. L'élément de donnée d'entrée est un nombre entier compris entre -30 et -1 qui représente la valeur du coefficient de pôle de filtre pour le déphasage d'écho spécifiquement sélectionné.

Dans la mise en oeuvre du système VIRIS, le niveau de sautillement est spécifié par le facteur de répétition de trame (FRF, *frame repetition factor*), introduit sous la forme d'une variable dans une ligne de commande. Au moyen du fichier de données, VIRIS injecte des dégradations dans le fichier d'image SIF, sur la base des niveaux spécifiés. Après chaque dégradation de trame, le programme calcule le rapport PSNR pour la trame actuelle et le rapport PSNR moyen sur toutes les trames traitées.

## I.5 Plan d'essai subjectif avec le système VIRIS

Des essais subjectifs ont été effectués afin de déterminer la relation entre le niveau de dégradation et l'opinion subjective au sujet de dégradations simulées par VIRIS. En général, les procédures d'essai et les recommandations données dans la Recommandation UIT-R BT.500-6 ont été utilisées comme directives pour concevoir le plan d'essai. La présente Recommandation donne des directives sur les méthodes d'essai, sur les conditions de visualisation et sur les échelles de cotation à utiliser pour évaluer la qualité des images de télévision. La méthode de base utilisée dans le plan d'essai a consisté à faire évaluer, par des observateurs non experts, la qualité de séquences d'images de télévision de 10 s contenant différents niveaux de dégradations VIRIS. Les observateurs disposaient de l'échelle de cotation de qualité à cinq points discrets, allant du jugement EXCELLENT à INSATISFAISANT. Le plan d'essai et les résultats présentés dans les deux paragraphes suivants décrivent les relations entre niveau de dégradation et opinion subjective pour un sous-ensemble des perturbations étudiée en I.2. Plus particulièrement, des essais subjectifs n'ont pas été effectués cette fois pour développer les relations entre niveau de dégradation et opinion au sujet des dégradations par distorsion en mosaïque, bruit corrélé au signal et sautillement, fondées sur les algorithmes décrits dans le cadre du présent appendice. Les informations présentées ci-dessous ont été extraites des articles correspondants de [2].

### I.5.1 Séquences d'images traitées par VIRIS

Trois séquences d'images ont été sélectionnées pour les essais afin de caractériser les dégradations par bruit et par flou. Le nombre spécifiquement choisi a été le résultat d'un compromis entre, d'une part, le souhait de disposer d'un vaste effectif d'échantillons pour représenter différents types d'images (avec l'avantage supplémentaire d'offrir de la variété aux sujets participants à l'essai) et, d'autre part, le temps de traitement requis pour préparer le matériel d'essai. Chacune des séquences avait une durée de 5 s et, après traitement, a été répétée au cours du processus de mise en images afin de créer une séquence de 10 s.

Les séquences ont été sélectionnées afin d'obtenir divers degrés de mouvement et de détail. Leurs titres, assortis d'une brève description, sont les suivants:

a) "bond" – Cette séquence a été extraite d'un disque laser. Elle montre deux personnages, Bond et une femme, montant à deux un cheval dans un désert. Cette séquence comporte peu de détails et le mouvement du cheval au pas est lent;

b) "poursuite" – Cette séquence a été extraite d'un film sur disque laser. Elle montre un jeune homme sur une planche à roulettes qui est poursuivi par plusieurs autres jeunes gens. La séquence comporte beaucoup de détails et de mouvements, avec trois coupures de scène;

c) "football" – Cette séquence a été extraite d'une séquence de 20 s qui avait servi lors d'essais antérieurs pour évaluer des algorithmes MPEG-1. C'est une séquence extraite d'une partie de football, qui contient beaucoup de détails et peu de mouvements.

### I.5.2 Conditions d'essai VIRIS

Pour les dégradations par bruit et par flou, chaque séquence d'images a été traitée 12 fois par le système VIRIS (2 dégradations × 6 niveaux). Les niveaux de dégradation ont été choisis sur la base des résultats d'essais de visualisation informels et d'un essai pilote limité, avec des observateurs non experts. Le Tableau I.3 montre les rapports PSNR calculés par VIRIS pour chacune des 3 séquences d'images ainsi que le rapport PSNR moyen des 3 séquences d'images, pour tous les niveaux de dégradation. Ce tableau montre également les données introduites dans VIRIS (lues dans un fichier de données) pour chacun des 6 niveaux des 2 dégradations.

TABLEAU I.3/P.930

**Rapports PSNR pour niveaux de dégradation VIRIS**

| Niveau de dégradation | Entrées dans VIRIS | Rapport PSNR moyen (dB) | | | |
|---|---|---|---|---|---|
| | | Bond | Poursuite | Football | Moyenne sur 3 pics |
| QN1 | 1 | 60,8 | 60,4 | 59,6 | 60,3 |
| QN2 | 3 | 55,3 | 54,9 | 54,2 | 54,8 |
| QN3 | 7 | 52,2 | 51,9 | 51,1 | 51,7 |
| QN4 | 15 | 48,7 | 48,4 | 47,8 | 48,3 |
| QN5 | 62 | 42,6 | 42,3 | 41,6 | 42,2 |
| QN6 | 125 | 39,4 | 39,1 | 38,5 | 39,0 |
| BLR1 | 1 | 47,2 | 41,8 | 42,7 | 43,9 |
| BLR2 | 2 | 43,4 | 38,1 | 38,1 | 39,9 |
| BLR3 | 3 | 40,0 | 35,4 | 34,2 | 36,5 |
| BLR4 | 4 | 38,6 | 33,7 | 32,6 | 35,0 |
| BLR5 | 5 | 36,3 | 31,8 | 30,1 | 32,7 |
| BLR6 | 6 | 34,2 | 30,2 | 28,0 | 30,8 |
| QN bruit de quantification<br>BLR flou | | | | | |

Le rapport PSNR varie considérablement entre les trois séquences vidéo pour les dégradations par flou et par bruit. La variation de la dégradation par flou est de l'ordre de 1 à 7 dB selon le niveau. La variation du rapport PSNR entre les séquences pour la dégradation par bruit est nettement moins importante (de l'ordre de 1 dB).

Les Figures I.2 et I.3 montrent les courbes du rapport PSNR moyen des 3 séquences d'images, en fonction des données d'entrée dans VIRIS pour deux niveaux de dégradation. Une fonction exponentielle ajustée sur les données est également représentée sur chacune des courbes afin d'offrir un passage entre PSNR et entrée VIRIS pour chacune des dégradations.

Comme indiqué plus haut, les 150 trames de chaque séquence d'images de 5 s traitée par VIRIS ont été converties du format SIF au format CCIR 601 puis transférées sur le disque. Chaque séquence a

ensuite été transférée deux fois de suite sur une bande numérique afin d'obtenir une séquence de 10 s. Comme cet essai subjectif pour caractériser les effets de bruit et de flou caractérisait également les précédentes dégradations par distorsion en mosaïque et par gigue de contour, il y a eu finalement 72 séquences d'image traitées par VIRIS (3 séquences × 4 dégradations × 6 niveaux de dégradation) en plus des 3 séquences d'images de référence au format SIF non traitées. Ces séquences ont été conservées sur bande magnétique. Une fois transféré numériquement sur bande magnétique, tout le matériel d'essai a été enregistré sur vidéodisque laser. Ce système permet un enregistrement NTSC de haute qualité et une reproduction à accès aléatoire au matériel enregistré. Il est piloté par ordinateur afin de reproduire le matériel d'essai dans différents ordres au hasard, comme prescrit par le protocole d'essai.

### I.5.3 Méthode d'essai VIRIS

Pour ces essais, on a utilisé une méthode d'essai à stimulus unique, avec une échelle qualitative de 5 degrés distincts. Les observateurs ont évalué chacune des conditions d'essai en cochant la catégorie de jugement appropriée sur une échelle de commentaires intitulés EXCELLENT, BON, SATISFAISANT, MEDIOCRE et INSATISFAISANT. Chaque condition d'essai faisait l'objet d'une feuille de vote.

Un total de 23 observateurs non experts, choisis dans un volant de sujets extérieurs au laboratoire, ont participé aux essais. Il y avait trois hommes et vingt femmes, d'âge compris entre 25 et 67, la moyenne étant de 52,9 ans. Chaque observateur avait une acuité visuelle normale (égale ou supérieure à 8 dixièmes) et une vision chromatique normale.

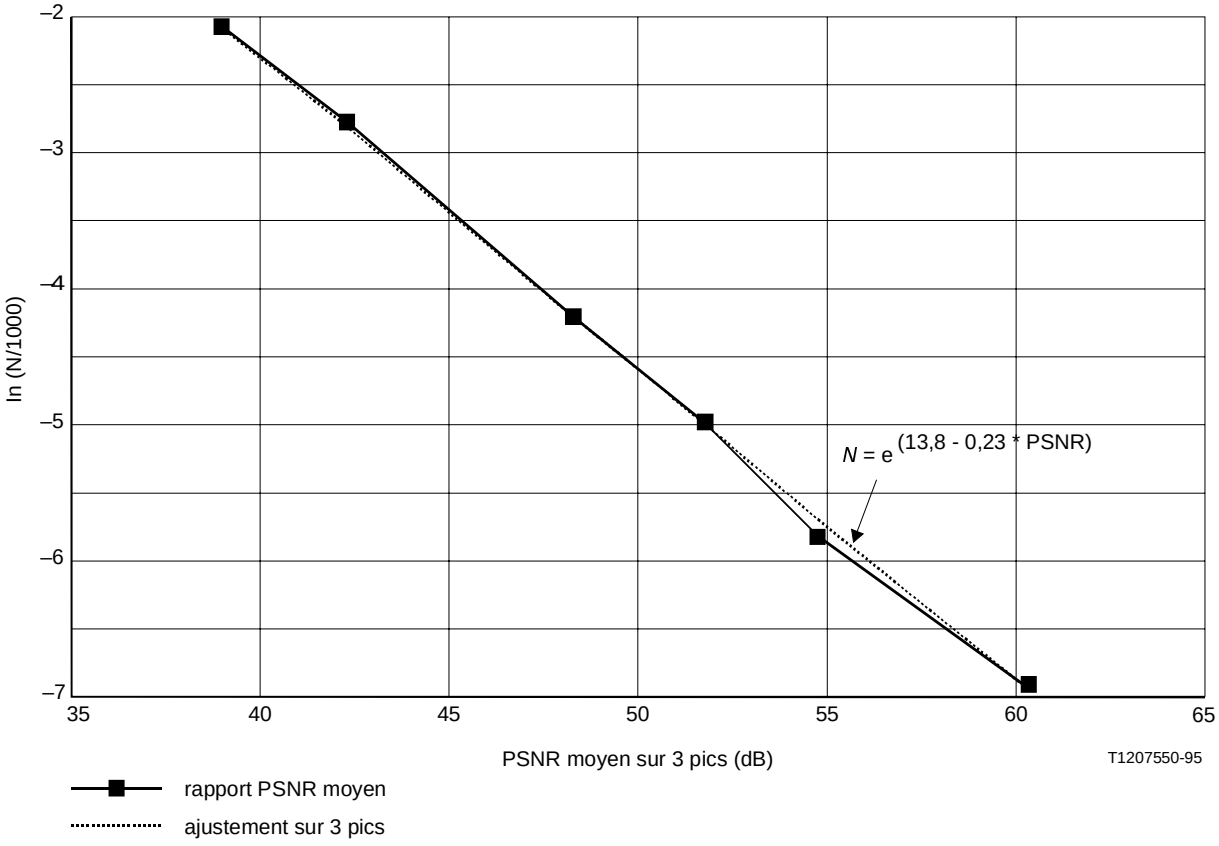


FIGURE I.2/P.930

**Rapport PSNR du bruit de quantification en fonction des données de bruit de quantification introduites dans VIRIS**

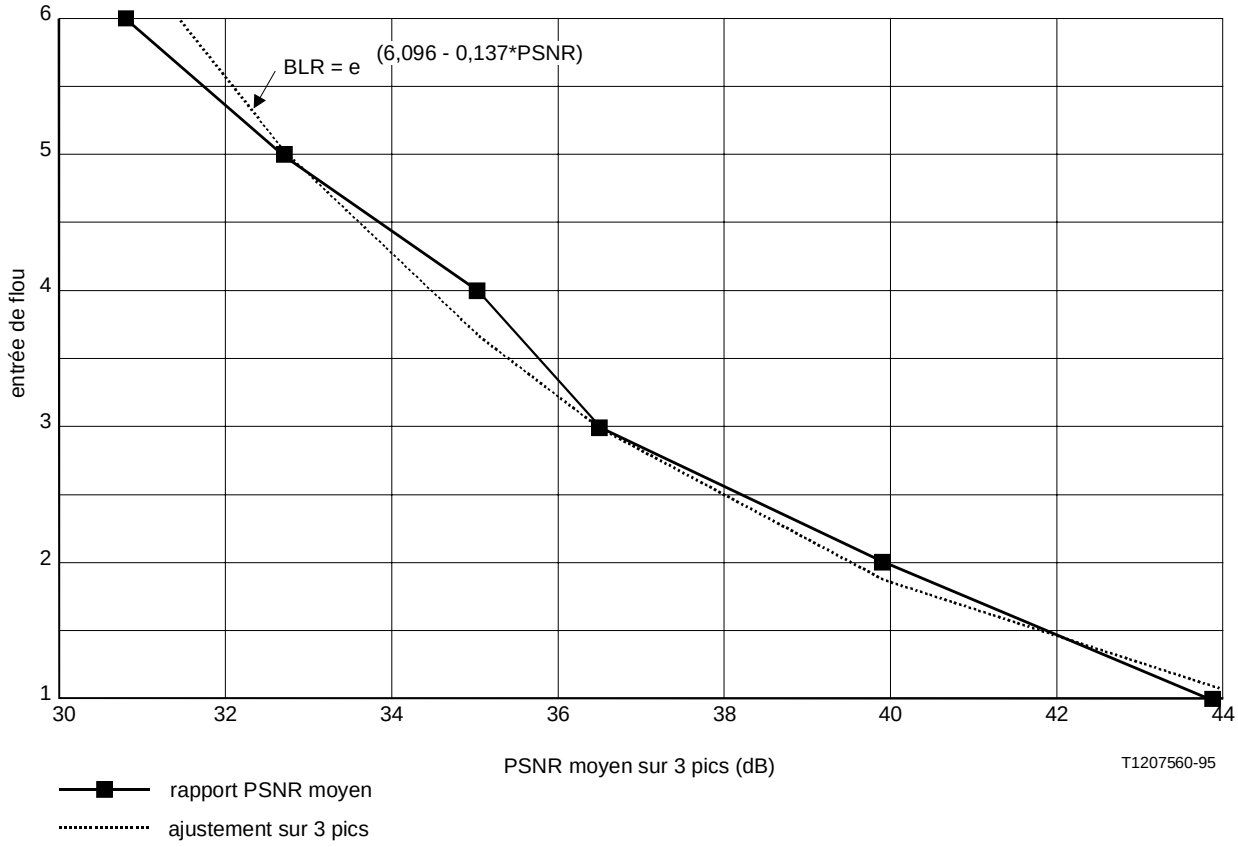


FIGURE I.3/P.930

**Rapport PSNR du flou en fonction des données de flou introduites dans VIRIS**

### I.5.4 Conditions de visualisation VIRIS

Les observateurs qui participaient à l'essai regardaient un poste de télévision. Ils étaient assis à une distance de 7,5 fois la hauteur (7,5 *H*) de l'image (soit environ 2,3 m de l'écran). Cette distance est à mi-chemin entre les deux distances de visualisation qui avaient été utilisées lors d'essais précédents pour évaluer des algorithmes à débit comprimé. Une distance de visualisation de 5*H* est au milieu de l'étendue recommandée par l'UIT-R et une distance de 10*H* est à peu près la distance de visualisation moyenne recommandée aux Etats-Unis d'Amérique. La distance de 7,5*H* utilisée dans l'étude a été un compromis entre une distance (5*H*) que l'auteur jugeait comme trop rapprochée pour une observation confortable lors des essais actuels et le souhait d'obtenir des jugements plus critiques que ceux qui pourraient être obtenus à la distance 10*H*.

La salle d'essai était d'environ 3,66 m × 6,1 m avec des murs gris clair absorbant les pressions acoustiques ainsi qu'un éclairage réglable par plafonnier à lampes incandescentes. L'éclairement ambiant et la luminance de l'écran ont été mesurés avec un photomètre numérique. L'éclairement ambiant de la salle d'essai a été atténué jusqu'à environ 6,5 lx aux positions où les sujets étaient assis, par réglage des gradateurs de plafonnier. L'éclairement ambiant mesuré au niveau du récepteur était d'environ 1,1 lx. Les commandes du récepteur ont été réglées de façon à donner un éclairement lumineux de crête d'environ 24,8 millilamberts (248 lx) et un éclairement lumineux minimal d'environ 0,16 millilamberts (1,6 lx) pour un contraste d'environ 150. Des signaux d'essai avec salves de noir et 100 unités IRE issus d'un générateur numérique ont été utilisés pour faire ces réglages.

### I.5.5 Procédures d'essai VIRIS

Un total de 81 conditions d'essai de base ont été utilisées au cours de l'essai subjectif. Ces conditions se composaient de 72 séquences au format SIF (3 séquences × 4 dégradations × 6 niveaux de dégradation) traitées par le système VIRIS (les anciens algorithmes de distorsion en mosaïque et de gigue de contour, ainsi que les actuels algorithmes de flou et de bruit de quantification) avec 3 répétitions de chacune des séquences vidéo SIF non traitées. Chaque condition durait 20 s et comportait un message de 2 s sur écran pour indiquer le numéro de la condition, la séquence vidéo de 10 s et une période de vote de 8 s avec message à l'écran pour demander à l'observateur de voter au sujet de la qualité de l'image. Chacune de ces 81 conditions d'essai a été essayée deux fois, soit un total de 162 conditions.

Trois observateurs ont été contrôlés à la fois, sauf pour un seul essai où deux observateurs seulement ont été contrôlés (un des sujets convoqués ne s'était pas présenté). Ils étaient assis en face du récepteur, à une distance de visualisation de 7,5H. Ils ont inscrit leur nom et le numéro de leur siège sur la première page d'un livret contenant une page de vote pour chaque effet et chaque condition d'essai. L'administrateur de l'essai a ensuite donné lecture des instructions d'essai. Après avoir répondu aux questions, une session pratique de 11 conditions a été menée pour familiariser les sujets avec les procédures d'essai et avec le matériel d'essai.

L'essai a été conduit en deux parties séparées par une pause de rafraîchissement de 20 minutes ainsi qu'une brève pause d'étirement de 2 à 3 minutes toutes les 40 séquences de chaque partie. Chaque partie comprenait les 81 conditions d'essai de base et durait environ 30 minutes. Les conditions de chaque partie étaient présentées dans un ordre pseudo-aléatoire, avec les deux contraintes suivantes:

a) le premier, le deuxième et le troisième groupe de 27 conditions contenaient chacun tous les niveaux de dégradation des quatre types de dégradation, y compris les trois images SIF de référence;

b) chacun des trois groupes de 27 conditions contenait un nombre égal d'images des trois séquences vidéo.

Ces contraintes ont été conçues pour compenser d'éventuels effets d'apprentissage pouvant affecter les résultats, comme une soudaine prise de conscience d'un aspect particulier d'une dégradation alors que l'essai est déjà bien avancé ou, par exemple, un changement progressif dans les critères utilisés par le sujet pour juger de la qualité, au fur et à mesure de la progression de l'essai. Les ordres aléatoires étaient différents dans chaque moitié de l'essai et étaient changés pour chaque groupe de sujets.

### I.5.6 Résultats des essais subjectifs VIRIS

#### I.5.6.1 Analyse des données

L'analyse des données a consisté à calculer la note moyenne d'opinion (MOS, *mean opinion score*) et les intervalles de confiance à 95 % pour chacun des niveaux de dégradation et pour les images SIF de référence, pour chacune des trois séquences vidéo. Les opinions exprimées par vote des sujets dans les termes suivants: EXCELLENT, BON, SATISFAISANT, MEDIOCRE et NON SATISFAISANT ont d'abord reçu, respectivement, les numéros 5, 4, 3, 2 et 1. La note MOS a été calculée pour chaque condition comme suit:

$$MOS = \frac{1}{N}\sum_{i=1}^{N} V_i \qquad \text{(I.5-1)}$$

où:

$V_i$ vote

$N$ nombre de votes

Le nombre total de votes sur les conditions de dégradation a été de 46 (23 sujets × 2 séries) et de 138 pour les conditions de référence SIF (23 sujets × 6 séries). Les intervalles de confiance à 95 % (CI) sur les notes MOS ont été calculés à partir des 23 moyennes des sujets pour chaque niveau de dégradation et chaque référence SIF, au moyen de l'équation (I.5-2) [4]:

$$95\%\ CI = \mu \pm \frac{t \cdot \sigma}{\sqrt{N}} \qquad \text{(I.5-2)}$$

où:

95% CI intervalle de confiance à 95 %

$\mu$ note MOS

$t$ 2,069 (valeur au test bilatéral du t de Student pour un niveau de signification de 0,05)

$\sigma$ écart type des moyennes des sujets

$N$ effectif des moyennes des sujets (23)

### I.5.6.2 Résultats de la simulation VIRIS

Les résultats, avec les rapports PSNR, les notes MOS et les intervalles de confiance à 95 %, sont représentés dans le Tableau I.4 pour tous les niveaux de dégradation et références SIF dans chacune des trois séquences d'images.

Seuls sont présentés les résultats obtenus par les algorithmes de simulation de dégradation décrits en I.2. L'essai subjectif réel caractérisait un ensemble de quatre dégradations (distorsion en mosaïque, gigue de contour, flou et bruit de quantification); mais les algorithmes permettant de simuler les deux premières dégradations ont été modifiés et les résultats obtenus par cet essai subjectif ne sont donc pas applicables aux algorithmes de simulation actuellement utilisés. Le paragraphe décrit un deuxième essai subjectif permettant de caractériser la dégradation par gigue de contour au moyen de la technique de simulation décrite en I.2.

Pour les trois images de référence (Bond, Poursuite et Football), les notes MOS ont été respectivement de 4,0, 3,8 et 4,2. Ces notes d'opinion sont un peu faibles pour les conditions de référence mais non aberrantes car le format d'image SIF limite la qualité d'image en résolution horizontale et verticale. Lors d'un essai précédent, utilisant une échelle de qualité continue, l'évaluation moyenne de six séquences d'images SIF a été d'environ 60, la plage 60-80 couvrant la région des opinions qualitatives de niveau BON. Les résultats présents avec des séquences d'images SIF ne sont donc pas incompatibles avec les résultats déjà obtenus.

La largeur des intervalles de confiance à 95 % variait de 0,3 à 0,7. En général, les intervalles de confiance devenaient plus étroits lorsque la qualité des images diminuait, ce qui indique que les sujets étaient plus en accord lorsqu'ils jugeaient des images de qualité médiocre que lorsqu'ils jugeaient des images de bonne qualité. Normalement, lorsque la dégradation couvre la plage complète des commentaires, les largeurs des intervalles de confiance seront plus étroits aux extrémités de la plage de commentaires et plus larges pour les qualités intermédiaires.

TABLEAU I.4/P.930

**Résultats d'essai subjectif pour dégradations VIRIS**

| Dégradation | Séquence d'images | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Bond | | | Poursuite | | | Football | | |
| Type/ Niveau | PSNR (dB) | MOS | Largeur *X* de CI | PSNR (dB) | MOS | Largeur *X* de CI | PSNR (dB) | MOS | Largeur *X* de CI |
| Images SIF de référence | - | 4,0 | 0,7 | - | 3,8 | 0,6 | - | 4,2 | 0,7 |
| QN1 | 60,8 | 3,9 | 0,6 | 60,4 | 3,8 | 0,6 | 59,6 | 3,9 | 0,6 |
| QN2 | 55,3 | 3,4 | 0,6 | 54,9 | 3,6 | 0,6 | 54,2 | 3,3 | 0,6 |
| QN3 | 52,2 | 2,7 | 0,5 | 51,9 | 3,4 | 0,6 | 51,1 | 3,3 | 0,6 |
| QN4 | 48,7 | 2,5 | 0,4 | 48,4 | 3,0 | 0,5 | 47,8 | 2,7 | 0,5 |
| QN5 | 42,6 | 1,8 | 0,4 | 42,3 | 1,8 | 0,4 | 41,6 | 2,2 | 0,4 |
| QN6 | 39,4 | 1,8 | 0,4 | 39,1 | 1,7 | 0,3 | 38,5 | 2,0 | 0,4 |
| BLR1 | 47,2 | 3,8 | 0,7 | 41,8 | 3,3 | 0,6 | 42,7 | 3,7 | 0,6 |
| BLR2 | 43,4 | 3,4 | 0,6 | 38,1 | 2,7 | 0,5 | 38,1 | 3,0 | 0,5 |
| BLR3 | 40,0 | 3,0 | 0,5 | 35,4 | 2,0 | 0,4 | 34,2 | 2,6 | 0,5 |
| BLR4 | 38,6 | 2,3 | 0,5 | 33,7 | 1,7 | 0,4 | 32,6 | 2,1 | 0,4 |
| BLR5 | 36,3 | 1,9 | 0,4 | 31,8 | 1,5 | 0,3 | 30,1 | 1,3 | 0,3 |
| BLR6 | 34,2 | 1,6 | 0,3 | 30,2 | 1,3 | 0,3 | 28,0 | 1,3 | 0,3 |

PSNR valeur de crête du rapport signal sur bruit
MOS note moyenne d'opinion
CI intervalle de confiance à 95 %
QN bruit de quantification
BLR flou

Les courbes des notes MOS en fonction des rapports PSNR sont représentées sur les Figures I.4 et I.5 pour les trois séquences d'images, respectivement dans le cas des dégradations par bruit et par flou. Une fonction logistique ajustée sur les données à 3 images de chacune des dégradations est également représentée sur chacune des figures. La fonction logistique est de la forme indiquée dans l'équation (I.5-3) ci-après.

$$MOS = K_1 + \frac{K_2}{1 + e^{-K_3(PSNR - K_4)}} \qquad \text{(I.5-3)}$$

où:

$K_1$ asymptote inférieure des notes MOS = 1,0

$K_2$ asymptote supérieure des notes MOS - $K_1$ = 3,2

$K_3$ paramètre de pente

$K_4$ valeur du rapport PSNR lorsque MOS = $K_1 + \frac{K_2}{2}$

Pour ajuster la fonction logistique, l'asymptote inférieure, $K_1$, a été arrêtée à 1,0 et l'asymptote supérieure, $K_1 + K_2$, a été fixée à 4,2 dans tous les cas. La valeur de l'asymptote inférieure, (1,0) correspond à la note MOS la plus basse qui peut être donnée avec l'échelle de commentaire à 5 degrés. La valeur asymptotique supérieure (4,2) représente la note MOS moyenne la plus élevée qui a été obtenue dans la présente étude pour l'une des conditions de référence SIF (séquence de football). Les deux paramètres d'ajustement étaient donc $K_3$, paramètre de pente, et $K_4$, valeur du rapport PSNR lorsque la note MOS est sur la médiane entre les valeurs asymptotiques inférieure et supérieure.

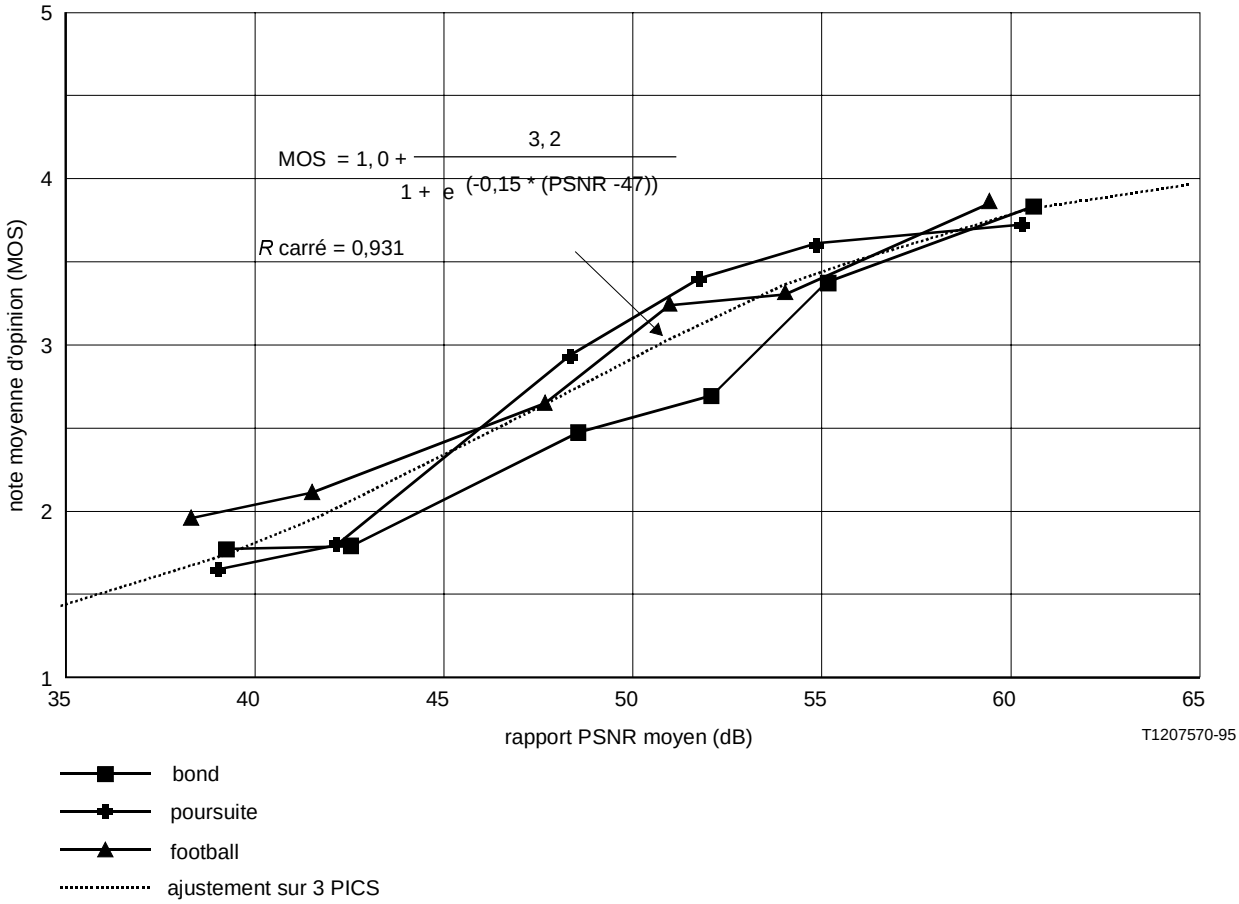


FIGURE I.4/P.930

**Rapport PSNR en fonction des notes MOS pour le bruit de quantification**

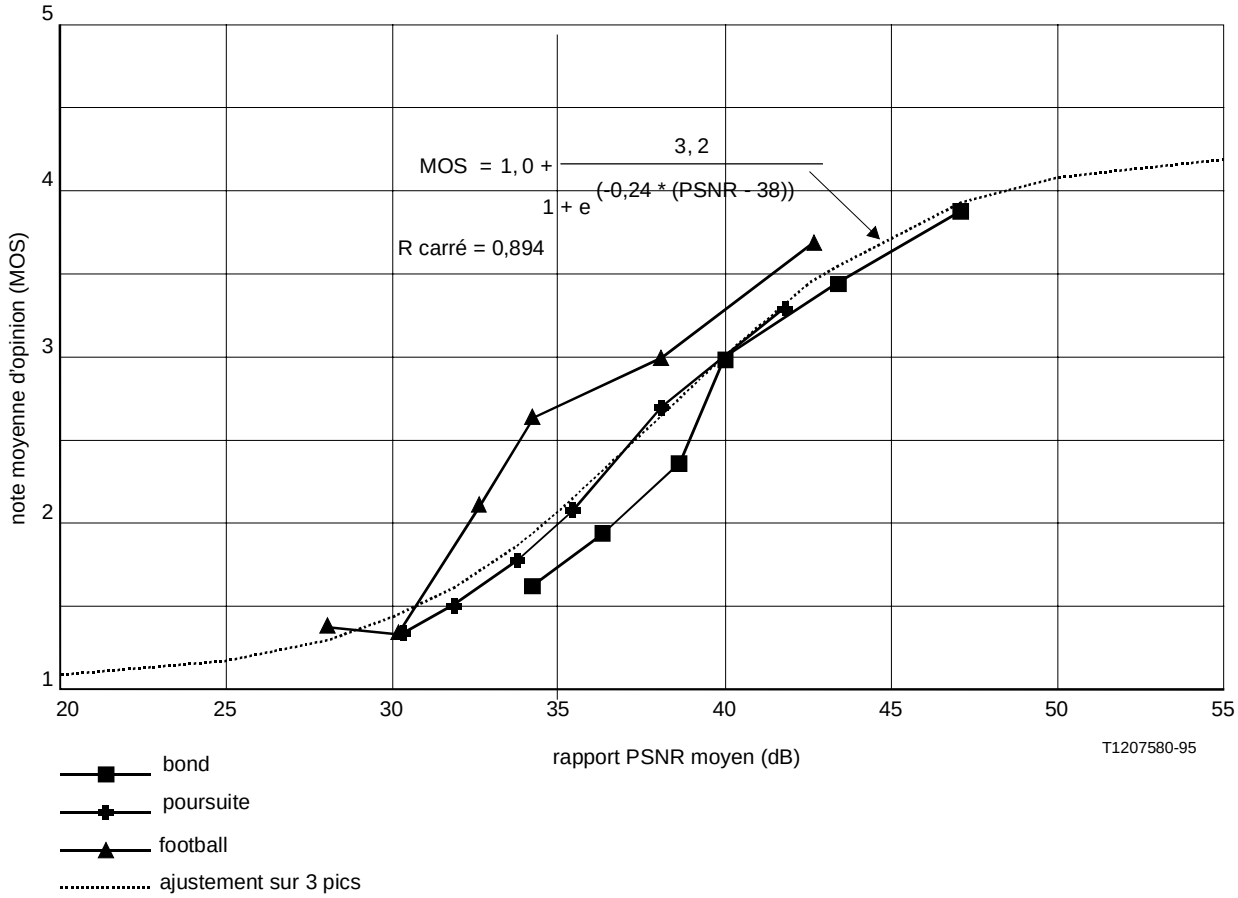


FIGURE I.5/P.930

**Rapport PSNR en fonction des notes MOS pour le flou**

Concernant le bruit de quantification, les résultats indiqués sur la Figure I.4 (réaction subjective aux rapports PSNR de bruit de quantification) ont été un peu différents selon l'image particulière qui était visualisée. L'étendue des notes MOS pour les trois images est cependant assez compacte et la fonction logistique semble être un bon ajustement sur les données à 3 images avec *R* carré = 0,9309 (*R* carré est une statistique de validité de l'ajustement qui va de 0 à 1, la valeur 1 représentant un ajustement optimal).

La Figure I.5 indique que la réaction subjective aux rapports PSNR de flou a été plus sensible à l'image que dans le cas des rapports PSNR de bruit et de distorsion en mosaïque, avec une dispersion un peu plus grande entre les notes MOS des 3 séquences d'images. La fonction logistique est encore un bon ajustement sur les données des 3 séquences d'images avec un *R* carré = 0,8941.

## I.6 Plan d’essai subjectif de gigue de contour

Un deuxième essai subjectif a été effectué afin de déterminer la relation entre niveau de dégradation et opinion subjective pour la dégradation par gigue de contour, comme décrit en I.2. D'autres conditions ont aussi été testées, dont une condition de référence SIF, une condition de référence NTSC et deux conditions MPEG-1, l'une à 1,3 Mbit/s et l'autre à 1,0 Mbit/s.

En général, les procédures d'essai et les recommandations données dans la Recommandation UIT-R BT.500-6 ont été utilisées comme directives pour concevoir le plan d'essai. La présente Recommandation donne des directives relatives aux méthodes d'essai, aux conditions de visualisation et aux échelles de cotation à utiliser pour évaluer la qualité d'image de télévision. La

méthode fondamentale utilisée dans le plan d'essai a consisté à faire juger, par des observateurs non experts, la qualité de séquences d'images de télévision de 10 s, représentant divers niveaux de dégradation VIRIS, les conditions de référence SIF et NTSC et les conditions MPEG-1. Ces essais faisaient appel à l'échelle de cotation de qualité à 5 degrés discrets, allant du qualificatif EXCELLENT à INSATISFAISANT. Les paragraphes suivants détaillent ce plan d'essai.

### I.6.1 Séquences d'images avec gigue de contour

Trois séquences d'images ont été sélectionnées pour les essais. Le nombre spécifiquement choisi a été le résultat d'un compromis entre d'une part le souhait de disposer d'un vaste effectif d'échantillons pour représenter différents types d'images (avec l'avantage supplémentaire d'offrir de la variété aux sujets participants à l'essai) et d'autre part le temps de traitement requis pour préparer le matériel d'essai. Chacune des séquences avait une durée de 5 s et, après traitement, a été répétée au cours du processus de mise en images afin de créer une séquence de 10 s.

Les séquences ont été sélectionnées afin d'obtenir divers degrés de mouvement et de détail. Leurs titres, assortis d'une brève description, sont les suivants:

"course d'autos" – Cette séquence a été extraite d'une séquence de 20 s qui montre une course de voitures sur un arrière-plan fixe de piste. La séquence comporte quatre coupures de scène pour passer à des voitures différentes. Elle donne certains détails (surtout en raison des numéros inscrits sur les voitures de course) et un mouvement modéré. Elle est issue d'un "clip" de TVHD qui a été téléchargé en format NTSC;

"tennis de table" – Cette séquence a été extraite d'une séquence de 20 s montrant deux hommes jouant au ping-pong. Elle comporte peu de détails et de mouvement. Il y a trois coupures de scène avec arrière-plan fixe dans deux des scènes et avec arrière-plan en mouvement lent dans l'autre scène. Cette séquence provient d'une bande numérique de magnétoscope D1 et était l'une des séquences "normales" qui ont été utilisées pour évaluer la performance de l'algorithme MPEG;

"football" – Cette séquence a été utilisée dans le précédent essai VIRIS. Elle a été prise lors d'une partie de football. Elle contient beaucoup de détails et un mouvement modéré. Elle ne comporte pas de coupures de scène et son arrière-plan se déplace lentement. Elle provient d'un "clip" de démonstration de TVHD qui a été téléchargé en format NTSC.

### I.6.2 Conditions d'essai VIRIS et préparation du matériel d'essai

Sur la base d'essais de visualisation officieux en laboratoire, huit niveaux de gigue de contour ont été sélectionnés comme conditions d'essai. Le Tableau I.5 montre les rapports PSNR calculés par VIRIS pour chaque niveau de dégradation par gigue de contour, d'abord sous la forme d'une moyenne pour chaque séquence d'images puis sous la forme d'une moyenne pour les trois images.

TABLEAU I.5/P.930

**Rapports PSNR pour niveaux de dégradation par gigue de contour VIRIS**

| Niveau de gigue de contour (EB) | Rapport PSNR moyen (dB) | | | |
|---|---|---|---|---|
| | Course d'autos | Football | Tennis de table | Moyenne sur 3 pics |
| −1 | 42,9 | 43,5 | 41,1 | 42,5 |
| −3 | 41,8 | 40,7 | 39,5 | 40,7 |
| −5 | 40,3 | 37,8 | 37,3 | 38,5 |
| −10 | 36,5 | 32,1 | 32,6 | 33,7 |
| −15 | 33,2 | 28,2 | 29,1 | 30,2 |
| −20 | 30,3 | 25,1 | 26,4 | 27,3 |
| −25 | 27,7 | 22,7 | 24,1 | 24,8 |
| −30 | 25,3 | 20,5 | 22,1 | 22,6 |
| NOTE – Le niveau de gigue de contour (EB, *edge busyness*) correspond aux valeurs des coefficients de pôle des trois filtres multibandes à gamme variable de coefficients indiqués dans le Tableau I.2. | | | | |

Les rapports PSNR des séquences de football et de tennis de table étaient assez proches, ne différant que de 0,5 à 2,4 dB sur l'étendue des niveaux de gigue de contour. Le rapport PSNR de la séquence de course automobile différait des deux autres séquences de 0,6 à 5,2 dB sur l'étendue des niveaux de gigue de contour. La Figure I.6 montre une courbe du rapport PSNR moyen sur 3 images en fonction du niveau d'entrée de gigue de contour dans VIRIS. La courbe montre également une fonction linéaire d'ajustement aux données afin d'offrir un passage entre rapport PSNR et niveau d'entrée de gigue de contour dans VIRIS.

En plus de 8 huit conditions de dégradation par gigue de contour, l'essai comportait des conditions de qualité NTSC, de qualité SIF et deux conditions de format MPEG-1, l'une à 1,3 Mbit/s et l'autre à 1,0 Mbit/s, soit un total de 12 conditions d'essai. Chacune des 3 séquences d'images de la séquence d'essai a été soumise à chacune des 12 conditions d'essai, soit un total de 36 séquences d'essai. Au cours du traitement et de la mise en image, les séquences étaient enregistrées sur bande numérique D1 puis transférées sur vidéodisque laser. L'enregistreur sur vidéodisque laser permettait d'effectuer des enregistrements NTSC de haute qualité avec reproduction par accès aléatoire au matériel enregistré. L'appareil était commandé par un programme informatique de façon à reproduire le matériel d'essai dans différents ordres aléatoires, selon les exigences du protocole d'essai.

### I.6.3 Méthode d'essai subjectif par gigue de contour

Une méthode d'essai à stimulus unique et échelle quantitative à 5 degrés distincts a été utilisée pour les essais. Un total de 27 observateurs non experts, sélectionnés dans un volant de sujets extérieurs au laboratoire, ont participé aux essais. Il y avait 3 hommes et 24 femmes, d'âge compris entre 19 et 71 ans, la moyenne étant de 47,3 ans. Les observateurs ont évalué chacune des conditions d'essai en cochant la catégorie de jugement appropriée sur une échelle de commentaires intitulés EXCELLENT, BON, SATISFAISANT, MEDIOCRE et INSATISFAISANT. Chaque condition d'essai faisait l'objet d'une feuille de vote.

### I.6.4 Méthode d'essai subjectif pour la gigue de contour

Les conditions de visualisation étaient celles qui avaient été utilisées lors du premier essai VIRIS (voir I.5.4). Trois observateurs ont été contrôlés à la fois, sauf pour un seul essai où deux observateurs seulement ont été contrôlés (un des sujets convoqués ne s'était pas présenté). L'administrateur de l'essai a donné lecture des instructions d'essai puis a répondu aux questions avant d'engager une session pratique de 12 conditions pour familiariser les sujets avec les procédures d'essai et avec le matériel d'essai. Chaque condition d'essai durait 20 s et comportait un message de 2 s écran pour indiquer le numéro de la condition, la séquence vidéo de 10 s et une période de

vote de 8 s avec message à l'écran pour demander à l'observateur de voter au sujet de la qualité de l'image. Chacune de ces 36 conditions d'essai a été essayée deux fois, soit un total de 72 conditions.

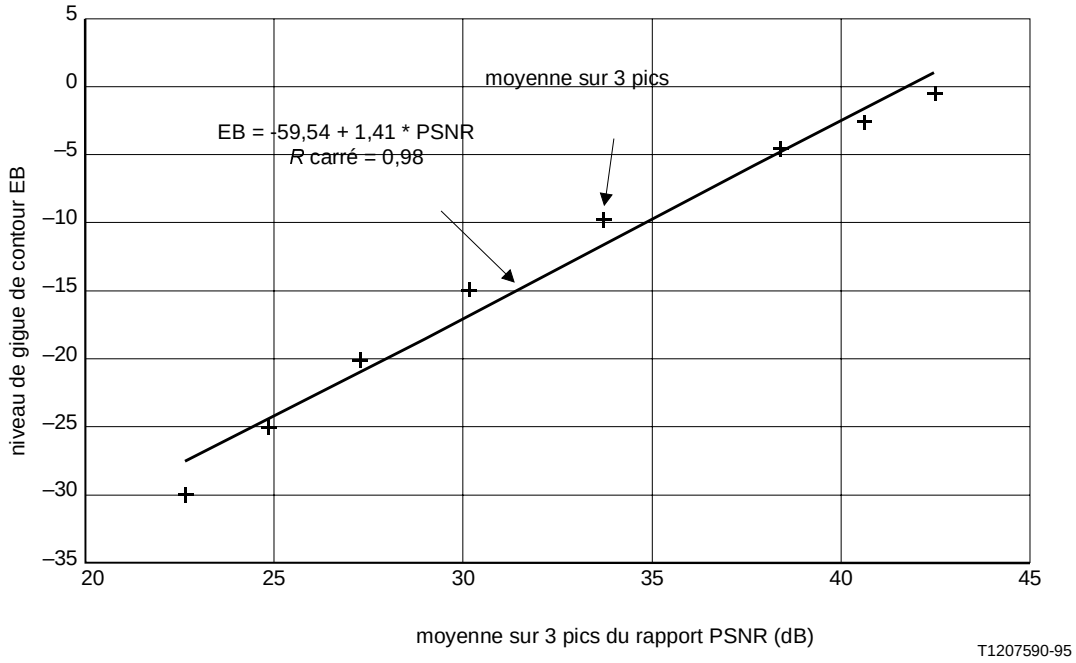


FIGURE I.6/P.930

**Rapport PSNR en fonction des données d'entrée dans VIRIS pour la gigue de contour**

L'essai a été conduit en deux parties séparées par une pause de 10 min. Chaque partie comprenait les 36 séquences d'essai différentes et durait environ 12 min. Les conditions de chaque partie étaient présentées dans un ordre pseudo-aléatoire, avec les deux contraintes suivantes:

a) le premier, le deuxième et le troisième groupe de 12 conditions contenaient chacun l'ensemble des 12 conditions d'essai de base;

b) chacun des trois groupes de 12 séquences d'images contenait un nombre égal d'images des trois séquences vidéo.

Ces contraintes ont été conçues pour compenser d'éventuels effets d'apprentissage pouvant affecter les résultats, comme une soudaine prise de conscience d'un aspect particulier d'une dégradation alors que l'essai est déjà bien avancé ou, par exemple, un changement progressif dans les critères utilisés par le sujet pour juger de la qualité, au fur et à mesure de la progression de l'essai. Les ordres aléatoires étaient différents dans chaque moitié de l'essai et étaient changés pour chaque groupe de sujets.

### I.6.5 Résultats d'essai subjectif de gigue de contour

La note moyenne d'opinion (MOS) a été calculée pour chacune des 12 conditions d'essai et chacune des 3 séquences vidéo. Les notes MOS dans les conditions de référence NTSC ont été respectivement de 4,7, 4,6 et 4,6 pour les séquences de course automobile, de football et de tennis de table. Les notes MOS dans les conditions de référence SIF pour les séquences correspondantes ont été respectivement de 3,2, 3,6 et 3,8, ce qui indique une nette baisse de qualité lorsque l'on passe du format NTSC au format SIF. Cette perte de qualité est essentiellement due à la baisse de résolution horizontale et verticale qui se produit lors de la conversion du format CCIR 601 au format SIF. Comme les algorithmes MPEG-1 utilisent le format SIF pour l'entrée des images, la qualité associée

au format SIF représente également la limite de qualité qui peut être obtenue avec les algorithmes MPEG-1.

La condition de référence SIF avec la séquence de football a obtenu une note MOS de 3,6, qui est inférieure de 0,6 à la note obtenue lors de l'essai VIRIS. Contrairement à cet essai, l'essai antérieur n'utilisait cependant pas les conditions de référence NTSC. Par conséquent, le matériel de stimulation ne couvrait pas toute l'étendue de l'échelle qualificative. Comme les sujets essayent probablement d'utiliser tous les qualificatifs de l'échelle pour juger le matériel de stimulation, il est possible qu'ils aient, lors de l'essai antérieur, donné une meilleure note aux conditions de référence SIF qu'ils ne l'auraient fait si une gamme complète de matériels de stimulation leur avaient été présentée. L'objet de l'inclusion des conditions NTSC dans le deuxième essai subjectif était d'obtenir une couverture complète de l'échelle qualificative pour réduire ou éliminer ce type de problème d'ancrage. Il est donc probable que la qualité SIF obtenue lors du deuxième essai soit plus représentative de la qualité SIF réelle que les résultats obtenus lors du premier essai subjectif.

La Figure I.7 montre, pour les 3 séquences d'images, des courbes de notes MOS en fonction des rapports PSNR de dégradation par gigue de contour VIRIS. Une fonction logistique ajustée sur les données des 3 séquences d'images et de la même forme que dans la précédente étude est également représentée sur cette figure. La Figure I.7 montre que la réaction subjective à la dégradation par gigue de contour a été un peu différente pour les trois séquences d'images d'essai. Les résultats obtenus avec les séquences de football et de tennis de table ont été similaires mais nettement différents de ceux qui ont été obtenus avec la séquence de course automobile, surtout aux rapports PSNR les plus élevés. Malgré les différences entre les résultats pour les 3 séquences d'images, la fonction logistique offre un bon ajustement sur les données des 3 séquences d'images avec un *R* carré = 0,90. Le *R* carré est une "statistique de validité de l'ajustement" qui va de 0 à 1, la valeur 1 représentant un ajustement optimal. Il indique aussi la proportion de variance dans les données qui est pris en compte par le modèle.

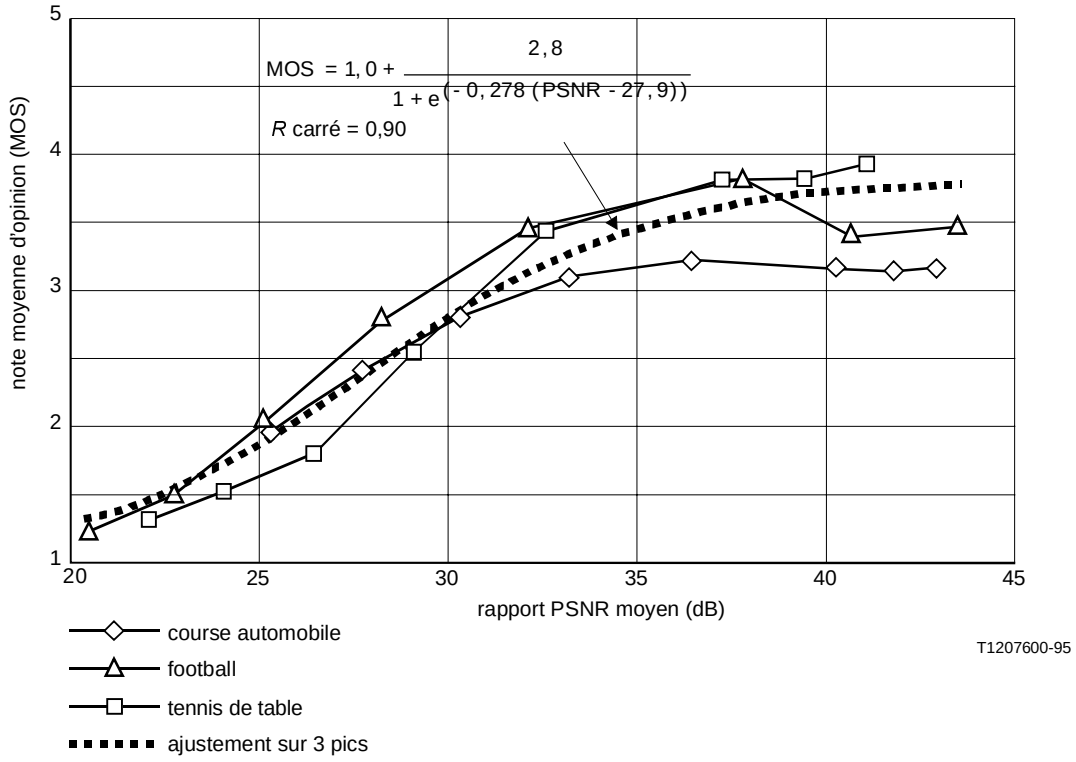


FIGURE I.7/P.930

**Rapport PSNR en fonction des notes MOS pour la gigue de contour**

## I.7 Conclusions

Un système réglable de dégradation vidéo de référence (RISV) est nécessaire pour évaluer la qualité des images produites par les systèmes vidéonumériques à compression. Un tel système doit produire une large gamme de qualités d'image en simulant les dégradations d'image dues aux codeurs numériques. Le système de dégradation vidéo de référence (VIRIS) décrit dans le présent appendice est un pas dans cette direction.

Le système RISV VIRIS peut introduire des simulations de dégradation par distorsion en mosaïque, flou, gigue de contour, bruit corrélé au signal, bruit de quantification et sautillement, dans des séquences d'images SIF. Ces dégradations peuvent être ajoutées séparément ou en combinaison. Toutes ces dégradations peuvent être caractérisées objectivement en termes de niveau de dégradation et de valeur de crête du rapport signal sur bruit (PSNR) sur toutes les trames de la séquence d'images. Les résultats d'essais subjectifs effectués pour trois de ces dégradations (la distorsion en mosaïque, bruit corrélé au signal et sautillement n'ont pas fait l'objet d'évaluations subjectives) ont indiqué que la réaction subjective aux dégradations simulées couvrait une large gamme de qualités d'image. Ces relations sont représentées sur les Figures I.4, I.5 et I.7 pour, respectivement, le bruit de quantification, le flou et la gigue de contour, avec chacune des séquences d'images testées. Une fonction logistique ajustée sur les données des 3 séquences d'images établit la relation entre notes MOS et rapports PSNR dans le cas de chacune des trois dégradations.

Les Figures I.2, I.3 et I.6 montrent des courbes de rapports PSNR dont la moyenne a été calculée sur les trois séquences d'images, en fonction du niveau d'entrée dans le système VIRIS de chacune des trois dégradations. Une fonction exponentielle d'ajustement aux données est également représentée sur chacune des courbes pour établir une relation entre rapports PSNR et données d'entrée VIRIS dans le cas des quatre dégradations. Une fonction linéaire a également été utilisée pour le système VIRIS. L'utilisation de ces fonctions avec les fonctions logistiques représentées sur les Figures I.4, I.5 et I.7 offre un passage entre notes MOS désirées et données d'entrée VIRIS pour chacune des quatre dégradations avec les scènes utilisées dans les résultats d'essai. L'utilisation de ces fonctions pour prédire les résultats pour d'autres scènes n'est pas recommandée actuellement. Elle fera l'objet d'un complément d'étude.

Des compléments d'étude sont également nécessaires sur le système VIRIS. Il s'agit de la simulation et des essais subjectifs de diverses combinaisons des dégradations. Le but est de relier l'opinion subjective désirée (les notes MOS) aux données d'entrée dans le système VIRIS pour diverses combinaisons de dégradations. Le résultat peut se présenter sous la forme d'une relation unique dans le cas de la combinaison de toutes les dégradations, ou de relations multiples (c'est-à-dire selon les cinq phénomènes étudiés dans la présente Recommandation) dans le cas de combinaisons spécifiques de dégradations. Des travaux futurs porteront sur l'approfondissement des dégradations créées réellement par les algorithmes de compression mis en oeuvre dans des codeurs vidéonumériques, sur la façon de simuler ces dégradations et sur la conception de nouvelles méthodes permettant d'établir une corrélation entre mesures objectives de la qualité et opinions subjectives.

## Bibliographie


[1] Contribution UIT-T COM 12-53 (1995), *Addition of Motion Correlated Artifacts to VIRIS (Adjonction au système VIRIS d'effets de corrélation cinétique),* Bellcore.

[2] Contribution UIT-T COM 12-21 (1993), *VIRIS, An Experimental Video Reference Impairment System (VIRIS, système expérimental de dégradation vidéo de référence),* Bellcore.

[3] Contribution UIT-T COM 12-35 (1994), *Modification of Video Reference Impairment System (VIRIS) to Improve Edge Busyness Impairment Simulation (Modification du système de dégradation vidéo de référence (VIRIS) pour améliorer la simulation de la dégradation par gigue de contour)*, Bellcore.

[4] HAYS (William L.): Statistics for Psychologists, *Holt, Rinehart and Winston*, 1963.

[5] ANIL (Jain): Fundamentals of Digital Image Processing, *Prentice Hall*, 1989.

[6] *Digital Filter Design Package*, Version 2.12, Atlanta Signal Processors Incorporated, décembre 1986.

## SÉRIES DES RECOMMANDATIONS UIT-T

| | |
|---|---|
| Série A | Organisation du travail de l'UIT-T |
| Série B | Moyens d'expression |
| Série C | Statistiques générales des télécommunications |
| Série D | Principes généraux de tarification |
| Série E | Réseau téléphonique et RNIS |
| Série F | Services de télécommunication non téléphoniques |
| Série G | Systèmes et supports de transmission |
| Série H | Transmission des signaux autres que téléphoniques |
| Série I | Réseau numérique à intégration de services |
| Série J | Transmission des signaux radiophoniques et télévisuels |
| Série K | Protection contre les perturbations |
| Série L | Construction, installation et protection des câbles et autres éléments des installations extérieures |
| Série M | Maintenance: systèmes de transmission, de télégraphie, de télécopie, circuits téléphoniques, et circuits loués internationaux |
| Série N | Maintenance: circuits internationaux de transmission radiophoniques et télévisuels |
| Série O | Spécifications des appareils de mesure |
| **Série P** | **Qualité de transmission téléphonique** |
| Série Q | Commutation et signalisation |
| Série R | Transmission télégraphique |
| Série S | Equipements terminaux de télégraphie |
| Série T | Equipements terminaux et protocoles des services télématiques |
| Série U | Commutation télégraphique |
| Série V | Communications de données sur le réseau téléphonique |
| Série X | Réseaux pour données et communication entre systèmes ouverts |
| Série Z | Langages de programmation |