

RECOMMANDATION UIT-R BT.710-3

**MÉTHODES D'ÉVALUATION SUBJECTIVE DE LA QUALITÉ D'IMAGE
DE TÉLÉVISION À HAUTE DÉFINITION**

(Question UIT-R 211/11)

(1990-1992-1994-1997)

L'Assemblée des radiocommunications de l'UIT,

considérant

- a) qu'un certain nombre d'administrations et d'organisations dans le monde évaluent actuellement des systèmes de télévision à haute définition (TVHD) et que dans de nombreuses régions du monde la radiodiffusion TVHD va vraisemblablement devenir le premier média du XXI^e siècle;
- b) que les évaluations subjectives sont vitales pour la conception et le choix de systèmes de TVHD;
- c) que la Recommandation UIT-R BT.500 décrit des méthodes subjectives générales, dont de nombreux détails méthodologiques sont également adaptés au contexte de la TVHD;
- d) que, cependant, il peut sembler utile de clarifier les méthodes d'évaluation et les conditions d'observation appropriées à la TVHD, dans les domaines clés actuellement à l'étude, au moyen d'une Recommandation particulière,

recommande

- 1** que les évaluations subjectives de la qualité des images des systèmes de TVHD, soient effectuées avec les conditions d'observation indiquées dans l'Annexe 1;
- 2** que l'évaluation subjective de la qualité globale des images de TVHD fournies par un système d'émission soit effectuée en utilisant la méthode à double stimulus utilisant une échelle continue de qualité (Recommandation UIT-R BT.500) avec la norme studio TVHD comme référence;
- 3** que l'évaluation des caractéristiques de défaillance d'un système d'émission analogique TVHD en cas de détérioration de la transmission soit effectuée au moyen de la méthode à double stimulus utilisant une échelle de dégradation (Recommandation UIT-R BT.500) avec, comme référence, soit l'image du studio TVHD, soit l'image émission non dégradée;
- 4** qu'il convient d'utiliser la méthode exposée dans la Recommandation UIT-R BT.1129 pour calculer les caractéristiques de défaillance en fonction du contenu d'image d'un système de TVHD numérique;
- 5** qu'en cas d'absence de référence de haute qualité, la méthode utilisant une échelle graphique ou la méthode utilisant une échelle d'évaluation des rapports soit retenue pour l'évaluation de la qualité globale des images fournies par un système de studio de TVHD (avant ou après traitement);
- 6** que, si une référence de haute qualité est disponible, la méthode à double stimulus utilisant une échelle continue de qualité (Recommandation UIT-R BT.500) soit retenue pour l'évaluation de la qualité globale des images fournies par un système de studio de TVHD (avant ou après traitement);
- 7** que, lorsqu'il est question de la variation de la qualité avec le contenu du programme ou des conditions de transmission rencontrées dans la pratique, l'on tienne compte de la description de la fonction caractéristique de dégradation composite que donne l'Appendice 2 de l'Annexe 1 à la Recommandation UIT-R BT.500;
- 8** que, pour l'interprétation de résultats d'études particulières, il soit dûment tenu compte de toute limitation effective que peuvent imposer les techniques courantes sur les résultats de l'étude (par exemple, les effets de bord des équipements de prise de vue ou de présentation d'images);
- 9** que l'on prenne soin de mettre en évidence l'influence du format de présentation des images quand il n'est pas celui du système de base (par exemple, suite à conversion vers le haut). Des évaluations pourront être faites pour tenir compte des différents formats, s'il y a lieu.

NOTE 1 – Des informations sur les méthodes d'évaluation subjective de la qualité de l'image dans un environnement TVHD sont données dans l'Annexe 2.

NOTE 2 – Des informations sur les facteurs d'évaluation propres à l'évaluation globale de la TVHD sont données dans l'Annexe 3.

ANNEXE 1

TABLEAU 1

Conditions d'observation pour l'évaluation subjective de la qualité des images de TVHD

Condition	Sujet	Valeurs ⁽¹⁾
a	Rapport de la distance d'observation sur la hauteur de l'image	3
b	Luminance maximale de l'écran (cd/m ²) ⁽²⁾	150-250
c	Rapport de la luminance du tube image inactif (faisceaux coupés) sur la luminance maximale ⁽³⁾	≤ 0,02
d	Rapport de la luminance de l'écran affichant seulement le niveau du noir dans une salle complètement noire à celle du blanc maximum ⁽⁴⁾	environ 0,01
e	Rapport de la luminance de l'arrière-plan derrière le moniteur image à la luminance maximale de l'écran	environ 0,15
f	Eclairement de la salle dû à d'autres sources ⁽⁵⁾	faible
g	Chromaticité de l'arrière-plan	D_{65}
h	Angle sous-tendu par la zone d'arrière-plan satisfaisant les conditions ci-dessus ⁽⁶⁾ . Cela doit être respecté pour tous les observateurs	53° (hauteur) × 83° (largeur)
i	Placement des observateurs	A l'intérieur d'un angle horizontal de ± 30° dont le sommet est le centre de l'écran. La limite verticale est à l'étude
j	Dimensions de l'écran ⁽⁷⁾	1,4 m (55 pouces)

- (1) Sachant qu'il peut ne pas être possible de réaliser toutes ces conditions pour les tests, des valeurs de remplacement sont données à titre temporaire. Il doit être mentionné, cependant, que les résultats des tests effectués dans ces conditions provisoires ne seront en général pas comparables avec ceux obtenus dans des situations pour lesquelles les conditions nominales s'appliquent.
- (2) La luminance maximale de l'écran correspond à un signal vidéo ayant une amplitude de 100%. Des valeurs ≥ 70 cd/m² pourront être utilisées jusqu'à ce que ce niveau spécifié soit techniquement réalisable.
- (3) Il est possible que cette caractéristique soit influencée par l'éclairement de la salle et par la gamme de contraste sur l'écran.
- (4) Le niveau du noir correspond à un signal ayant une amplitude de 0%.
- (5) L'éclairement de la salle doit être ajusté de sorte à rendre possible les conditions c et e.
- (6) Un minimum de 28° (hauteur) × 48° (largeur) est recommandé.
- (7) Des valeurs ≥ 76,2 cm (30 pouces) pourront être utilisées si des écrans à la dimension spécifiée ne sont pas disponibles.

ANNEXE 2

Méthodes d'évaluation subjective de la qualité de l'image dans un environnement TVHD

1 Introduction

Les méthodes utilisées dans les essais subjectifs des systèmes classiques de télévision sont décrites dans la Recommandation UIT-R BT.500. Les principaux concepts de la méthodologie d'évaluation s'appliquent également à toutes les formes de télévision mais il faut étudier avec soin comment appliquer à la TVHD les spécifications détaillées des méthodes destinées à la télévision classique.

L'UIT-R examine les progrès accomplis en matière de TVHD pour déterminer les modifications qu'il sera nécessaire, le cas échéant, d'apporter aux méthodes d'essais subjectifs pour tenir compte de ces nouveaux développements. Dans ce domaine, les études ne sont pas encore terminées.

Avant d'entrer dans le vif du sujet, il est important de souligner les points suivants:

- La qualité de l'image n'est pas le seul facteur qui doit être pris en considération dans le choix des normes. D'autres facteurs, tels que la complexité du système, la disponibilité, les possibilités futures, etc., doivent être incorporés dans l'équation globale;
- Les résultats des expériences d'évaluation subjective ne sont pas en soi des lois physiques. Ils offrent simplement des directives pour un ensemble donné de conditions d'essai et ne constituent pas des faits absolus concernant un système ou un autre;
- Les différences conceptuelles entre les appréciations des échelles de qualité et de dégradation actuellement utilisées ne sont pas uniformes; mais l'usage veut que, lors des traitements des résultats, on se base sur l'hypothèse selon laquelle ces différences sont uniformes. Des études portant sur d'autres méthodes d'évaluation qui présentent moins d'inconvénients sont actuellement en cours mais l'interprétation des résultats des méthodes actuelles doit tenir compte des insuffisances;
- L'élément clé des évaluations subjectives est souvent le choix du matériel d'essai. Selon les directives, il faut utiliser un matériel «critique» sans toutefois exagérer dans ce sens. Pour déterminer les cas qui entrent dans cette catégorie, il faut une parfaite compréhension de la manière dont fonctionnent les systèmes de TVHD.

2 Evaluation de la qualité de l'image dans un environnement TVHD

2.1 Eléments d'évaluation de la qualité de l'image

2.1.1 Evaluation des formats de studio de TVHD

Il y aura lieu d'évaluer:

- la qualité fondamentale de l'image,
- la qualité de l'image après un traitement en aval, à savoir l'incrustation-couleur, le ralenti et la manipulation des images ainsi que les conversions possibles vers d'autres formats, y compris le film.

2.1.2 Evaluation des formats de studio classiques (et des formats de film) obtenus à partir de sources de studio de TVHD

Il y aura lieu d'évaluer si, en termes de qualité d'image, les formats de studio classiques et de film obtenus à partir de sources de studio de TVHD sont adéquats.

2.1.3 Evaluation des formats d'émission de TVHD

Il y a lieu d'évaluer:

- la qualité fondamentale de l'image,
- les fonctions caractéristiques de dégradation,
- les fonctions caractéristiques d'écho, et
- la sensibilité au brouillage.

2.1.4 Evaluation des images de télévision classique insérées dans des émissions de TVHD

Certains des formats d'émission de TVHD actuellement étudiés comprennent un format de télévision classique («compatibilité vers l'arrière»). Il y aura donc lieu d'évaluer si, en termes de qualité d'image, les images de télévision classique insérées dans des émissions de TVHD sont adéquates.

2.2 Questions relatives à l'évaluation de la qualité de l'image

2.2.1 Méthodes d'évaluation

2.2.1.1 Evaluation de la qualité de l'image

Les appréciations de qualité à cinq notes actuellement utilisées dans les évaluations subjectives ne sont pas uniformément espacées du point de vue conceptuel et des difficultés ont été constatées dans la comparaison des résultats obtenus dans différents laboratoires, en particulier lorsque la traduction des appréciations était nécessaire. Par ailleurs, en raison de la sensibilité des évaluations de qualité (à l'aide de termes d'appréciation conceptuels) par rapport à la gamme de conditions utilisées dans l'essai, il n'est pas indiqué d'interpréter les termes de manière absolue ou de comparer des résultats provenant d'essais conduits selon différentes gammes de qualité (par exemple, la TVHD et la télévision classique).

Une échelle de qualité à sept notes a été utilisée avec succès pour établir la signification de la qualité de la TVHD et de telles techniques peuvent se révéler utiles à l'avenir. En outre, des solutions de remplacement à la méthode de qualité à cinq notes sont présentées au § 6 de la Recommandation UIT-R BT.500. Néanmoins, après examen de ces solutions, il est suggéré d'utiliser, d'une manière générale, pour les évaluations de qualité dans un environnement de TVHD, la méthode de la qualité continue à double stimulus décrite dans la Recommandation UIT-R BT.500.

2.2.1.2 Evaluation des dégradations de l'image

Dans une certaine mesure, les mêmes problèmes ont été constatés dans l'échelle de dégradation à cinq notes que dans l'échelle de qualité à cinq notes. Cependant, il est recommandé que, pour l'évaluation des dégradations de la qualité de l'image, on utilise d'une manière générale la méthode à double stimulus utilisant une échelle de dégradation décrite dans la Recommandation UIT-R BT.500.

2.2.2 Conditions d'observation pour l'évaluation subjective dans un environnement de TVHD

2.2.2.1 Evaluation des formats de studio de TVHD

Le Rapport UIT-R BT.801 définit les objectifs de présentation de l'image pour des formats de studio de TVHD.

2.2.2.2 Evaluation des formats de studio classiques obtenus à partir de sources de studio de TVHD

Comme ces évaluations concernent des systèmes de télévision déjà étudiés dans les textes de l'UIT-R, l'évaluation des formats de studio classiques doit utiliser les conditions d'observation déjà approuvées et présentées dans la Recommandation UIT-R BT.500.

2.2.2.3 Evaluations des formats d'émission de TVHD

Il n'est pas établi clairement si les objectifs de présentation de l'image pour les images de studio de TVHD correspondent bien à des conditions d'observation susceptibles d'être rencontrées dans les foyers. Cependant, les évaluations subjectives des formats d'émission de TVHD doivent tenir compte, dans une certaine mesure, des performances supérieures du studio de TVHD.

Il est probable, en raison de contraintes imposées à l'émission, que les formats d'émission de TVHD ne soient pas en mesure de reproduire entièrement le niveau de qualité de l'image que peut atteindre un studio de TVHD. Cependant, compte tenu du fait que les formats d'émission tentent de reproduire aussi fidèlement que possible l'image de studio originale et afin de préserver la cohérence des essais subjectifs dans toute la chaîne d'émission de studio de TVHD, il est suggéré que les conditions d'observation données dans l'Annexe 1 soient également utilisées pour des essais de formats d'émission de TVHD ainsi que pour des essais de formats de studio de TVHD.

2.2.2.4 Evaluation d'images de télévision classique insérées dans des émissions de TVHD

Etant donné qu'il s'agit d'images de télévision classique, les conditions d'observation spécifiées dans les Recommandations UIT-R BT.500 et UIT-R BT.1129 sont applicables.

3 Evaluation de la qualité de l'image du format de studio de TVHD

3.1 Evaluation de la qualité fondamentale de l'image

La question qui se pose ici est celle de la qualité des images du format de studio de TVHD avant leur post-traitement. Les facteurs susceptibles d'influer sur la qualité fondamentale de l'image comprennent, entre autres, la résolution spatiale, la résolution temporelle, la palette de couleurs et les caractéristiques de linéarité. L'Annexe 3 résume les travaux relatifs aux facteurs d'évaluation de la qualité de l'image de TVHD.

Il est généralement admis qu'une augmentation de la palette de couleurs et l'insertion d'un codage de luminance constante constituent des objectifs souhaitables pour le système de studio de TVHD. Toutefois, ces objectifs exigent un traitement du signal plus complexe dans la caméra et dans l'équipement d'affichage; il peut donc devenir nécessaire d'établir un compromis entre les avantages de ces objectifs et les inconvénients résultant de la complexité du traitement du signal.

Pour estimer la valeur de la palette de couleurs élargie et l'incidence du traitement supplémentaire du signal, il faut disposer d'un équipement de visualisation ayant une palette sensiblement plus étendue que les équipements de visualisation à tube cathodique actuels et un signal source convenablement traité pour cet affichage d'une palette de couleurs plus étendue. De plus, il faut disposer d'un équipement de visualisation à tube cathodique avec le traitement non linéaire pour transformer le signal source à palette élargie en un signal approprié pour l'équipement de visualisation à palette plus réduite. Des images fixes contenant une gamme de couleurs normale plus quelques couleurs en dehors de la palette réduite devraient être évaluées par comparaison entre les deux visualisations. Pour les couleurs très pures, une source possible est fournie par des boules de fil avec des couleurs saturées. Les matériaux de ce type peuvent donner des couleurs très saturées qui se situent en dehors de la gamme des équipements de visualisation à tube cathodique actuels tout en se prêtant assez bien à une composition de scène. L'évaluation subjective d'une telle scène sur un équipement de visualisation normal à tube cathodique et sur des équipements de visualisation de grande pureté devrait fournir une estimation du gain de qualité.

L'évaluation de méthodes de codage de luminance constante par rapport aux méthodes de codage de luminance non constante devrait se faire par comparaison entre un équipement de visualisation *RGB* pleine bande et un équipement sur lequel on pourrait présenter des signaux à luminance constante ou des signaux à luminance non constante. Le sujet de la scène devrait comprendre des détails apparaissant en couleurs saturées aussi bien que des éléments d'une scène normale. Les ombres formées sur les boules de fils rouges, verts et bleus saturés constituent un moyen permettant de faire apparaître les détails en couleurs saturées.

Les méthodes normalement utilisées pour évaluer la qualité de l'image (c'est-à-dire les méthodes à double stimulus) exigent généralement une condition de référence qui fournit une qualité supérieure à celle du système à l'essai. En raison de la haute qualité d'un système de studio de TVHD, il est toutefois difficile de réunir les conditions de référence appropriées. C'est pourquoi il peut être opportun d'utiliser des scènes observées directement (fixes et animées) pour créer les conditions de référence nécessaires à l'évaluation des systèmes de studio de TVHD.

3.1.1 Méthodologie

La méthode à double stimulus utilisant une échelle de qualité continue pourrait être employée. La référence utilisable pour les évaluations de qualité de l'image pourrait être la scène observée directement (sous réserve d'un cadrage approprié). L'essai pourrait porter sur la même scène observée par l'intermédiaire du système à l'essai.

3.1.2 Conditions d'observation

Voir l'Annexe 1.

3.1.3 Matériel d'évaluation

Le matériel d'essai pourrait comprendre un certain nombre d'images fixes et de séquences animées. Les sources d'images fixes pourraient être soit des diapositives (rétroprojection) ou des photographies (projetées directement). Pour les séquences animées, les sources pourraient être des dioramas. La condition de référence serait réalisée avec une source observée directement alors que la condition testée serait établie lorsque la même source est observée par l'intermédiaire d'une caméra et d'un moniteur. On maintiendrait un cadrage identique pour les deux conditions en utilisant la réflexion du matériel d'essai correspondant aux deux conditions sur le même miroir d'observation de format 16:9. On pourrait passer d'une condition à l'autre au moyen de volets de fermeture sur les trajets optiques. Le passage devrait se faire sous le contrôle d'un expérimentateur.

Les essais comprennent des comparaisons du matériel d'essai pris par une caméra vidéo avec le même matériel observé directement. Pour réduire au minimum les biais éventuels dus aux différences inhérentes à la télévision par rapport au «monde réel», il sera nécessaire de contrôler un certain nombre de facteurs qui sont notamment les suivants:

- *différences de parallaxe*: pendant l'essai, l'observateur ne devrait pas se déplacer de manière significative car cela entraînerait un certain degré de parallaxe dû au mouvement dans la scène observée directement mais non dans celle qui apparaît sur le moniteur;
- *profondeur visible*: le miroir d'observation affichera alternativement l'image de télévision et la scène source. La composition et l'éclairage des scènes sources devraient être réglés de manière à réduire au minimum les différences de profondeur entre l'image de télévision et la scène observée directement;
- *éclairage de la scène*: le miroir d'observation affichera alternativement l'image de télévision et la scène source. L'éclairage de la scène source devra être ajusté lorsqu'on modifie le trajet d'affichage pour maintenir l'intensité et la température de couleur (D_{65}) à un niveau constant pour les deux images. La température de couleur peut devoir être réglée scène par scène.

On a suggéré un certain nombre de critères pour la composition des scènes sources. Ceux-ci comprennent:

- la résolution spatiale statique,
- la résolution spatiale dynamique,
- le rendu de luminance,
- le rendu des couleurs,
- le rendu de mouvement.

En outre, il pourrait être utile de compléter cette liste avec d'autres scènes à des fins spéciales. Celles-ci permettraient d'évaluer:

- les effets de profondeur apparente (par exemple, dans les scènes panoramiques),
- le rendu de tons familiers (par exemple, les couleurs de la peau),
- le sentiment de présence (par exemple, dans un mouvement panoramique), et
- la performance de papillotement (par exemple, avec de grandes plages secondaires blanches).

Il est important d'établir l'ensemble standard de séquences d'essai à utiliser pour diverses évaluations subjectives de la qualité d'image de la TVHD, comme cela a été fait pour les séquences 4:2:2 (voir la Recommandation UIT-R BT.1210).

3.1.4 Interprétation des résultats

Le système à l'essai devrait s'approcher d'aussi près que possible du niveau de qualité fourni par la référence observée directement. En examinant les résultats, il convient de tenir compte de deux points:

- Un système de studio de TVHD résultera vraisemblablement d'un compromis entre les diverses caractéristiques qui se rapportent à la qualité. Outre l'examen de la qualité sur la base d'une moyenne des divers éléments du matériel d'essai, il serait judicieux d'étudier séparément les réactions aux différentes scènes afin d'identifier les caractéristiques qui pourraient être améliorées;
- Pour l'interprétation des résultats, il est nécessaire de déterminer, et si possible, de corriger les biais éventuels des résultats dus aux différences de maturité technique des équipements.

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

Deux éléments sont pris en considération: le traitement postproduction et la conversion des normes.

3.2.1 Traitement postproduction

Les principaux éléments du traitement de postproduction sont l'incrustation couleur, le ralenti et la manipulation d'image. Des évaluations effectuées au moment où les normes 4:2:2 de la Recommandation UIT-R BT.601 ont été établies tendent à montrer que l'incrustation couleur est l'opération de postproduction la plus exigeante. Pour une fréquence de trame et un système de balayage donnés, cela s'appliquera vraisemblablement à la TVHD.

3.2.1.1 Evaluation de l'incrustation couleur

3.2.1.1.1 Méthodologie

La méthode avec échelle de dégradation à double stimulus devrait être utilisée à condition de disposer d'une gamme complète de différentes qualités d'image. La référence pour l'évaluation pourrait être une image en incrustation utilisant un signal *RGB* à largeur de bande complète, en tant qu'image de premier plan. L'essai d'incrustation et les images de référence doivent avoir une qualité optimisée image par image, comme cela serait le cas dans la pratique. On étudie actuellement la méthodologie qu'il convient d'appliquer si une gamme complète de qualités d'image n'est pas disponible.

3.2.1.1.2 Conditions d'observation

Voir l'Annexe 1.

3.2.1.1.3 Matériel d'évaluation

Le matériel d'essai doit représenter des cas critiques pour les types de dégradation susceptibles de faire l'objet d'un traitement d'incrustation couleur. Le matériel qui sera vraisemblablement le plus exigeant devra contenir des détails fins animés. Pour la TVHD, l'on ne connaît actuellement aucune séquence d'essai spécifique pour l'incrustation couleur mais il pourrait être approprié d'utiliser des peignes en mouvement, des rubans entrelacés et du verre (transparent). Toutefois une image fixe est disponible. La qualité de l'incrustation dépend fortement de l'éclairage de la scène et il convient de veiller tout particulièrement à garantir que cet éclairage soit optimal et stable (voir la Recommandation UIT-R BT.1210).

3.2.1.1.4 Interprétation des résultats

Le matériel d'essai ne doit pas être notablement dégradé par rapport au matériel de référence.

3.2.1.2 Evaluation du ralenti et de la manipulation d'image

3.2.1.2.1 Méthodologie, conditions d'observation, matériel d'évaluation, interprétation des résultats

Les évaluations de cette catégorie posent des problèmes liés au fait qu'il est improbable que l'on puisse disposer à l'heure actuelle d'un signal de référence de haute qualité. C'est l'introduction d'un signal de référence qui donne aux méthodes à double stimulus leurs propriétés utiles. On étudie actuellement une méthode à échelle proportionnelle qui pourrait se révéler suffisamment stable et reproductible sans le recours à une référence. Par ailleurs, il peut y avoir, dans certains cas, un moyen de produire des séquences de référence de haute qualité. Par exemple, le ralenti de haute qualité peut être réalisé si l'on effectue séparément le tournage de la séquence source à fréquence d'image supérieure.

3.2.2 Qualité de l'image après conversion de normes TVHD vers TVHD

3.2.2.1 Méthodologie

L'objectif déclaré de toutes les administrations est de parvenir à une norme mondiale unique de studio de TVHD pour permettre, entre autres, l'échange international de programmes sans nécessité de conversion de normes. Cependant, il se présentera sans doute des cas où une conversion sera nécessaire à partir d'autres formats ou de films de TVHD. En outre, des conversions similaires pourraient être nécessaires avant la production de formats d'émission ayant une fréquence de trame différente de celle de la source. En pareil cas, il convient d'étudier le format d'émission approprié à ces types de conversions.

Une conversion de norme de fréquence de trame peut faire apparaître des défauts temporels fugitifs et, dans ce cas, une méthode d'évaluation en deux étapes est proposée pour permettre une meilleure évaluation globale du système.

3.2.2.1.1 Evaluations principales

Il s'agit des évaluations qui sont considérées comme étant les évaluations les plus utiles. Une méthode d'échelle de qualité continue à double stimulus devrait être utilisée. Le signal de référence devrait être théoriquement la même image ou la même séquence utilisée comme entrée dans le convertisseur de normes mais ayant été tournée cette fois en utilisant les paramètres de balayage du signal de sortie du convertisseur. Si cela n'est pas possible ou si de tels essais devaient se révéler utiles pour d'autres raisons, le signal de référence devrait être le signal d'entrée dans le convertisseur.

3.2.2.1.2 Evaluations auxiliaires

Au moyen de la méthode à un seul stimulus (voir la Recommandation UIT-R BT.500), on demande à un certain nombre d'observateurs experts d'attribuer une note de qualité globale à plusieurs programmes convertis représentatifs. Il est également possible d'évaluer la fréquence de détection d'artefacts mais cela nécessite un complément d'étude.

3.2.2.2 Conditions d'observation

Comme indiqué dans l'Annexe 1.

3.2.2.3 Matériel d'essai

3.2.2.3.1 Evaluations principales

On doit utiliser un nombre relativement élevé d'images fixes et d'images animées. Les séquences pouvant aussi servir pour l'évaluation de la conversion TVHD-TVHD sont indiquées au § 3.1.3 (voir la Recommandation UIT-R BT.1210).

3.2.2.3.2 Evaluation auxiliaire

On pourrait demander à un certain nombre d'observateurs experts d'évaluer la qualité globale de divers programmes d'une durée de 5 à 20 min, comprenant des exemples de différents types de mouvements et de scènes à détails fins.

3.2.2.3.3 Caractéristiques du matériel d'essai

Le matériel d'essai critique pour la conversion des normes est susceptible de comporter des zones de détail fin ayant des vitesses et des directions de mouvement différentes.

3.2.2.4 Interprétation des résultats

En interprétant les résultats, il convient de veiller à ce que l'on n'attribue pas au processus de conversion des différences de qualité inhérentes aux deux normes de studio de TVHD. L'utilisation de séquences de référence enregistrées directement selon la norme de sortie serait utile à cet égard.

La qualité subjective de l'image convertie doit être «virtuellement équivalente» à l'image d'entrée à moins qu'elle ne soit limitée par les paramètres de l'une ou l'autre norme.

4 Evaluations de la qualité des images de studio classiques obtenues à partir d'images TVHD dans un environnement de studio

4.1 Eléments d'évaluation

L'interface entre la TVHD et la télévision classique peut impliquer des conversions du nombre de lignes, de la fréquence de trame et du format d'image, bien que des cas puissent se présenter qui ne nécessitent pas de conversions de la fréquence de trame. La qualité de l'image classique doit être la même que pour la production directe dans la norme classique.

4.2 Dégradations dues à la conversion des normes

4.2.1 Dégradations dues à la conversion du nombre de lignes

Les conversions faisant intervenir des changements dans le nombre de lignes actives peuvent se traduire par des perturbations perceptibles des contours des objets qui se déplacent verticalement. Ces perturbations peuvent être plus prononcées pour les conversions qui augmentent le nombre de lignes par rapport à celles qui le diminuent.

4.2.2 Dégradations dues à la conversion de la fréquence de trame

Les conversions faisant intervenir des changements dans la fréquence de trame se traduiront par l'apparition de défauts artificiels tels que le bruitage, limités aux zones animées de l'image. Le niveau de ces dégradations est lié au rapport fréquence de trame/complexité de l'algorithme de conversion. Certaines techniques, telles que la compensation adaptative du mouvement, peuvent ramener ces défauts à des niveaux très bas.

4.2.3 Dégradations dues à la conversion du format d'image

La conversion entre le grand format de la TVHD et le format 4:3 classique peut se traduire par une perte du contenu significatif de l'image ou de résolution. Il ne s'agit cependant pas d'un domaine où les évaluations subjectives sont susceptibles de fournir des directives utiles.

4.3 Evaluation de la qualité de la télévision classique obtenue à partir d'un signal de TVHD

4.3.1 Méthodologie

Il est évident que la qualité des convertisseurs de télévision classique ne peut être complètement évaluée au moyen de séquences d'images animées. Pour l'évaluation des petites dégradations de portée limitée auxquelles on peut s'attendre, on estime que la méthode à double stimulus utilisant l'échelle de qualité continue est la plus efficace. On demande aux observateurs de visionner une paire de séquences, l'une en format de studio classique direct et l'autre en format classique approprié, mais issu de la TVHD.

4.3.2 Conditions d'observation

Les conditions d'observation sont celles que spécifient les Recommandations UIT-R BT.500 et UIT-R BT.1129.

4.3.3 Matériel d'évaluation

Il convient d'utiliser une large gamme de matériel d'évaluation représentant des cas critiques (voir la Recommandation UIT-R BT.1210). Il peut être utile d'utiliser aussi des images fixes.

Deux autres types de séquences, susceptibles d'être plus critiques, pourraient également faire partie de l'essai:

- scènes avec mouvement de zoom,
- scènes avec mouvements dans des directions opposées comme sur une place de marché.

Des essais devraient également être effectués avec du matériel traité en aval.

4.3.4 Interprétation

Théoriquement, une interface TVHD-télévision classique devrait donner la même qualité que la télévision classique directe. Mais cette exigence ne sera probablement pas complètement satisfaite pour le rendu du mouvement, et la fréquence des dégradations de l'image évaluées dans les programmes de télévision devrait être étudiée. Cela peut supposer une étude en deux étapes comme celle appliquée à la conversion TVHD-TVHD.

5 Evaluation de la qualité des systèmes d'émission de TVHD obtenue à partir d'une norme de studio de TVHD

5.1 Eléments d'évaluation

Les caractéristiques du système qui présentent de l'intérêt pour cette évaluation sont les suivantes:

5.1.1 Qualité fondamentale

Il s'agit de la qualité de l'image dans des conditions de réception parfaites, c'est-à-dire lorsque le rapport signal/bruit, S/N , ou le rapport porteuse/bruit, C/N , sont élevés.

5.1.2 Fonction caractéristique de dégradation

Il s'agit de la relation entre la qualité de l'image et le bruit (qui a une caractéristique dépendant du système de modulation utilisé). On peut déterminer la gamme devant faire l'objet de l'évaluation en effectuant un essai préliminaire et en essayant d'obtenir entre 5 et 8 points sur l'échelle. Pour les systèmes à modulation d'amplitude, le rapport S/N est généralement compris entre 25 et 55 dB et pour les systèmes à modulation de fréquence, le rapport C/N est compris entre 0 et 30 dB.

5.1.3 Fonction caractéristique d'écho

Il s'agit de la relation entre la qualité de l'image et l'amplitude et le temps de propagation de l'écho. Cela concerne plus particulièrement les systèmes à modulation d'amplitude. La gamme d'évaluation doit être déterminée grâce à des essais préliminaires mais une solution appropriée consisterait à obtenir de l'information sur trois courbes ayant un signal à retard ajouté à un signal sans retard, avec respectivement, un retard de 150 ns, 1 μ s et 5 μ s, et avec des amplitudes d'écho comprises entre -5 et -25 dB par rapport au signal utile.

5.1.4 Fonction caractéristique d'interférence

Il convient d'évaluer les fonctions caractéristiques d'interférence dans le même canal et dans le canal adjacent.

Il y aurait peut-être lieu d'évaluer les caractéristiques ci-dessus (§ 5.1.2 à 5.1.4) à la fois avec et sans embrouillage.

5.2 Méthodologie

5.2.1 Qualité fondamentale

Le problème de conception fondamental pour une émission TVHD est de respecter, aussi fidèlement que possible, les exigences visuelles de la TVHD à l'intérieur de la largeur de bande disponible. Pour cela, on peut utiliser le sous-échantillonnage spatial ou temporel, ou les deux.

De telles techniques peuvent donner lieu à des dégradations notables ou à des pertes de qualité qui dépassent celles que l'on peut attribuer au format de studio. Le sous-échantillonnage spatial peut se traduire par des pertes notables dans la résolution horizontale, verticale ou diagonale. Le sous-échantillonnage temporel peut produire des réductions notables de la qualité du rendu du mouvement. Le sous-échantillonnage spatio-temporel peut donner lieu à des pertes notables de la résolution spatiale pour les séquences d'images animées.

Il est clair que l'on a besoin d'images à haute résolution et de séquences d'images animées pour l'évaluation des formats d'émission de TVHD. Cependant, afin de garantir une évaluation globale adéquate et représentative, une méthode d'évaluation à deux étapes est proposée pour ce qui est de la qualité fondamentale.

5.2.1.1 Evaluations principales

Il s'agit des évaluations qui sont considérées comme étant les évaluations les plus utiles. Une méthode à double stimulus utilisant une échelle de qualité continue devrait être utilisée. La référence doit être le signal de source de studio et le signal d'essai devrait être le signal d'émission.

5.2.1.2 Evaluations auxiliaires

On demande à un certain nombre d'observateurs experts d'attribuer une note de qualité globale à plusieurs programmes représentatifs dans le format d'émission. Il est également possible d'évaluer la fréquence de détection des défauts artificiels mais cela nécessite un complément d'étude.

5.2.2 Fonction caractéristique de dégradation, d'écho et d'interférence

Il convient d'utiliser une méthode à double stimulus utilisant une échelle de dégradation conforme à la Recommandation UIT-R BT.500.

Deux méthodes peuvent être adoptées:

- *fonction caractéristique de dégradation cumulative*: on considère les points auxquels se produisent des pertes inacceptables par rapport à la référence de haute qualité non dégradée;
- *fonction caractéristique de dégradation non cumulative*: on considère les points auxquels se produisent des pertes inacceptables par rapport aux formats d'émission non dégradés.

5.3 Conditions d'observation

Voir l'Annexe 1.

5.4 Matériel d'essai

5.4.1 Qualité fondamentale

Le matériel d'essai doit être choisi dans une gamme d'images fixes et de séquences d'images animées à haute résolution représentant des cas critiques mais pas excessivement.

Le matériel susceptible de représenter des cas critiques doit comporter des scènes à nombreux détails avec des vitesses et des directions de mouvement différentes (voir la Recommandation UIT-R BT.1210).

5.4.2 Fonction caractéristique de dégradation, d'écho et d'interférence

On devrait atteindre des résultats satisfaisants en utilisant seulement un ensemble limité d'images fixes et animées. La note globale moyenne peut généralement être calculée et utilisée de façon significative (voir la Recommandation UIT-R BT.1210).

5.5 Interprétation des résultats

5.5.1 Qualité fondamentale

Il semble raisonnable de dire que, pour être efficace, la qualité du signal d'émission de TVHD doit être plus proche de celle du signal de studio de TVHD que de celle du signal *RGB* de la télévision classique.

En règle générale, il faut que, pour la plupart du matériel, on obtienne un gain de qualité suffisant par rapport à la télévision classique; par ailleurs, les défauts artificiels temporels doivent être suffisamment faibles de manière à ne pas s'écarter de la qualité TVHD.

5.5.2 Fonction caractéristique de dégradation, d'écho et d'interférence

Cette question doit faire l'objet d'un complément d'étude.

6 Evaluation de la qualité des images compatibles insérées dans des formats d'émission de TVHD

6.1 Éléments d'évaluation

Certains systèmes d'émission de TVHD sont conçus pour permettre la réception simultanée sur des récepteurs TVHD et des récepteurs classiques. Le § 5 concerne l'évaluation de la qualité d'émission TVHD proprement dite. Le présent paragraphe concerne la qualité du signal classique simultanément reçu.

Au niveau de la conception, un compromis doit généralement être établi entre la qualité obtenue en visualisation TVHD et la qualité obtenue en visualisation classique, ce qui se traduit par un certain degré de compatibilité dépendant du niveau de dégradation introduit. Cela implique a priori l'étude des mêmes facteurs qui sont énumérés au § 5 mais cette fois appliqués à l'image compatible.

Les systèmes d'émission de TVHD proposés impliquent un traitement temporel ainsi que d'autres mécanismes qui pourraient causer des dégradations aux images compatibles.

6.2 Méthodologie

Pour la qualité de base, on pourra appliquer la méthode à double stimulus utilisant une échelle de qualité continue avec du matériel préparé directement au format d'émission classique ou du matériel directement converti à partir du format TVHD studio comme référence. En ce qui concerne les fonctions caractéristiques de dégradation, d'écho et d'interférence, on pourra utiliser la méthode à double stimulus utilisant une échelle de dégradation avec du matériel préparé directement au format d'émission classique (sans dégradation additionnelle) ou du matériel converti directement à partir du format TVHD studio (sans dégradation additionnelle) comme référence. Dans tous les cas, le signal d'essai sera l'image reçue compatible.

6.3 Conditions d'observation

Comme indiqué dans les Recommandations UIT-R BT.500 et UIT-R BT.1129 pour la télévision classique.

6.4 Matériel d'essai

Il convient d'utiliser une série d'images fixes et animées.

Les caractéristiques du matériel d'essai doivent correspondre généralement à celles des évaluations décrites au § 5 (c'est-à-dire critiques mais sans excès) (voir aussi la Recommandation UIT-R BT.1210).

6.5 Interprétation des résultats

En termes quantitatifs, il est difficile de déterminer ce que doit être la qualité des images «compatibles», ne serait-ce qu'en raison du caractère continu des échelles d'évaluation.

Les résultats pour chaque image ou séquence d'essai doivent être présentés séparément.

La qualité de l'image insérée doit en principe être «équivalente» à celle du signal de référence. Dans la pratique, il convient d'atteindre un degré de compatibilité approuvé.

7 Evaluation de la qualité du film cinématographique obtenue à partir de matériel source TVHD

A étudier.

8 Comparaisons entre différents formats de TVHD en compétition

Occasionnellement, il pourra être nécessaire de comparer différents formats en compétition pour la TVHD afin de faire un choix. On considère que ces comparaisons pourraient être particulièrement utiles pour identifier les meilleures caractéristiques des différents formats à l'essai.

8.1 Comparaisons des systèmes de studio de TVHD

On a identifié trois manières de comparer les différents formats de studio en compétition:

- directement, par comparaison côte à côte;
- indirectement, par comparaison implicite à une condition de référence commune dans une seule expérience;
- théoriquement, en établissant différents classements en fonction de conditions optimales déterminées par des moyens psychophysiques.

8.2 Comparaisons directes

A étudier.

8.3 Comparaisons indirectes

Pour faire une comparaison indirecte, il faut une référence commune qui serve à évaluer chaque système à l'essai. Les méthodes subjectives normalement utilisées pour les comparaisons indirectes (c'est-à-dire les méthodes à double stimulus) emploient des références qui permettent généralement d'obtenir une meilleure qualité que celle des conditions à l'essai.

Toutefois, du fait de la haute qualité des systèmes de TVHD de studio en compétition, il est difficile de trouver de telles références. C'est pourquoi il peut être indiqué d'avoir recours à des scènes observées directement pour obtenir la référence.

Pour qu'un essai indirect soit valable, la référence directement observée doit rester constante pour tous les systèmes soumis à l'essai. S'agissant des images fixes, on peut bien entendu utiliser à cet effet des transparents ou des photographies. En revanche, pour les images animées, il faut employer pour la référence des séquences animées entièrement reproductibles. Pour cela, on peut utiliser des scènes commandées mécaniquement (par exemple des dioramas).

Il est tout aussi important, sauf pour les différences inhérentes aux formats eux-mêmes, de faire en sorte que le matériel d'essai soit le même pour tous les systèmes à l'essai. Cela sera possible si la caméra vidéo du système à l'essai sert à enregistrer l'image fixe ou la séquence de référence, à la condition que la référence reste constante.

Il convient de noter que tous les systèmes à l'étude devraient être évalués dans un contexte expérimental unique (c'est-à-dire que les observateurs devraient voir, au cours de l'expérience, une séquence aléatoire des systèmes à l'essai). Pour cela, on peut alterner les caméras utilisées pour les systèmes à l'essai. La valeur qui devrait rester constante doit être choisie de façon à convenir à tous les systèmes à l'essai. Il ne sera peut-être pas toujours possible de déduire des données générales applicables aux conditions énoncées dans l'Annexe 1 d'après les résultats obtenus avec les conditions d'observation que permettent les équipements actuels. Il faudrait donc interpréter les résultats des essais de manière à faire la différence entre les valeurs liées à la norme du système et celles concernant la mise en œuvre pratique.

Les scènes observées directement peuvent fournir une référence dont la qualité est considérablement supérieure à celle des systèmes testés. Si tel est le cas, deux problèmes peuvent se poser:

- Les différences entre les notes subjectives attribuées aux systèmes testés peuvent être réduites au minimum de manière artificielle. Lorsque les observateurs portent un jugement, ils sont en général influencés par la gamme et la répartition des qualités observées. Lorsque la gamme de qualité (y compris celle indiquée par la référence) est très étendue, les images ayant à peu près la même qualité sont généralement perçues comme ayant une qualité analogue, alors que cette similitude n'est pas si marquée lorsque la gamme de qualité est plus limitée ou en cas de comparaison directe.

- La méthode d'essai préférée peut varier: si les conditions (y compris les conditions de référence et d'essai) portent sur une large gamme de qualité, la méthode à double stimulus utilisant une échelle de dégradation peut être utilisée pour les comparaisons indirectes. Toutefois, si les conditions couvrent une gamme de qualité plus limitée, la méthode préférée est celle à double stimulus utilisant une échelle de qualité continue.

En conséquence, deux méthodes sont possibles selon l'objectif de l'essai: si l'essai est censé mettre des systèmes en rapport avec une norme «parfaite», une référence de qualité nettement supérieure et la méthode à double stimulus utilisant une échelle de dégradation peuvent être utilisées. Toutefois, en pareil cas, des différences ténues entre les systèmes risquent de ne pas être décelées. Si les essais ont pour objet d'établir des distinctions subtiles entre les systèmes, l'emploi d'une référence de qualité nettement supérieure devrait être évité et la méthode à double stimulus utilisant une échelle de qualité continue devrait être utilisée. On sera peut-être amené, dans ce dernier cas, à limiter la qualité de la scène directement observée au moyen de la composition, de l'éclairage, du filtrage optique, etc.

8.3.1 Méthodologie

Selon la gamme de qualité utilisée dans l'essai, on peut faire appel à la méthode à double stimulus utilisant une échelle de qualité continue ou à la méthode à double stimulus utilisant une échelle de dégradation.

Si la première méthode est retenue, il sera peut-être utile d'examiner une variante de cette méthode, qui utilise des expositions relativement longues afin d'inciter les observateurs à déceler les effets difficiles à percevoir, notamment dans les séquences animées.

Si l'on utilise la méthode à double stimulus utilisant une échelle de qualité continue, chaque essai comprendra plusieurs affichages alternés des conditions de référence et des conditions à tester. Pour la moitié des essais (déterminés de manière aléatoire), il faudra présenter la condition de référence en premier; pour les autres, on présentera la condition à tester en premier. Si l'on emploie la méthode d'évaluation à double stimulus utilisant une échelle de dégradation, chaque essai comprendra une seule alternance entre la condition de référence et la condition d'essai, la condition de référence étant présentée en premier.

Les matériels d'essai et les précautions à prendre pour réduire au minimum les effets d'une contamination possible des résultats sont ceux donnés au § 3.1.3.

Chaque observateur devrait voir le miroir d'observation à travers une ouverture qui permettrait une observation binoculaire, avec peu de mouvements de la tête ou pas de mouvement du tout. Les observateurs pourraient être appelés individuellement ou par petits groupes. Toutefois, si plusieurs observateurs passent en même temps, l'angle de vision (par rapport à la scène) doit être maintenu constant pour tous les observateurs.

Etant donné que l'on sait que des groupes linguistiques différents utilisent différemment les termes relatifs à l'échelle de qualité et de dégradation, tous les essais devraient être faits dans une seule langue avec des observateurs ayant une bonne connaissance de cette langue.

8.3.2 Conditions d'observation

Voir l'Annexe 1.

8.3.3 Matériels d'évaluation

En cours de rassemblement.

8.3.4 Interprétation des résultats

L'interprétation des résultats se fait sur la base du classement relatif des systèmes en compétition par rapport à la référence commune observée directement. Il convient de tenir compte des points relevés dans le § 3.1.4.

8.4 Comparaisons théoriques

Sur le fond, cette méthode consiste à examiner, paramètre par paramètre, le classement des systèmes possibles en fonction des conditions psychophysiques pertinentes. C'est la méthode proposée.

9 Comparaisons de formats d'émission de TVHD en compétition

Comme dans le cas de systèmes de studio de TVHD, les comparaisons peuvent être directes, indirectes ou théoriques. Ici, on n'a tenu compte que des comparaisons indirectes.

9.1 Qualité fondamentale

L'approche est généralement la même que pour les formats de studio de TVHD (voir le § 8.3). Dans ce cas, la méthode à double stimulus utilisant une échelle de qualité continue peut être utilisée, avec une seule référence de haute qualité.

9.2 Fonction caractéristique de dégradation

Ces essais sont en général ceux qui sont indiqués au § 5. Toutefois, le but est de comparer les fonctions caractéristiques de dégradation de tous les systèmes en compétition.

10 Autres questions

D'autres questions connexes sont envisagées: méthodes d'évaluation pour la conversion de la TVHD en film 35 mm, utilisation des descripteurs de qualité de l'UIT-R, interprétation des objectifs de qualité en termes de résultats numériques d'évaluation, qualité subjective et caractéristique S/N des signaux de TVHD, relations entre les aspects image et son de la TVHD.

Selon les résultats d'une série d'essais subjectifs effectués en Italie afin d'évaluer, pour divers matériels et sur des écrans de dimensions différentes, la distance d'observation préférée dans le cas de programmes de TVHD, la distance moyenne préférée est de cinq à six fois la hauteur de l'image (5 à 6 H).

D'après des études antérieures de la télévision à résolution classique, la distance d'observation préférée y serait de 8 à 9 H .

Il faut continuer à étudier la relation entre la distance moyenne d'observation préférée et celle des évaluations subjectives.

ANNEXE 3

Facteurs d'évaluation propres à l'évaluation globale de la TVHD

Dans une étude récente de grande envergure effectuée au Canada et aux Etats-Unis d'Amérique, des observateurs ont évalué la TVHD (système MUSE-E par satellite) à la fois en termes absolus et par comparaison avec des images NTSC de qualité studio. Les évaluations portaient à la fois sur la qualité globale de l'image et sur des facteurs d'évaluation spécifiques, dont la netteté de l'image, la qualité de la couleur, le rendu du mouvement, le rendu de la profondeur, la luminosité de l'image, les dimensions et le format de l'écran. Les résultats ont montré que:

- les jugements absolus de la TVHD seulement sont concentrés à la fin de l'échelle de qualité, ce qui laisse supposer des problèmes éventuels si les résultats des essais séparés avec différents systèmes de TVHD devaient être comparés sur la base des jugements de qualité absolus;
- des observateurs ont pu répondre en discriminant aux différents facteurs d'évaluation, ce qui suggère que la méthode des facteurs spécifiques pourrait être utile dans les évaluations futures;
- les appréciations de la qualité de l'image globale étaient fortement liées à la plupart, mais non à l'ensemble, des facteurs d'évaluation en fonction desquels la TVHD a été perçue comme étant différente de la norme NTSC, ce qui suggère que la qualité globale de l'image peut ne pas entièrement refléter les réactions de l'observateur;
- les appréciations relatives aux facteurs spécifiques ont été, dans une certaine mesure, interdépendantes, ce qui suggère des hiérarchies possibles entre les facteurs utilisés dans les évaluations, et l'existence possible de facteurs de qualité fondamentaux d'ordre inférieur;
- les appréciations de la qualité globale de l'image ont été, dans une certaine mesure, affectées par les facteurs d'évaluation différents en fonction de la distance d'observation, ce qui suggère la nécessité d'un examen approfondi des distances d'observation à utiliser dans les évaluations.