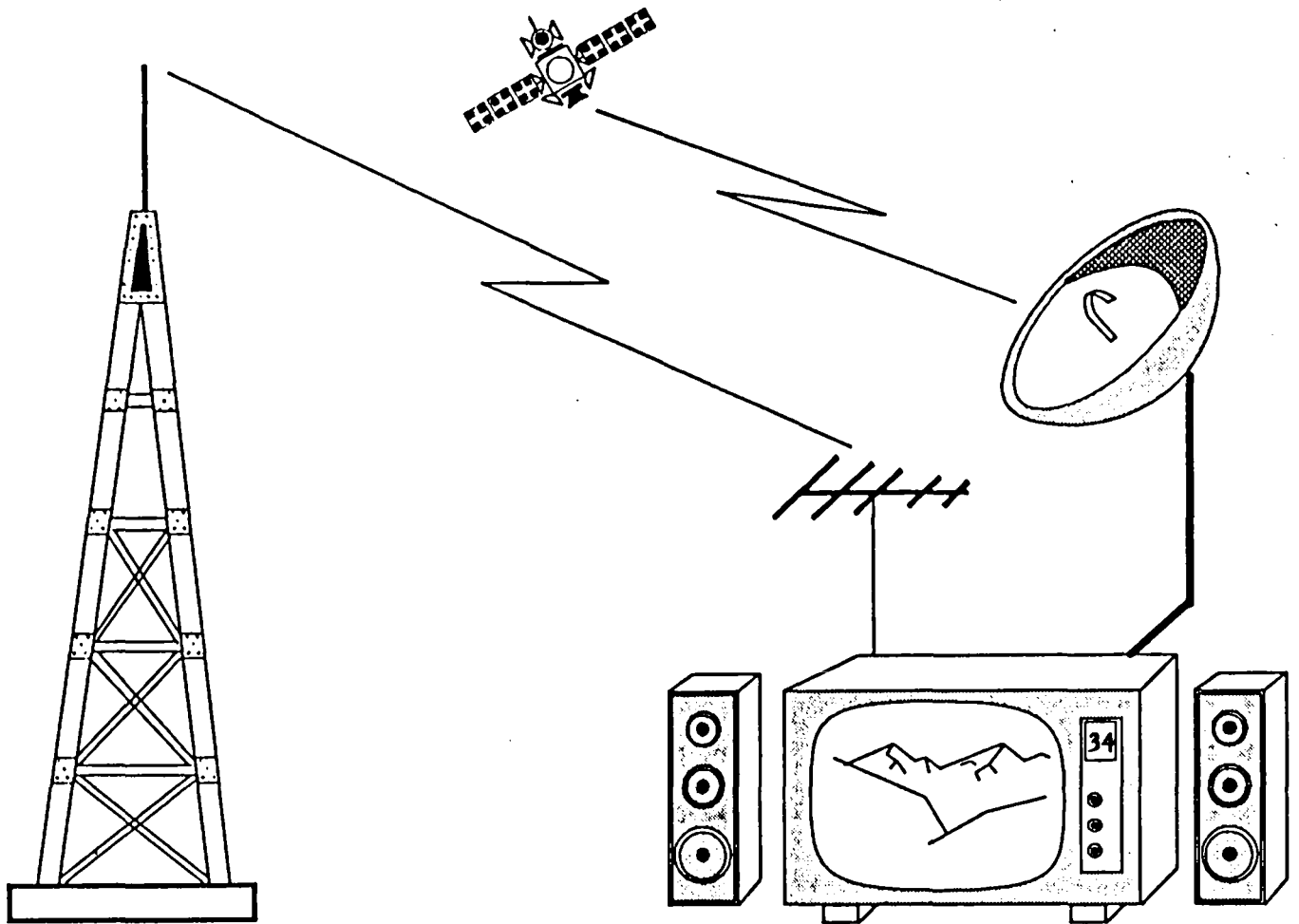




UNION INTERNATIONALE DES TÉLÉCOMMUNICATIONS

1992 -RECOMMANDATIONS DU CCIR

(Nouvelles et révisées en date du 15 septembre 1992)



Série RBT

SERVICE DE RADIODIFFUSION (TÉLÉVISION)



COMITÉ CONSULTATIF INTERNATIONAL DES RADIOCOMMUNICATIONS
ISBN 92-61-04592-8



Genève, 1992

© UIT 1992

Tous droits de reproduction réservés. Aucune partie de cette publication ne peut être reproduite ni utilisée sous quelque forme que ce soit et par aucun procédé, électronique ou mécanique, y compris la photocopie et les microfilms, sans l'accord écrit de l'UIT.



Recommandation 500-5 (1992)

Méthode d'évaluation subjective de la qualité des images de télévision

Extrait de la publication :

Recommandations CCIR : Série RBT : Service de radiodiffusion (télévision)
(Genève : UIT, 1992), pp. 166-189

This electronic version (PDF) was scanned by the International Telecommunication Union (ITU) Library & Archives Service from an original paper document in the ITU Library & Archives collections.

La présente version électronique (PDF) a été numérisée par le Service de la bibliothèque et des archives de l'Union internationale des télécommunications (UIT) à partir d'un document papier original des collections de ce service.

Esta versión electrónica (PDF) ha sido escaneada por el Servicio de Biblioteca y Archivos de la Unión Internacional de Telecomunicaciones (UIT) a partir de un documento impreso original de las colecciones del Servicio de Biblioteca y Archivos de la UIT.

(ITU) للاتصالات الدولي الاتحاد في والمحفوظات المكتبة قسم أجراه الضوئي بالمسح تصوير نتاج (PDF) الإلكترونية النسخة هذه والمحفوظات المكتبة قسم في المتوفرة الوثائق ضمن أصلية ورقية وثيقة من نقلاً.

此电子版（PDF版本）由国际电信联盟（ITU）图书馆和档案室利用存于该处的纸质文件扫描提供。

Настоящий электронный вариант (PDF) был подготовлен в библиотечно-архивной службе Международного союза электросвязи путем сканирования исходного документа в бумажной форме из библиотечно-архивной службы МСЭ.

RECOMMANDATION 500-5

MÉTHODE D'ÉVALUATION SUBJECTIVE DE LA QUALITÉ
DES IMAGES DE TÉLÉVISION

(Question 119/11)

(1974-1978-1982-1986-1990-1992)

Le CCIR,

considérant

- a) qu'une grande quantité d'informations a été recueillie sur les méthodes utilisées dans divers laboratoires pour l'évaluation de la qualité de l'image;
- b) que l'examen de ces méthodes montre qu'il existe un large accord entre les différents laboratoires sur certains aspects de ces essais;
- c) que l'adoption de méthodes normalisées est essentielle pour l'échange d'informations entre divers laboratoires;
- d) que les évaluations de la qualité ou de la dégradation de l'image faites en exploitation normale ou spéciale par certains techniciens chargés du contrôle, en utilisant des échelles à cinq notes, peuvent également s'inspirer de certains aspects des méthodes recommandées pour les essais en laboratoire;
- e) que l'introduction de nouvelles méthodes de traitement des signaux de télévision telles que le codage numérique et la réduction du débit binaire, les nouveaux types de signaux de télévision utilisant des composantes multiplexées dans le temps et, éventuellement, de nouveaux services tels que la télévision améliorée, la télévision à haute définition (TVHD) peuvent entraîner des modifications dans les méthodes d'évaluation subjective,

recommande

1. que les méthodes générales d'essai, les échelles et les conditions d'observation pour l'évaluation de la qualité des images décrites ci-après soient utilisées pour les expériences de laboratoire et aussi pour les évaluations en exploitation chaque fois que cela est possible;
2. que, dans un proche avenir, et cela malgré l'existence d'autres méthodes et la mise au point de nouvelles méthodes, les méthodes décrites aux § 2 et 3 de l'Annexe 1 à la présente Recommandation soient utilisées chaque fois que possible; et
3. que, puisqu'il est important de définir la base des évaluations subjectives, tous les rapports d'essai donnent la description la plus complète possible des configurations d'essai, du matériel d'essai, des observateurs et des méthodes.

Note 1 – L'Annexe 1 fournit des informations sur les méthodes d'évaluation subjective servant à définir la qualité des systèmes de télévision.

Note 2 – L'Annexe 2 fournit des informations sur les méthodes d'évaluation subjective des dégradations résultant du codage numérique des signaux de télévision.

ANNEXE 1

1. Introduction

Les méthodes d'évaluation subjective servent à définir la qualité des systèmes de télévision au moyen de mesures qui laissent prévoir avec précision les réactions de ceux qui observeront les systèmes à l'essai. On sait bien qu'à cet égard les méthodes objectives ne peuvent rendre exactement compte de la qualité d'un système et qu'il faut donc les compléter par des mesures subjectives.

On considère en général deux catégories d'évaluations subjectives. En premier lieu, celles qui établissent la qualité d'un système dans les meilleures conditions. On les appelle évaluations de la qualité. En second lieu, celles qui établissent la faculté qu'ont les systèmes de conserver leur qualité dans des conditions non idéales de transmission ou d'émission. On les appelle couramment évaluations des dégradations.

Pour procéder à des évaluations subjectives appropriées, il faut d'abord choisir parmi les diverses options disponibles celles qui conviennent le mieux aux objectifs et aux circonstances des problèmes d'évaluation en question. Cela conduit en pratique à décider du choix des méthodes et matériels d'essai et des conditions d'observation.

1.1 Choix des méthodes d'essai

Pour les évaluations de la télévision, on a recouru à des méthodes d'essai fondamentales très diverses. Cependant, dans la pratique, chaque problème d'évaluation particulier suppose des méthodes particulières. Le Tableau 1 indique les problèmes classiques d'évaluation et les méthodes qui servent à les traiter.

TABLEAU 1
Choix des méthodes d'essai

Problème d'évaluation	Méthode utilisée	Origine (Recommandation 500)
Mesurer la qualité des systèmes par rapport à une référence	Méthode à double stimulus utilisant une échelle de qualité continue	§ 3
Mesure quantitative de la qualité des systèmes (quand on ne dispose d'aucune référence)	Méthode utilisant une échelle de rapports (1)	§ 4
Comparer la qualité de plusieurs systèmes (quand on ne dispose d'aucune référence)	Méthode de comparaison directe <i>ou</i> méthode utilisant une échelle de rapports (1)	§ 4
Identifier les paramètres qui permettent de déterminer les différences entre les systèmes	Méthode à échelle multidimensionnelle <i>ou</i> méthode d'analyse des paramètres	§ 4
Mesurer les différences entre systèmes en fonction de paramètres spécifiques	Méthode multivalente	
Mesurer la résistance des systèmes (c'est-à-dire vis-à-vis des défaillances)	Méthode à double stimulus utilisant une échelle de dégradation	§ 2
Mesure quantitative de la résistance des systèmes (c'est-à-dire vis-à-vis des défaillances)	Méthode utilisant une échelle de rapports (1)	§ 4
Définir le moment où une dégradation devient visible	Estimation du seuil par choix forcé <i>ou</i> méthode d'ajustement	§ 4
Déterminer si on remarque une différence entre systèmes	Méthode de choix forcé	§ 4

(1) Certaines études donnent à penser que cette méthode est plus stable lorsqu'on dispose d'une large gamme de qualité.

1.2 Choix du matériel d'essai

Plusieurs méthodes ont servi à définir le type de matériel d'essai nécessaire aux évaluations de la télévision. Cependant, dans la pratique, il faut utiliser certains types de matériel d'essai pour traiter des problèmes d'évaluation particuliers. Le Tableau 2 indique les problèmes classiques d'évaluation et le matériel d'essai qui sert à les traiter.

1.3 Choix des conditions d'observation

Pour les évaluations de la télévision conventionnelle, il existe un ensemble particulier de conditions d'observation. Toutefois, on peut se placer à des distances d'observation différentes selon que les évaluations sont normales ou critiques (voir le Tableau 3).

TABLEAU 2

Choix du matériel d'essai*

Problème d'évaluation	Matériel utilisé	Origine (Recommandation 500)
Qualité globale avec matériel moyen	Général, «critique sans excès»	§ 2
Capacité, applications critiques (par exemple: contribution, post-traitement, etc.)	Gamme étendue, y compris matériel très critique pour l'application à l'essai	Annexe 1
Qualité des systèmes «adaptatifs»	Matériel très critique pour le schéma «adaptatif» utilisé	Annexe 1
Recenser les points vulnérables et les améliorations possibles	Matériel critique propre à la caractéristique	
Identifier les paramètres qui distinguent les systèmes	Large gamme de séquences complexes	
Conversion de normes	Critique pour ce qui les distingue (par exemple, fréquence de trame)	

* Il va de soi que tout matériel d'essai est du type qu'on pourrait rencontrer dans des programmes de télévision. Voir les Appendices 1 et 2 pour plus de renseignements sur le choix du matériel d'essai.

TABLEAU 3

Choix des conditions d'observation

Problème d'évaluation	Conditions d'observation	Origine (Recommandation 500)
Evaluer les systèmes classiques	Observer à une distance de 6 fois la hauteur de l'image	§ 2
Evaluer les systèmes classiques dans des conditions critiques	Observer à une distance de 4 fois la hauteur de l'image pour les systèmes à 625 lignes et 4 ou 5 fois pour les systèmes à 525 lignes	§ 2

2. La méthode à double stimulus utilisant une échelle de dégradation (Méthode UER)

2.1 Description générale

Dans le cadre d'une évaluation typique, on peut chercher à évaluer soit un nouveau système, soit l'effet d'une dégradation due à la transmission. La personne chargée d'organiser l'évaluation devra tout d'abord choisir un matériel d'essai suffisant pour que l'évaluation puisse être significative et définir les conditions d'observations à utiliser. Si la variation de paramètres présente de l'intérêt, il est nécessaire de choisir un jeu de valeurs de paramètres réparties sur l'échelle de dégradation à intervalles plus ou moins égaux. Si on évalue un nouveau système dont on ne peut faire varier de cette façon les valeurs des paramètres, il faut alors, soit ajouter de nouvelles dégradations subjectivement identiques, soit utiliser une autre méthode, par exemple celle indiquée au § 3.

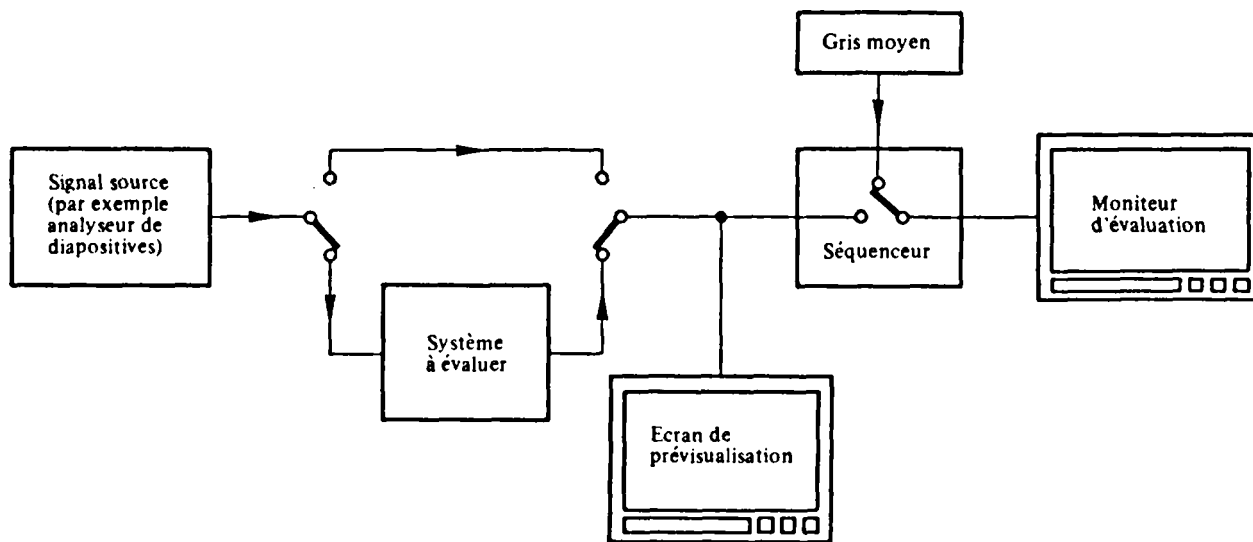
La méthode à double stimulus (UER) est cyclique en ce sens que l'on présente d'abord une image de référence non dégradée puis la même image dégradée à l'observateur qui est ensuite prié de donner son avis sur la seconde image, tout en gardant à l'esprit la première. Au cours des séances, qui durent au plus une demi-heure, une série d'images ou de séquences d'images, couvrant toutes les combinaisons requises, sont présentées à l'observateur. Les ordres de présentation des images et des dégradations sont aléatoires. L'image de référence fait partie des images ou des séquences d'images à évaluer. A l'issue de la série de séances, on calcule la note moyenne pour chaque condition d'essai et chaque image d'essai.

Cette méthode fait appel à l'échelle de dégradation qui donne généralement des résultats plus stables pour des dégradations faibles que pour des dégradations importantes. Bien qu'elle ait parfois été utilisée avec des gammes de dégradations limitées, cette méthode convient mieux pour une gamme de dégradations complète.

2.2 Mode opératoire général

Le système d'essai aura la configuration indiquée à la Fig. 1 ci-dessous.

FIGURE 1
Configuration du système d'essai pour la méthode à double stimulus
utilisant une échelle de dégradation



On présente aux observateurs un moniteur d'évaluation qui reçoit un signal au moyen d'un séquenceur. Le trajet du signal jusqu'au séquenceur peut être, soit direct (le signal vient de la source), soit indirect (le signal passe par le système à évaluer). Les observateurs voient défiler devant eux une série d'images ou de séquences d'images d'essai. Ces images sont présentées par paires: la première image de la paire provient directement de la source, la seconde est la même image qui est passée par le système à évaluer.

2.3 Signaux source

Le signal source fournit directement l'image de référence et le signal source pour le système à évaluer. Sa qualité doit être optimale pour la norme de télévision utilisée. Il est essentiel que l'image de référence de la paire d'images présentée n'ait pas de défauts si l'on veut obtenir des résultats stables.

Les images et les séquences d'images mémorisées numériquement constituent les signaux source les plus facilement reproductibles et ce sont donc elles que l'on préfère. Elles peuvent être échangées entre différents laboratoires pour obtenir de meilleures comparaisons entre systèmes. Le format de bande D-1 4:2:2 (Recommandation 657) devra servir de base pour l'échange d'images et de séquences d'images source lorsque ces machines seront disponibles sur le marché, à des prix économiques. Les formats de bande pour ordinateur conviennent également.

A court terme, les analyseurs de diapositives 35 mm sont préférables pour des images fixes. La résolution disponible est satisfaisante pour l'évaluation des systèmes de télévision traditionnels. La colorimétrie et d'autres caractéristiques du film peuvent donner une apparence subjective différente des images issues de caméra de studio. Si ces paramètres ont une influence sur les résultats, il convient d'utiliser des sources de signaux provenant directement du studio, bien qu'elles soient souvent beaucoup moins pratiques. En règle générale, les analyseurs de diapositives seront adaptés pour chaque image afin d'obtenir la meilleure qualité subjective possible de l'image puisque telle serait la situation dans la pratique.

On évalue souvent les possibilités de traitement en aval grâce à la technique d'incrustation d'image. En studio, l'incrustation d'image est très sensible à l'éclairage. Pour les évaluations, il faut donc utiliser de préférence une paire de diapositives spécialement conçues pour l'incrustation d'image qui donnera toujours des résultats de haute qualité. On peut introduire, le cas échéant, un mouvement dans la diapositive de premier plan.

2.4 Conditions d'observation

Les conditions d'observation doivent être les suivantes:

2.4.1 Conditions générales

a) Rapport de la distance d'observation à la hauteur de l'image	4H et 6H*
b) Luminance de crête de l'écran	70 cd/m ²
c) Rapport de la luminance du tube image inactif à la luminance de crête	≤ 0,02
d) Rapport de la luminance de l'écran, quand on reproduit seulement le niveau du noir dans une salle complètement noire, à celle qui correspond au blanc maximal	environ 0,01
e) Rapport de la luminance de l'arrière-plan, derrière le moniteur-image, à la luminance de crête de l'image	environ 0,15
f) Eclairage de la salle dû à d'autres sources	faible
g) Chromaticité de l'arrière-plan	D ₆₅
b) Rapport de l'angle solide sous-tendu par la zone d'arrière-plan satisfaisant cette spécification, à l'angle solide sous-tendu par l'image	≥ 9

2.4.2 Conditions particulières

a) Nombre type d'observateurs à 4H par moniteur d'évaluation	2 (pour la première moitié des séances) 3 (pour l'autre moitié)
b) Nombre type d'observateurs à 6H par moniteur d'évaluation	comme ci-dessus
c) Moniteur d'évaluation**	haute qualité, taille d'écran: 50-60 cm (22"-26")
d) Brilliance et contraste de la visualisation	établi via PLUGE (voir la Recommandation 814)
e) Nombre type d'observateurs par moniteur d'évaluation	5 (2 à 4H et 3 à 6H pour la première séance, 3 à 4H et 2 à 6H pour la séance suivante, etc.)
f) Nature de la (des) salle(s) d'observation	pièce dont 3 côtés sont tendus de rideaux blancs, le 4 ^e (derrière les observateurs) étant tendu de rideaux gris.

2.5 Séance d'évaluation

Une séance ne devra pas dépasser une demi-heure et comprendra au plus 40 présentations (voir le § 2.6).

Les séances sont organisées par groupes de deux afin de permettre à tous les observateurs de voir les images ou les séquences d'images tant à 4H qu'à 6H. S'il y a trop de conditions expérimentales pour une seule paire de séances, il faudra organiser d'autres paires de séances. Il convient de choisir un ordre aléatoire pour la présentation des images (par exemple, déduit de carrés gréco-latins); quoi qu'il en soit, les conditions d'essai doivent être présentées dans un ordre permettant d'équilibrer, séance après séance, tous les effets que les phénomènes de fatigue et d'adaptation peuvent avoir sur les notations. Certaines présentations peuvent être répétées d'une séance à l'autre pour vérifier la cohérence. Chaque condition expérimentale devra être présentée deux fois au cours de la même séance.

* 6H est la distance préférée pour les évaluations de systèmes de télévision traditionnels (625/50, 525/60); toutefois, une distance de 4H peut également convenir à condition que les résultats soient consignés séparément pour qu'il n'y ait pas de différence significative dans les moyennes obtenues.

** Lorsqu'on utilise plus d'une salle d'observation, les moniteurs d'évaluation devront être soigneusement synchronisés.

Les images et les dégradations seront présentées dans un ordre pseudo-aléatoire et, de préférence, dans un ordre différent pour chaque séance. En tout état de cause, la même image ne devra jamais être présentée deux fois consécutivement, que les niveaux de dégradation soient les mêmes ou qu'ils soient différents.

La gamme des dégradations devra être choisie de telle sorte que toutes les notes soient utilisées par la majorité des observateurs; l'objectif est de tendre vers une moyenne générale (moyenne de tous les jugements émis au cours de l'expérience) voisine de 3.

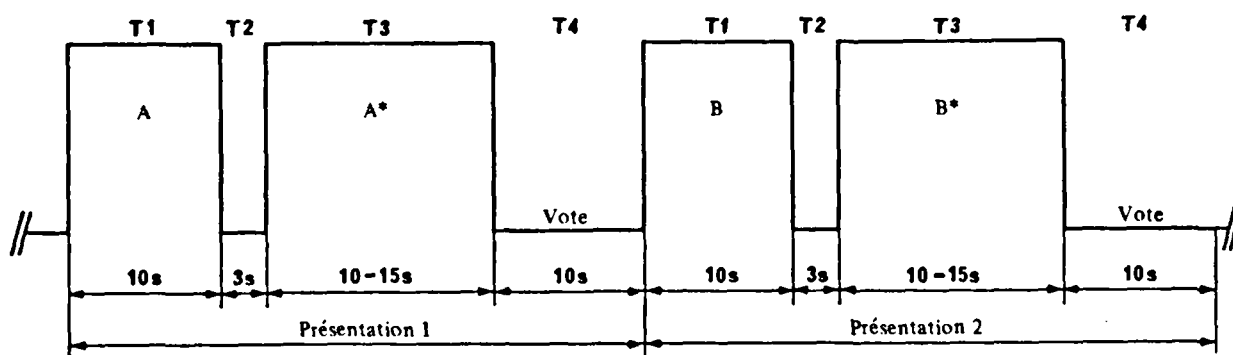
Une séance ne dépassera pas une trentaine de minutes, y compris les explications et présentations préliminaires; la séquence d'essai pourra commencer par quelques images représentatives de la gamme des dégradations; les jugements relatifs à ces images ne seront pas pris en considération dans les résultats finals.

L'Appendice 2 donne des indications supplémentaires pour le choix des niveaux de dégradation.

2.6 Présentation du matériel d'évaluation

Une séance d'évaluation comprend un certain nombre de présentations, organisées de la façon indiquée à la Fig. 2 ci-dessous.

FIGURE 2
Organisation des présentations du matériel d'évaluation



A, B: image ou séquence d'images de référence
A*, B*: image ou séquence d'images à évaluer

Chaque présentation comporte quatre phases:

T1 = 10 s	Image de référence
T2 = 3 s	Gris moyen produit par un niveau vidéo d'environ 200 mV
T3 = 10-15 s	Matériel d'évaluation
T4 = 10 s	Gris moyen

La phase T3 peut durer entre 10 et 15 s. Même pour des images animées, il a été démontré que le fait de prolonger la phase T3 au-delà des 15 s ne permettait pas aux observateurs de mieux évaluer les images.

2.7 Observateurs

Il faut au moins 15 observateurs qui ne seront ni des spécialistes, en ce sens qu'ils ne s'occupent pas directement, dans le cadre de leur travail habituel, des questions liées à la qualité des images de télévision, ni des observateurs expérimentés*. Avant chaque séance, les observateurs seront sélectionnés pour leur acuité visuelle normale ou rendue normale par correction et leur vision normale des couleurs, cela à l'aide de diagrammes choisis à cet effet.

* Selon des résultats préliminaires, des observateurs non-experts confrontés à une qualité de transmission et à des technologies de visualisation supérieures, donneraient des résultats plus critiques.

2.8 *Echelles d'évaluation*

Il convient d'utiliser une échelle de dégradation à cinq notes:

- 5 imperceptible
- 4 perceptible mais non gênant
- 3 légèrement gênant
- 2 gênant
- 1 très gênant.

Les observateurs utiliseront un formulaire représentant très clairement l'échelle, avec des cases numérotées ou un autre moyen pour consigner les notes.

2.9 *Choix du matériel d'évaluation*

Pour certains paramètres, les dégradations observées pour la plupart des images ou des séquences d'images peuvent être plus ou moins identiques. Dans ces conditions, les résultats obtenus à partir d'un très petit nombre d'images ou de séquences d'images (par exemple, 2) peuvent rester significatifs.

Toutefois, l'impact des nouveaux systèmes dépend souvent largement de la scène et du contenu de la séquence. En pareil cas, pendant la totalité des heures de programme, il y aura une distribution statistique des probabilités de dégradation et du contenu des images ou des séquences d'images. Etant donné qu'en règle générale on ne connaît pas la forme de cette distribution statistique, le choix du matériel d'évaluation et l'interprétation des résultats doivent être faits avec beaucoup de soin.

En général, il est essentiel d'avoir un matériel de caractère critique car il est possible de tenir compte de ce facteur lors de l'interprétation des résultats mais il n'est pas possible d'extrapoler des résultats à partir d'un matériel non critique. Lorsque la scène ou le contenu de la séquence influence les résultats, il convient de choisir un matériel «critique mais sans excès» pour le système à évaluer. Par «sans excès» on entend que les images pourront raisonnablement faire partie de programmes normaux. En pareil cas, il faut utiliser au moins quatre éléments: par exemple, deux d'entre eux sont véritablement critiques, les deux autres le sont modérément.

Un certain nombre d'organisations ont mis au point des images ou des séquences d'images d'essai fixes. A l'avenir, on espère traiter ce problème dans le cadre du CCIR.

Le CCIR a proposé un matériel pour évaluer des systèmes numériques dans lesquels une réduction du débit binaire à 30-33 Mbit/s est appliquée aux signaux conformes à la Recommandation 601. L'évaluation de ces systèmes doit porter également sur la possibilité de diverses opérations de traitement en aval telles que l'incrustation d'image. Dans ce cas, le système d'incrustation d'image doit être inclus à la fois dans le trajet direct du signal et dans le trajet du signal via le système à évaluer. Ces signaux peuvent ensuite être inclus dans les présentations d'évaluation. Avec cette méthode, il est toutefois important d'éviter des images ou séquences d'images de référence qui sont elles-mêmes dégradées. Si l'évaluation de la dégradation supplémentaire d'une image déjà dégradée présente un intérêt, il convient de les utiliser comme séquences à évaluer.

Les Appendices 1 et 2 donnent des indications supplémentaires pour le choix du matériel d'essai.

2.10 *Introduction aux évaluations*

La méthode d'évaluation et les types de dégradations susceptibles de se produire seront présentés avec soin aux observateurs. Les questions visant à faciliter la compréhension seront autorisées mais les instructions ne doivent pas varier d'une séance à l'autre et en répondant aux questions, il faudra veiller à ne pas faire preuve de partialité.

Au début de chaque séance, une explication est donnée aux observateurs sur le type d'évaluation, l'échelle d'évaluation, la séquence elle-même et le séquençement (image de référence, gris, image d'essai, vote). Il convient que la dynamique et le type de dégradations à évaluer soient illustrés sur des images autres que celles utilisées dans les essais mais ayant une sensibilité comparable. Il ne faut pas en déduire que la plus mauvaise qualité observée correspond nécessairement à la notation subjective la plus basse. Les observateurs seront priés de fonder leur jugement sur l'impression globale donnée par l'image et d'exprimer ce jugement à l'aide des termes utilisés pour définir l'échelle d'évaluation subjective.

Les observateurs seront priés de regarder l'image pendant toute la durée de T1 et T3. Le vote sera autorisé uniquement pendant T4.

2.11 Présentation des résultats

La cohérence des résultats sera vérifiée en étudiant les notes données par le même observateur à la même image pendant la même séance. Si les notations diffèrent de deux points ou plus, les deux notes seront rejetées.

Pour chaque paramètre d'essai, il faut donner la moyenne et l'écart type de la distribution statistique des notes d'évaluation. Si l'évaluation portait sur la variation de la dégradation en fonction de la variation de la valeur d'un paramètre, il conviendra d'utiliser des courbes de régression. Une courbe de régression appropriée en coordonnées logarithmiques permettra de représenter les résultats sous forme d'une droite. C'est là le mode de présentation préféré.

Les résultats doivent être donnés avec les informations suivantes:

- détails de la configuration de l'expérience,
- détails du matériel d'évaluation,
- type de source d'image et de moniteur d'évaluation,
- nombre et type d'observateurs,
- systèmes de référence utilisés,
- moyenne générale de l'expérience,
- résultats moyens originaux et corrigés et écarts types si un ou plusieurs observateurs ont été éliminés selon la procédure donnée dans le paragraphe ci-après.

Il faut calculer, après chaque séance, les valeurs moyennes $E(X_j)$ et les écarts types $\sigma(X_j)$ associés à chaque niveau de dégradation ou système de traitement à évaluer j . Ces valeurs moyennes reposent sur une distribution dont les deux variables sont les scènes et les observateurs. Il faut vérifier, au moyen du test β_2 , si cette distribution est normale ou non (en calculant le coefficient de kurtosis de la fonction, c'est-à-dire le rapport moment du quatrième ordre sur carré du moment du deuxième ordre). Si β_2 est compris entre 2 et 4, on peut considérer que la distribution est normale. Les résultats X_{ij} de chaque distribution j sont alors à comparer avec la valeur moyenne associée, plus l'écart type associé multiplié par 2 (normale) ou par $\sqrt{20}$ (non normale), P_i , et à la valeur moyenne associée, moins ce même écart type multiplié par 2 ou $\sqrt{20}$, Q_i . Chaque fois que les résultats donnés par un observateur se situent hors de cet intervalle, il faut les enregistrer sur un compteur associé à chaque observateur; il faut deux compteurs séparés, un pour les valeurs supérieures P_i , l'autre pour les valeurs inférieures Q_i . On calcule ensuite les deux rapports suivants: $P_i + Q_i$ sur le nombre total de résultats donnés par chaque observateur au cours de toute la séance et $P_i - Q_i$ sur $P_i + Q_i$ en valeur absolue. Si le premier est supérieur à 5% et le second inférieur à 30%, il faut éliminer l'observateur i .*

La procédure ci-dessus peut également être exprimée mathématiquement comme suit:

$$\text{si } X_{ij} \geq E(X_j) + 2 \cdot \sigma(X_j) \quad (\text{distribution normale})$$

ou

$$\text{si } X_{ij} \geq E(X_j) + \sqrt{20} \cdot \sigma(X_j) \quad (\text{distribution non normale})$$

$$\text{alors } P_i = P_i + 1.$$

$$\text{Si } X_{ij} \leq E(X_j) - 2 \cdot \sigma(X_j) \quad (\text{distribution normale})$$

ou

$$\text{si } X_{ij} \leq E(X_j) - \sqrt{20} \cdot \sigma(X_j) \quad (\text{distribution non normale})$$

$$\text{alors } Q_i = Q_i + 1.$$

$$\text{Si } \frac{P_i + Q_i}{\text{Total des résultats par observateur}} > 0,05 \quad \text{et} \quad \left| \frac{P_i - Q_i}{P_i + Q_i} \right| < 0,3$$

alors éliminer l'observateur i .

* Il ne faut pas appliquer cette procédure plus d'une fois aux résultats d'une expérience donnée. En outre, on la réservera aux cas où il y a relativement peu d'observateurs (moins que 20 par exemple) dont aucun n'est un expert.

3. Méthode à double stimulus utilisant une échelle de qualité continue

3.1 Description générale

Dans une évaluation typique, on peut évaluer soit un nouveau système, soit les effets de la transmission sur la qualité. On estime que la méthode à double stimulus est particulièrement utile lorsqu'il n'est pas possible de créer des conditions expérimentales et des stimulus d'essai représentant toute la gamme de qualité.

La méthode est cyclique en ce sens que l'on présente à l'observateur une paire d'images, chacune provenant de la même source, l'une ayant passé par le système à évaluer et l'autre venant directement de la source. L'observateur est prié d'évaluer la qualité des deux images.

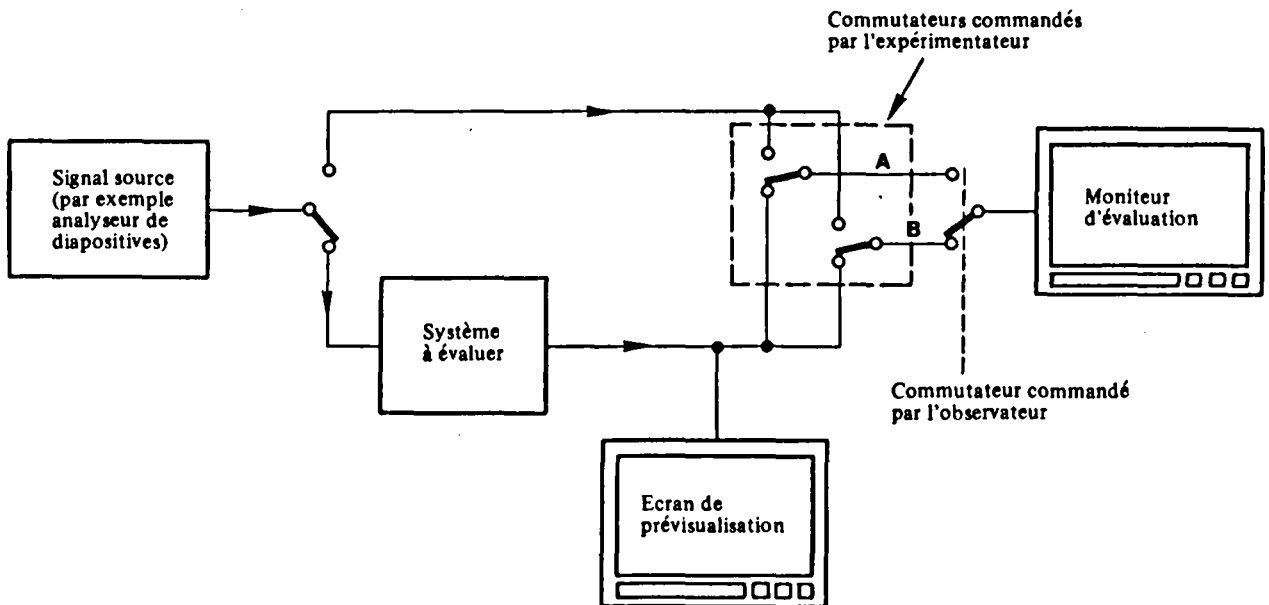
Au cours des séances qui durent au plus 30 min, on présente à l'observateur une série de paires d'images, les images constituant la paire alternant aléatoirement. Les images et les dégradations, couvrant toutes les combinaisons requises, sont également présentées dans un ordre aléatoire. A l'issue des séances, on calcule les moyennes pour chaque condition expérimentale et chaque image d'essai.

3.2 Mode opératoire général

Le système d'essai aura la configuration indiquée à la Fig. 3 ci-dessous.

FIGURE 3

Configuration du système d'essai pour la méthode à double stimulus utilisant une échelle de qualité continue



Il y a deux variantes de cette méthode, la variante (I) et la variante (II), exposées ci-après.

- (I) L'observateur, habituellement seul, est autorisé à passer de la condition A à la condition B et inversement jusqu'à ce qu'il se soit fait une opinion sur chacune d'elles. Les trajets A et B reçoivent l'image de référence directe ou l'image qui est passée par le système à évaluer. L'image et le trajet sont alternés de façon aléatoire d'une condition d'essai à l'autre. Ce phénomène est noté par l'expérimentateur mais non annoncé aux observateurs.
- (II) On présente consécutivement aux observateurs les images provenant des trajets A et B, afin qu'ils se fassent une opinion sur chacune d'elles. Pour chaque présentation, les trajets A et B reçoivent l'image comme dans la variante (I) ci-dessus. La stabilité des résultats obtenus avec cette variante, qui utilise une échelle de qualité limitée, est encore à l'étude.

3.3 Signaux source

Comme pour la méthode décrite au § 2; toutefois, il se peut qu'une image de référence dégradée n'ait pas le même effet sur la stabilité.

3.4 Conditions d'observation

Les mêmes que pour la méthode décrite au § 2. Toutefois, pour la variante (I), il y a un seul observateur par moniteur d'évaluation.

3.5 Séance d'évaluation

La même que pour la méthode décrite au § 2. Au moins pour la variante (I) il n'est pas nécessaire d'avoir une moyenne générale de 3.

3.6 Présentation du matériel d'évaluation

Une séance d'évaluation comprend un certain nombre de présentations. Dans le cas de la variante (I), qui ne nécessite qu'un seul observateur, l'observateur peut, pour chaque présentation, passer du trajet A au trajet B et inversement jusqu'à ce que l'observateur ait mentalement la mesure de la qualité associée à chaque signal. Il peut répéter cette opération deux ou trois fois pendant des laps de temps ne dépassant pas 10 s. Dans la variante (II), qui fait appel à plusieurs observateurs simultanément, avant d'enregistrer les résultats, chaque paire de conditions est présentée une ou plusieurs fois pendant un laps de temps égal, afin de permettre aux observateurs de mesurer mentalement les qualités associées à ces conditions. Ensuite, chaque paire est visualisée une ou plusieurs fois tandis que les résultats sont enregistrés. Le nombre de répétitions dépend de la longueur des séquences d'essai. Pour des images fixes, une séquence de 3 à 4 s et cinq répétitions (avec notation pendant les deux dernières) peut convenir. Pour des images en mouvement avec des défauts variant dans le temps, une séquence de 10 s avec deux présentations (et notation pendant la seconde) peut être appropriée.

Si des considérations pratiques limitent la durée des séquences disponibles à moins de 10 s, il est possible de recourir à des compositions utilisant ces séquences plus courtes sous forme de segments afin d'étendre jusqu'à 10 s le temps de visualisation. Pour réduire au minimum la discontinuité aux jonctions, des segments de séquence successifs peuvent être inversés dans le temps (appelés parfois visualisation «palindromique»). Il convient cependant de s'assurer que les conditions d'essai au cours de la visualisation de segments en sens inverse représentent des processus de causalité, c'est-à-dire qu'ils doivent être obtenus par le passage du signal source à l'envers à travers le système en cours d'évaluation.

3.7 Observateurs

Comme pour la méthode décrite au § 2.

3.8 Echelle d'évaluation

La méthode exige l'évaluation simultanée de deux versions de chaque image. Dans chaque paire d'images, l'une n'est pas dégradée alors que l'autre peut comporter ou non une dégradation. L'image non dégradée sert de référence mais les observateurs ignorent laquelle est l'image de référence. Dans la série d'essais, la position de l'image de référence est modifiée de façon pseudo-aléatoire.

Les observateurs doivent simplement évaluer la qualité globale de l'image pour chaque présentation en faisant une marque sur une échelle verticale. Les échelles verticales sont présentées par paires pour tenir compte de la double présentation de chaque image. Les échelles constituent un système de notation continu afin d'éviter les erreurs de quantification mais elles sont divisées en cinq segments égaux qui correspondent à l'échelle de qualité normale à cinq notes du CCIR. Les adjectifs qui caractérisent les différents niveaux sont les mêmes que ceux utilisés normalement; dans le cas présent, ils sont indiqués comme référence et imprimés uniquement à gauche de la première échelle de chaque rangée de dix colonnes doubles sur la feuille de notation. La Fig. 4 représente une partie d'une feuille de notation typique. On évite toute confusion possible entre les graduations de l'échelle et les résultats des essais en imprimant les échelles en bleu et en indiquant les résultats en noir.

3.9 Choix du matériel d'évaluation

Comme pour la méthode décrite au § 2.

3.10 Introduction à l'évaluation

Comme pour la méthode décrite au § 2, à l'exception du dernier alinéa du § 2.10.

FIGURE 4

Partie d'un formulaire de notation de la qualité
utilisant des échelles continues

	27		28		29		30		31	
	A	B	A	B	A	B	A	B	A	B
Excellent										
Bon										
Assez bon										
Médiocre										
Mauvais										

3.11 Présentation des résultats

Deux méthodes différentes sont possibles:

- Dans le premier cas, les résultats peuvent être exprimés sous la forme d'un essai comparatif, c'est-à-dire que l'on indique directement la différence de qualité par rapport à la condition de référence. Pour chaque paramètre évalué, il faut donner la moyenne et l'écart type de la distribution statistique de la différence mesurée.
- Dans le second cas (méthode de présentation préférée), les résultats peuvent être convertis et exprimés avec les termes utilisés pour décrire une qualité équivalente. Pour chaque condition d'essai, les deux évaluations (de référence et d'essai) sont converties: les barèmes inscrits sur la feuille de notation sont traduits en notes normalisées comprises entre 0 et 100. Pour chaque système à évaluer, ces notes sont ensuite pondérées compte tenu des différents groupes d'observateurs, distances d'observation, et images afin de donner les notes moyennes pour les conditions de référence et de test, cela pour chaque combinaison des variables.

Etant donné que les notes moyennes pour les conditions de référence sont toujours inférieures à 1,0, il faut adopter une nouvelle échelle applicable aux notes obtenues pour les conditions d'essai en retranchant la dégradation résiduelle. La note moyenne pour les conditions de référence est considérée comme la dégradation résiduelle. Les résultats de cette soustraction sont exprimés en unités de dégradation (imps) et peuvent être reconvertis en notes moyennes si on le souhaite.

Le rapport doit comporter les mêmes renseignements supplémentaires que pour la méthode indiquée au § 2, à l'exception de la note moyenne.

4. Autres méthodes d'évaluation

Lorsque les conditions le permettent, on utilisera les méthodes à un seul stimulus et les méthodes comparatives avec stimulus.

4.1 Méthodes à un seul stimulus

Dans ce type de méthode, une seule image ou séquence d'images est présentée à l'observateur qui fournit une notation de l'ensemble de la présentation.

4.1.1 Observateurs

Pour les essais en laboratoire, les observateurs sont choisis de la façon indiquée au § 2.7. Le nombre d'observateurs dépend de la sensibilité et de la fiabilité de la procédure d'essai retenue ainsi que de l'ampleur escomptée de l'effet évalué. Dans des conditions normales, on utilise un échantillon de 10 à 20 observateurs par évaluation.

4.1.2 Images d'essai

Pour les essais en laboratoire, le contenu des images d'essai sera choisi selon la description faite au § 2.9.

Une fois le contenu choisi, les images d'essai sont préparées de manière à refléter les différentes configurations à l'étude ou la/les gamme(s) d'un ou de plusieurs paramètre(s). Lorsqu'on veut évaluer deux paramètres ou plus, les images peuvent être préparées de deux façons. Dans la première variante, chaque image représente un niveau d'un seul paramètre. Dans la seconde, chaque image représente un niveau de tous les paramètres examinés, mais image après image, chaque niveau de chaque paramètre apparaît avec chaque niveau de tous les autres paramètres. Les deux méthodes permettent de connaître précisément les résultats pour chaque paramètre. La dernière méthode permet également de déceler les interactions entre les différents paramètres (c'est-à-dire les effets non additifs).

4.1.3 Conditions d'observation

On a remarqué que, livrés à eux-mêmes, les observateurs choisiraient des distances d'observation supérieures à celles des évaluations subjectives. Il faut poursuivre les études du rapport entre les distances préférées et celles qu'on utilise lors des évaluations subjectives.

4.1.4 Séance d'évaluation

Avant la séance d'évaluation, on explique aux observateurs ce que l'on attend d'eux en leur présentant, en général, des exemples d'images ou de séquences d'images. Les instructions sont normalement données sous forme écrite ou d'enregistrement. On prend soin de ne pas influencer les observateurs dans l'exécution de leur tâche.

La séance comporte une série de présentations qui seront présentées dans un ordre aléatoire et, de préférence, dans un ordre différent pour chaque observateur. Lorsqu'on utilise une séquence aléatoire unique, l'expérimentateur veille habituellement à ce que la même image ne soit pas présentée deux fois de suite avec le même niveau et le même type de dégradation.

Une présentation type comprend trois visualisations: une image d'adaptation gris moyen, une image de stimulus et de nouveau une image gris moyen. La durée de ces visualisations varie selon la tâche de l'observateur, le matériel (images fixes/images animées) et les options ou les paramètres considérés; 3, 10 et 10 s respectivement sont des durées courantes pour ces visualisations. L'avis ou les avis de l'observateur peuvent être recueillis pendant la visualisation de l'image de stimulus ou de la seconde image gris moyen.

4.1.5 Types de méthodes à un seul stimulus

Trois types de méthodes à un seul stimulus ont été généralement utilisés pour évaluer les systèmes de télévision.

4.1.5.1 Méthodes utilisant une échelle d'évaluation par catégorie

Dans ce cas, les observateurs attribuent à une image ou une séquence d'images une catégorie choisie parmi un ensemble de catégories définies d'un point de vue sémantique. Les catégories peuvent traduire la présence ou l'absence d'un attribut, par exemple, pour établir le seuil de dégradation. Les échelles par catégories permettant d'évaluer la qualité de l'image et la dégradation de l'image, ont été utilisées dans la plupart des cas; les échelles du CCIR sont indiquées au Tableau 4 ci-après. Dans la surveillance de l'exploitation, on utilise parfois des demi-notes. Des échelles permettant d'évaluer la lisibilité du texte, l'effort de lecture et l'utilité de l'image ont été utilisées dans des cas particuliers.

Cette méthode aboutit, pour chaque condition, à une distribution des évaluations selon les catégories de l'échelle. La façon dont les réponses sont analysées dépend du jugement (détection, etc.) et de l'information recherchée (seuil de détection, rangs ou tendance moyenne des conditions, «distances» psychologiques entre les différentes conditions). Un grand nombre de méthodes d'analyse sont disponibles.

TABLEAU 4

Echelles de qualité et de dégradation du CCIR

Echelle à cinq notes	
Qualité	Dégradation
5 Excellent	5 Imperceptible
4 Bon	4 Perceptible mais non gênant
3 Assez bon	3 Légèrement gênant
2 Médiocre	2 Gênant
1 Mauvais	1 Très gênant

4.1.5.2 Méthodes n'utilisant pas une échelle d'évaluation par catégorie

Dans ce cas, les observateurs attribuent une valeur à chaque image ou séquence d'images présentée. Cette méthode a deux variantes.

Dans le cas d'une échelle continue, qui constitue une variante de la méthode par catégorie, l'observateur attribue à chaque image ou chaque séquence d'images à un point situé sur une ligne tracée entre deux qualificatifs sémantiques (par exemple, les extrémités d'une échelle par catégorie comme au Tableau 4). Pour référence, l'échelle peut comporter d'autres qualificatifs, situés en des points intermédiaires. La distance qui sépare une extrémité de l'échelle sert d'indice pour chaque condition.

Dans le cas d'une échelle discrète, l'observateur attribue à chaque image ou séquence d'images une note qui reflète, pour un paramètre spécifique, le niveau de la qualité de l'image tel qu'il l'a apprécié (par exemple, la netteté de l'image). La gamme de notes utilisées peut être restreinte (par exemple, 0-100) ou non. Parfois, la note attribuée reflète le niveau apprécié en termes «absolus» (sans référence directe au niveau de qualité d'une quelconque autre image ou séquence d'images comme dans certaines formes de la méthode d'estimation des grandeurs). Dans d'autres cas, la note traduit le niveau apprécié par rapport au niveau considéré précédemment comme «type» (par exemple, méthode d'estimation des grandeurs, fractionnement et estimation par la méthode utilisant une échelle de rapport).

Dans un cas comme dans l'autre, on aboutit à une distribution des notes pour chaque condition d'essai. La méthode d'analyse utilisée dépend du type de jugement et de l'information requise (par exemple, rangs, tendance centrale, «distances» psychologiques).

4.1.5.3 Mesures de la performance

Certains aspects des conditions normales d'observation peuvent être évalués en termes de «performance» des tâches purement externes (informations ciblées, lecture d'un texte, identification d'objets, etc.). Ainsi, une mesure de la performance portant par exemple sur la précision ou la rapidité avec laquelle ces tâches sont exécutées peut servir d'indice de l'image ou de la séquence d'images.

Les mesures de la performance conduisent à une distribution des notes de précision ou de rapidité pour chaque condition. L'analyse s'attache avant tout à établir les relations entre les conditions dans la tendance centrale (et dispersion) des notes et utilise souvent l'analyse de variance ou une technique analogue.

4.1.6 Problèmes

4.1.6.1 Gamme de conditions et ancrage

Etant donné que la méthode utilisant une échelle d'évaluation par catégorie et certaines méthodes n'utilisant pas ce type d'échelle sont sensibles aux variations de la gamme et de la distribution des conditions observées, les séances d'évaluation subjective doivent inclure les gammes complètes de variation des facteurs. On peut atteindre toutefois plus ou moins le même objectif avec une gamme plus restreinte en présentant également certaines conditions qui se situeront aux extrémités des échelles. Elles peuvent être représentées comme exemples et identifiées comme étant les plus extrêmes (ancrage direct) ou réparties tout au long de la séance et non identifiées comme étant les plus extrêmes (ancrage indirect).

4.1.6.2 Signification des notes

Etant donné qu'elles varient en fonction de la gamme, il peut ne pas être judicieux d'interpréter en termes absolus les évaluations obtenues avec la méthode utilisant une échelle par catégorie ou certaines méthodes n'utilisant pas ce type d'échelle (par exemple, la qualité d'une image ou d'une séquence d'images).

4.2 Méthodes comparatives avec stimulus

Dans ce type de méthodes, on présente deux images ou séquences d'images à l'observateur qui fournit un indice de la relation entre les deux présentations.

4.2.1 Observateurs

Les observateurs sont choisis de la même façon que dans les méthodes à un seul stimulus.

4.2.2 Images d'essai

Les images ou séquences d'images utilisées sont produites de la même façon que dans les méthodes à un seul stimulus. Les images ou séquences d'images ainsi obtenues sont ensuite combinées pour former les paires utilisées dans les essais d'évaluation.

4.2.3 Conditions d'observation

Les conditions d'observation sont déterminées de la même façon que dans les méthodes à un seul stimulus.

4.2.4 Séance d'évaluation

L'évaluation fera intervenir soit un seul moniteur d'évaluation soit deux bien synchronisés et se déroulera généralement comme dans le cas des méthodes à un seul stimulus. Si on utilise un seul moniteur d'évaluation, la présentation élémentaire comportera un stimulus supplémentaire identique en durée au premier. Dans ce cas, on fera bien de s'assurer au fil des essais, que les deux membres d'une paire apparaissent un même nombre de fois en première et en seconde position. Si on utilise deux moniteurs d'évaluation, les images de stimulus sont présentées simultanément.

4.2.5 Types de méthodes comparatives avec stimulus

Trois types de méthodes comparatives avec stimulus ont été utilisés pour évaluer des systèmes de télévision.

4.2.5.1 Méthodes utilisant une échelle d'évaluation par catégorie

Dans ce genre de méthode, les observateurs attribuent la relation entre les membres d'une paire à une catégorie choisie parmi un ensemble de catégories définies d'un point de vue sémantique. Ces catégories peuvent indiquer la présence de différences perceptibles (par exemple, IDENTIQUE, DIFFÉRENT), la présence et le degré de différences perceptibles (par exemple, MOINS, IDENTIQUE, PLUS) ou des appréciations de l'importance et du degré des différences. L'échelle comparative du CCIR est indiquée au Tableau 5 ci-dessous.

TABLEAU 5
Echelle de comparaison

-3	Beaucoup moins bon
-2	Moins bon
-1	Légèrement moins bon
0	Identique
+1	Légèrement mieux
+2	Mieux
+3	Beaucoup mieux

Cette méthode conduit, pour chaque paire de conditions, à une distribution des évaluations subjectives en fonction des catégories de l'échelle. La façon dont les réponses sont analysées dépend de l'appréciation (par exemple, différence) et de l'information requise (par exemple, différences juste perceptibles, rang des conditions, «distances» entre les conditions, etc.).

4.2.5.2 Méthodes n'utilisant pas une échelle d'évaluation par catégorie

Dans ce genre de méthode, les observateurs attribuent une valeur à la relation entre les membres d'une paire d'évaluations subjectives. Cette méthode présente deux variantes.

- Dans le cas d'une échelle continue, l'observateur attribue à chaque relation un point situé sur une ligne tracée entre deux qualificatifs (par exemple, IDENTIQUE-DIFFÉRENT ou les extrémités d'une échelle par catégorie comme dans le Tableau 5). Les échelles peuvent comporter d'autres qualificatifs de référence situés en des points intermédiaires. La distance qui sépare le point de l'extrémité de la ligne sert de référence pour chaque paire de conditions.
- Dans la seconde variante, l'observateur attribue à chaque relation une note qui reflète le niveau de l'image tel qu'il l'a perçu, cela pour un paramètre précis (par exemple, la différence de qualité). La gamme des notes utilisées peut être limitée ou non. La note attribuée peut décrire la relation en termes «absolus» ou en termes d'une paire «type».

Dans les deux cas, on obtient une distribution des valeurs pour chaque paire de conditions. La méthode d'analyse dépend de la nature de l'appréciation portée et de l'information requise.

4.2.6 Mesures de la performance

Dans certains cas, les mesures de la performance peuvent être obtenues à partir de méthodes comparatives avec stimulus. Dans la méthode du choix forcé, chaque paire d'images est préparée de telle sorte que l'une des images présente un niveau spécifique d'un attribut (par exemple, dégradation) alors que l'autre présente un niveau différent de ce même attribut ou ne présente pas cet attribut. L'observateur est prié d'indiquer l'image qui présente le niveau le plus élevé/le moins élevé de l'attribut ou l'image qui ne présente pas l'attribut; la précision et la rapidité de la performance servent à mesurer la relation entre les membres de la paire.

4.2.7 Problèmes

4.2.7.1 Formation des paires

Les méthodes comparatives avec stimulus permettent d'évaluer plus complètement les relations existant entre les conditions lorsque les évaluations portent sur toutes les paires possibles de conditions. Toutefois, s'il faut un trop grand nombre d'observations, on peut répartir les observations entre les observateurs, ou utiliser un échantillon de toutes les paires possibles.

4.2.7.2 Méthodes utilisant une échelle multidimensionnelle

Plusieurs chercheurs ont fait appel aux méthodes utilisant une échelle multidimensionnelle pour étudier les évaluations des systèmes de télévision obtenues à partir de méthodes comparatives avec stimulus.

4.3 Choix des méthodes

Toutes les méthodes décrites jusqu'ici présentent des avantages et des inconvénients; il n'est pas encore possible d'en recommander une plutôt qu'une autre. Il incombe donc au chercheur de choisir la méthode qui convient le mieux aux conditions qui prévalent.

Les limites inhérentes aux différentes méthodes donnent à penser qu'il pourrait être déraisonnable de trop insister sur une seule méthode. Il semble donc plus judicieux d'envisager des approches plus «complètes», c'est-à-dire d'utiliser plusieurs méthodes, ou une approche multidimensionnelle.

APPENDICE 1
DE L'ANNEXE 1

Caractéristiques de défaillance fonction du contenu de l'image

1. Introduction

Une fois mis en œuvre, un système devra traiter une gamme potentiellement étendue de programmes et il risque d'infliger à certains d'eux une perte de qualité. Pour voir si un système convient, il faut connaître la proportion des programmes qui lui causent des difficultés et la perte de qualité qui en résulte. On a en effet besoin, pour le système considéré, d'une caractéristique de défaillance fonction du contenu de l'image.

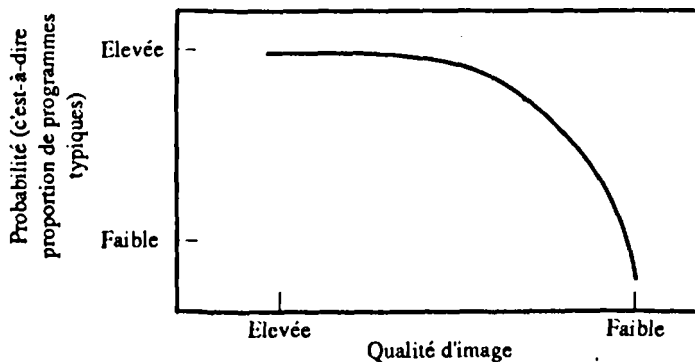
Cette caractéristique est particulièrement importante pour les systèmes dont la qualité ne se dégrade pas progressivement quand l'image devient de plus en plus critique. Par exemple, certains systèmes numériques et adaptatifs peuvent conserver une haute qualité sur toute une large gamme de types de programmes mais se détériorer au-delà.

2. Etablissement de la caractéristique de défaillance

Le concept de caractéristique fonction du contenu de l'image définit la proportion des programmes susceptibles de se présenter à long terme et pour lesquels le système donne un certain niveau de qualité. C'est ce qu'illustre la Fig. 5.

FIGURE 5

Représentation graphique d'un exemple de caractéristiques de défaillance fonction du contenu de l'image



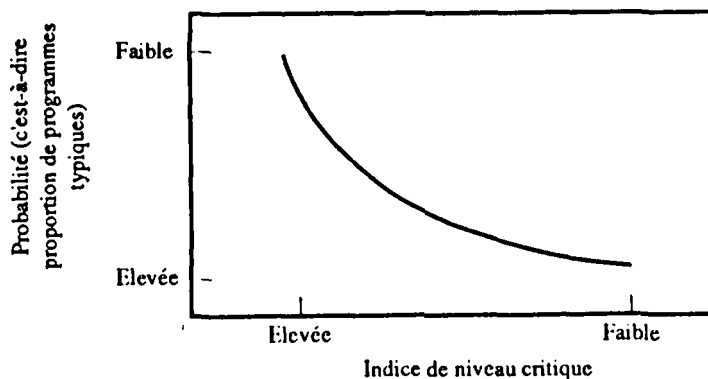
On peut obtenir, en quatre étapes, la caractéristique de défaillance fonction du contenu de l'image:

L'étape 1 détermine une mesure algorithmique du «niveau critique» qui doit pouvoir classer par ordre de mérite un certain nombre de séquences d'images auxquelles le système ou la catégorie de systèmes concernés ont infligé des distorsions, de sorte que le classement correspond à celui qu'auraient donné des observateurs humains chargés de cette tâche. La mesure du niveau critique peut prendre en compte une modélisation de la vue.

L'étape 2 établit, en appliquant la mesure du niveau critique à un grand nombre d'échantillons de programmes types de télévision, une distribution qui estime la probabilité que se présentent des images de niveaux critiques divers pour le système ou les catégories de systèmes considérés. La Fig. 6 présente un exemple de ce genre de distribution.

FIGURE 6

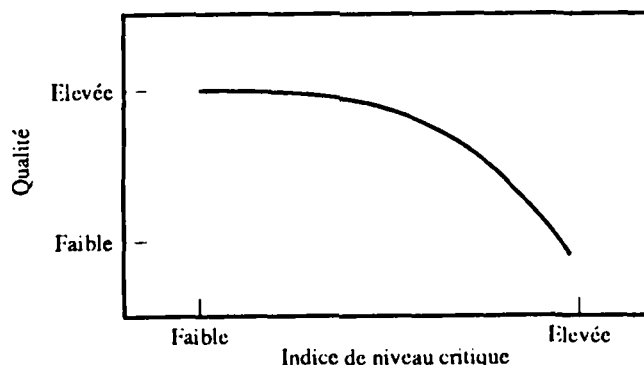
Probabilité que se présentent des Images
d'un certain niveau critique



L'étape 3 établit, de façon empirique, la faculté du système de conserver la qualité quand le niveau critique du programme augmente. Dans la pratique, il faut évaluer subjectivement la qualité que donne le système avec les programmes choisis pour échantillonner la gamme de niveaux critiques définie à l'étape 2. Il en résulte une fonction qui met en relation la qualité donnée par le système et le niveau critique du programme. La Fig. 7 donne un exemple de cette fonction.

FIGURE 7

Exemple de courbe de la qualité
en fonction du niveau critique du programme



L'étape 4 combine les résultats des étapes 2 et 3 pour établir une caractéristique de défaillance fonction du contenu de l'image, comme celle de la Fig. 5.

3. Utilisation de la caractéristique de défaillance

La caractéristique de défaillance, qui donne une idée globale de la qualité à espérer pour tous les types de programmes possibles, est un moyen essentiel pour étudier si un système convient. On peut s'en servir de trois façons:

- pour optimiser les caractéristiques d'un système (par exemple, la résolution à la source, le débit binaire, la largeur de bande) au moment de sa conception pour l'adapter au mieux aux exigences d'un service;
- étudier si un système donné conviendra (c'est-à-dire prévoir la conséquence et la gravité des défaillances en exploitation);
- voir, parmi plusieurs systèmes, ceux qui conviennent le mieux (c'est-à-dire comparer les caractéristiques de défaillance et déterminer quel système conviendra le mieux à l'usage envisagé). On notera que, si plusieurs systèmes possibles de type semblable peuvent avoir le même indice de niveau critique, des systèmes de types différents peuvent en avoir de distincts. Toutefois, bien que la caractéristique de défaillance n'exprime que la probabilité d'observer dans la pratique différents niveaux de qualité, on peut comparer directement les caractéristiques même si elles résultent d'indices de niveau critique distincts et propres au système.

Alors que la méthode décrite ci-dessus donne une façon de mesurer la caractéristique de défaillance fonction du contenu de l'image d'un système, il n'est pas sûr qu'elle puisse prévoir si un système sera acceptable pour un téléspectateur. On obtient cette information en faisant observer, par un certain nombre d'observateurs, des programmes codés selon le système en question et en examinant leurs commentaires.

APPENDICE 2 DE L'ANNEXE 1

Méthode de détermination d'une caractéristique de défaillance composite en fonction du contenu du programme et des conditions de transmission

1. Introduction

Une caractéristique de défaillance composite établit une relation entre la qualité de l'image perçue et la probabilité pratique de l'obtenir, en considérant explicitement le contenu du programme et les conditions de transmission.

On pourrait obtenir une telle caractéristique au moyen d'études subjectives avec un nombre suffisant d'observations, d'essais et de points de réception pour avoir un échantillon représentatif de la population des contenus de programmes possibles et des conditions de transmission. Dans la pratique, toutefois, une expérience semblable risque d'être irréalisable.

Le présent Appendice décrit une autre méthode, plus facile à mettre en œuvre, permettant de déterminer la caractéristique de défaillance composite. Cette méthode comprend trois étapes:

- analyse du contenu du programme;
- analyse de la voie de transmission; et
- établissement des caractéristiques de défaillance composite.

2. Analyse du contenu du programme

Cette étape comprend deux opérations. On définit d'abord une mesure appropriée du contenu du programme. Puis on évalue la façon dont se répartissent les probabilités des résultats de la mesure.

Une mesure du contenu du programme est une statistique qui révèle les aspects du contenu du programme, lesquels soulignent la faculté du ou des systèmes considérés de donner une reproduction du programme perçue comme étant fidèle. Il serait évidemment avantageux que cette mesure soit fondée sur un modèle de perception approprié. Toutefois, en l'absence d'un tel modèle, il peut suffire d'avoir une mesure qui rende compte de certains aspects de l'importance de la diversité spatiale, dans les trames ou images vidéo et entre elles, pourvu qu'elle présente une relation à peu près uniforme avec la qualité perçue de l'image. Il peut être nécessaire de recourir à des modes de mesure différents pour des systèmes (ou catégories de systèmes) qui ont des méthodes de représentation de l'image complètement différentes.

Une fois qu'un mode de mesure approprié a été choisi, il faut estimer avec quelle probabilité les valeurs statistiques possibles surviennent. Cela peut se faire de deux façons différentes:

- avec la méthode empirique, on analyse un échantillon de, par exemple, 200 segments de programme de 10 s, à un format de production qui, du point de vue de la résolution, de la fréquence d'image et du format d'image, convienne au(x) système(s) considéré(s). L'analyse de ces échantillons fournit les fréquences relatives d'apparition des valeurs statistiques que l'on prend comme estimations de la probabilité d'apparition dans la pratique; ou
- avec la méthode théorique, on estime les probabilités au moyen d'un modèle théorique. On notera que, bien que la méthode empirique soit préférée, il peut être nécessaire, dans certains cas, de recourir à la méthode théorique (par exemple, lorsqu'on n'a pas assez de renseignements sur le contenu du programme, notamment lorsque de nouvelles technologies de production apparaissent).

Les analyses ci-dessus aboutiront, pour les valeurs statistiques du contenu, à une distribution de probabilité (voir aussi l'Appendice 1). On les combinera avec les résultats de l'analyse des conditions de transmission pour préparer l'étape finale de la méthode.

3. Analyse de la voie de transmission

Cette étape comprend aussi deux opérations. On définit d'abord une mesure de la qualité de la voie de transmission. Puis on évalue la façon dont se répartissent les probabilités des résultats de la mesure.

Une mesure de la voie de transmission est une statistique qui révèle les aspects de la qualité de la voie, lesquels influencent la faculté du ou des systèmes considérés de donner une reproduction du programme perçue comme étant fidèle. Il serait évidemment avantageux que cette mesure soit fondée sur un modèle de perception approprié. Toutefois, en l'absence d'un tel modèle, il peut suffire d'avoir une mesure qui rende compte de certains aspects des contraintes qu'impose la voie, pourvu qu'elle présente une relation à peu près uniforme avec la qualité perçue de l'image. Il peut être nécessaire de recourir à des modes de mesure différents pour des systèmes (ou catégories de systèmes) qui ont des méthodes de codage de la voie complètement différentes.

Une fois qu'un mode de mesure approprié a été choisi, il faut estimer avec quelle probabilité les valeurs statistiques possibles surviennent. Cela peut se faire de deux façons différentes:

- avec la méthode empirique, la qualité de la voie est mesurée par exemple en 200 instants et points de réception choisis au hasard. L'analyse de ces échantillons fournit les fréquences relatives d'apparition des valeurs statistiques que l'on prend comme estimations de la probabilité d'apparition dans la pratique; ou
- avec la méthode théorique, on estime les probabilités au moyen d'un modèle théorique. On notera que, bien que la méthode empirique soit préférée, il peut être nécessaire, dans certains cas, de recourir à la méthode théorique (par exemple, lorsqu'on n'a pas assez de renseignements sur la qualité de la voie, notamment lorsque de nouvelles technologies de transmission apparaissent).

Les analyses ci-dessus aboutiront, pour les valeurs statistiques de la voie, à une distribution de probabilité. On les combinera avec les résultats de l'analyse du contenu du programme pour préparer l'étape finale de la méthode.

4. Etablissement des caractéristiques des défaillances composites

A cette étape, on procède à une expérimentation subjective au cours de laquelle le contenu du programme et les conditions de transmission varient tous deux selon les probabilités calculées au cours des deux premières étapes.

La méthode fondamentale mise en œuvre ici est la procédure à double stimulus utilisant une échelle de qualité continue et notamment la version à 10 s recommandée pour les séquences animées (voir l'Annexe 1, § 3). La référence y est une image de qualité studio au format approprié (par exemple, avec la résolution, la fréquence et le format d'image convenant au(x) système(s) considéré(s)). En revanche, au cours de l'essai, on présente la même image que celle qui serait reçue avec le ou les systèmes considérés dans des conditions de la voie choisies.

Le matériel d'essai et les conditions de la voie sont choisis d'après les probabilités établies au cours des deux premières étapes de la méthode. Parmi les segments de matériel d'essai, qui ont chacun été étudiés en vue de déterminer leur valeur essentielle selon la statistique du contenu, on a un ensemble de sélection. On prélève dans cet ensemble du matériel de façon qu'il couvre toute la gamme possible des valeurs statistiques et on en prend d'autant plus que le niveau est plus critique. On choisit de même les valeurs statistiques possibles pour la voie. Ensuite, ces deux causes de variation d'origine indépendante sont combinées de façon aléatoire pour former une combinaison de contenus et de conditions de la voie de probabilité donnée.

Les résultats de ces études, qui établissent une relation entre la qualité d'image perçue et sa probabilité d'apparition dans la pratique, servent ensuite à estimer si un système convient ou à comparer des systèmes selon qu'ils conviennent plus ou moins bien.

ANNEXE 2

**Mesures de la qualité des images en fonction des dégradations
qui résultent du codage numérique des signaux de télévision****1. Introduction**

L'Annexe 1 décrit les méthodes d'évaluation subjective de la qualité et des dégradations d'une image de télévision de définition normale et la Recommandation 710 celles de la TVHD. La présente Annexe est consacrée à l'application de ces méthodes aux codecs de télévision.

On a acquis récemment une grande expérience dans l'évaluation subjective des codecs de haute qualité pour la télévision en composantes 4:2:2 à 34, 45 et 140 Mbit/s. On a étudié dans ces expériences les performances des codecs, des points de vue de la qualité intrinsèque de l'image décodée, de la qualité après des post-traitements de studio (incrustations et ralenti) de l'image décodée et des dégradations dues à différents taux d'erreur binaires dans la voie. La présente Annexe est en partie basée sur ces expériences.

Pour les applications de distribution, les spécifications de la qualité peuvent être définies en termes de jugement subjectif de la part d'observateurs. De tels codecs peuvent théoriquement être évalués subjectivement en fonction des spécifications. En théorie, la qualité d'un codec prévue pour des applications de contribution ne pourrait cependant pas être définie en termes de caractéristiques subjectives car le signal qu'il produit n'est pas destiné au visionnage immédiat, mais à la postproduction, à l'enregistrement et/ou au codage pour transmission ultérieure. Du fait de la difficulté de définir la qualité pour une série d'opérations de postproduction différentes, le principe choisi a consisté à définir les performances d'une chaîne d'appareils, y compris pour la fonction de postproduction, considérée comme représentative d'une application de contribution réelle. Cette chaîne pourrait être constituée d'un codec, suivi par un dispositif de postproduction (ou par un autre codec dans le cas de l'évaluation de la qualité intrinsèque de la contribution), puis par un autre codec avant présentation du signal à l'observateur. L'adoption de cette stratégie pour les codecs destinés aux applications de contribution a pour conséquence que les procédures de mesure décrites dans la présente Recommandation peuvent aussi être utilisées pour les évaluer.

On insiste dans toute l'Annexe sur l'importance du choix de séquences d'essai critiques, surtout des scènes naturelles, et l'on donne quelques indications sur la manière dont ces séquences peuvent être produites ou choisies.

2. Evaluation subjective de la qualité d'image des codecs

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

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

2.1 Evaluation de la qualité intrinsèque

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

2.1.1 Conditions d'observation et choix des observateurs

Il est conseillé de respecter dans ce domaine les prescriptions figurant au § 2.4 de l'Annexe 1 pour la télévision de définition conventionnelle et dans la Recommandation 710 pour les codecs de télévision à haute définition.

2.1.2 Séquences d'images d'essai

Il est recommandé d'utiliser dans les essais au moins six séquences d'images, plus une supplémentaire destinée à une démonstration avant le début de l'expérience. La durée d'une séquence doit être d'environ 10 s, mais il convient de noter que les observateurs préféreront peut-être une durée de 15 à 30 s. Cette séquence doit être comprise entre «moyennement critique» et «critique» pour l'application à débit binaire réduit considérée.

2.1.3 Méthodologie des essais

Lorsque la plage de qualité à étudier est étroite, comme ce sera normalement le cas pour les codecs de télévision, la méthodologie à utiliser est celle du double stimulus avec échelle de qualité continue décrite dans le § 3 de l'Annexe 1. La séquence de source initiale est utilisée comme référence. Des discussions se poursuivent sur la durée des séquences de présentation. Dans de récentes expériences sur des codecs destinés à la vidéo en composantes 4:2:2, il a été jugé avantageux de modifier la présentation par rapport à celle qui est décrite dans la présente Recommandation. Des images composites ont été utilisées à titre de référence supplémentaire pour donner un niveau de qualité plus faible par rapport à celui auquel était estimée la qualité du codec.

2.2 Evaluation de la qualité après post-traitement

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

2.2.1 Conditions d'observation et choix des observateurs

Voir le § 2.1.1.

2.2.2 Séquences d'images d'essai

Du fait de contraintes matérielles ou de la nécessité d'évaluer un codec pour plusieurs opérations de post-traitement, le nombre des séquences d'images utilisées peut être au minimum de trois, avec une supplémentaire pour la démonstration. La nature des séquences dépendra du type de post-traitement étudié, mais elles devront être comprises entre «modérément critique» et «critique» pour la télévision à débit binaire réduit et pour l'opération en question. La durée des séquences devra être d'environ 10 s, mais il convient de noter que les observateurs préféreront peut-être une durée de 15 à 30 s. Pour évaluer le ralenti, une vitesse de reproduction égale au 1/10 de celle de la source conviendra sans doute.

2.2.3 Méthodologie des essais

La méthodologie à utiliser est celle du double stimulus à échelle de qualité continue. Dans ce cas, la référence sera cependant la séquence à la source soumise au même post-traitement que les images décodées. Si l'on juge avantageux de prévoir une référence de plus basse qualité, elle doit aussi être soumise au même post-traitement. Pour les essais effectués par le CCIR, on a légèrement modifié la présentation décrite dans la présente Recommandation.

3. Evaluation subjective de la dégradation de la qualité d'image des codecs du fait des erreurs de transmission

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

3.1 Séquences d'images d'essai

Comme l'on doit étudier la qualité du codec pour une série de taux d'erreur binaires de transmission, les contraintes matérielles conduisent à suggérer que trois séquences, plus une pour la démonstration, seraient peut-être suffisantes. La durée d'une séquence doit être d'environ 10 s, mais il convient de noter que les observateurs préféreront peut-être une durée de 15 à 30 s. Cette séquence doit être comprise entre «moyennement critique» et «critique» pour la télévision à débit binaire réduit.

3.2 Choix des taux d'erreur binaires

On doit choisir un minimum de cinq taux d'erreur binaires, mais de préférence davantage, répartis logarithmiquement et couvrant toute la gamme pour laquelle le codec donne des dégradations allant de «imperceptible» à «très gênant».

3.3 Méthodologie des essais

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

3.4 Note sur l'utilisation de très faibles taux d'erreur binaires

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

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

4. Comparaisons subjectives entre codecs

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

Telle qu'elle est décrite, la méthode permet des comparaisons précises et donne le moyen de déterminer la relation entre couples de systèmes. Une extension de cette méthode au classement de la qualité ou de la dégradation pour plus de deux systèmes est possible. Selon ce principe, on établit un classement global à partir de ceux donnés par les observateurs pour tous les couples possibles de séquences d'images.

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

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

5. Choix des images d'essai pour l'évaluation des codecs numériques

D'un bout à l'autre de la présente Annexe, on insiste sur l'importance qu'il y a de tester les codecs au moyen de séquences d'images critiques pour la réduction de débit binaire en télévision. On peut donc raisonnablement se demander quel est le niveau critique d'une séquence donnée pour une application déterminée de la réduction du débit binaire ou se préoccuper de savoir si une séquence est plus critique qu'une autre. Une réponse simple, mais pas particulièrement utile, est d'affirmer que la notion de niveau critique a une signification très différente selon les codecs. C'est ainsi qu'une image fixe très détaillée pourrait être critique pour un codec intratrame, tandis que pour un système intertrames capable d'exploiter les similitudes entre images, la même scène ne poserait aucun problème. Certains types de séquences où figurent des textures mouvantes et des mouvements complexes seront critiques pour toutes les catégories de codecs et ce sont donc eux qu'il est le plus utile de produire ou de reconnaître. Les mouvements complexes peuvent être d'une forme prévisible par l'observateur, mais non par les algorithmes de codage; c'est le cas des déplacements tortueux périodiques.

On a découvert une grandeur simple mais utile basée sur la mesure de l'entropie adaptative intertrames-intratrame en essayant de chiffrer statistiquement le niveau critique de l'image. Cette méthode a été utilisée pour «calibrer» les séquences d'images que le CCIR propose d'employer dans les essais de codecs à 34, 45 et 140 Mbit/s et s'est révélée utile dans le choix de celles à adopter. La manière la plus facile de réaliser ces mesures sur les séquences est de transférer celles-ci à des ordinateurs de traitement de l'image, puis de les soumettre à un logiciel d'analyse.

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

a) *Codecs intratrame à mots de longueur fixe*

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

b) *Codecs intertrames à mots de longueur fixe*

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

c) *Codecs intratrame à mots de longueur variable*

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

d) *Codecs intertrames à mots de longueur variable*

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

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

On a rassemblé une collection complète de séquences d'essai enregistrées au format DI en composantes 4:2:2. La Recommandation 802 donne des précisions à leur sujet et indique les critères (qui peuvent s'appliquer à d'autres normes de représentation d'image) sur la base desquels elles sont préparées.
