

RECOMMANDATION UIT-R BT.1788

**Méthode d'évaluation subjective de la qualité vidéo
dans les applications multimédias**

(Question UIT-R 102/6)

(2007)

Domaine de compétence

Les systèmes de radiodiffusion numérique vont permettre de fournir des applications de radiodiffusion multimédias et de données comportant des signaux vidéo, des signaux audio, des images fixes, du texte et des éléments graphiques. La présente Recommandation spécifie des méthodes non interactives d'évaluation subjective de la qualité vidéo dans les applications multimédias.

L'Assemblée des radiocommunications de l'UIT,

considérant

- a) que des systèmes de radiodiffusion numérique sont mis en place dans bon nombre de pays;
- b) que des services de radiodiffusion multimédias et de données comportant des signaux vidéo, des signaux audio, des images fixes, du texte, des éléments graphiques, etc., utilisant des systèmes de radiodiffusion numérique existent ou sont en projet;
- c) que les services multimédias seront fondés sur une infrastructure de radiodiffusion caractérisée par l'utilisation possible de récepteurs fixes et de récepteurs mobiles, de débits d'images fixes et de débits d'images variables, de différents formats d'image, de codecs vidéo évolués, par une éventuelle perte de paquets, etc.;
- d) qu'il sera nécessaire de spécifier la qualité de fonctionnement requise et de vérifier que les solutions techniques envisagées pour chaque service sont adaptées à la qualité de fonctionnement requise pour le service en question;
- e) que cette vérification reposera essentiellement sur une évaluation subjective de la qualité vidéo dans des conditions bien définies;
- f) que les méthodes d'évaluation subjective spécifiées dans la Recommandation UIT-R BT.500 peuvent être utilisées pour les applications multimédias;
- g) que des méthodes d'évaluation subjective autres que celles spécifiées dans la Recommandation UIT-R BT.500 peuvent aussi être utilisées;
- h) qu'il est important d'adopter des méthodes normalisées pour les échanges d'informations entre divers laboratoires,

recommande

- 1** que les méthodes générales d'essai, à savoir les échelles de notation et les conditions d'observation pour l'évaluation de la qualité des images décrites dans l'Annexe 1, soient utilisées pour les expériences en laboratoire ainsi que pour les évaluations en exploitation chaque fois que cela est possible en ce qui concerne les applications multimédias;

2 que tous les rapports d'essai donnent la description complète des configurations d'essai, du matériel d'essai, des observateurs et des méthodes;

3 que, pour faciliter les échanges d'informations entre différents laboratoires, les données recueillies soient traitées conformément aux techniques statistiques décrites à l'Annexe 2.

NOTE 1 – La Commission d'études 6 des radiocommunications doit élaborer une bibliothèque de matériel vidéo appropriée pour l'évaluation subjective de la qualité vidéo dans les applications multimédias.

Annexe 1

Description des méthodes d'évaluation

1 Introduction

De nombreux pays ont commencé à mettre en place des systèmes de radiodiffusion numérique permettant d'offrir des applications de radiodiffusion multimédias et de données comportant des signaux vidéo, des signaux audio, des images fixes, du texte et des éléments graphiques.

Des méthodes normalisées d'évaluation subjective sont nécessaires pour spécifier la qualité de fonctionnement requise et pour vérifier que les solutions techniques envisagées pour chaque application conviennent. Des méthodes subjectives sont nécessaires car elles donnent des mesures qui permettent aux industriels d'anticiper plus directement les réactions des utilisateurs finals.

Le système de radiodiffusion nécessaire pour fournir des applications multimédias est sensiblement différent de celui qui est actuellement utilisé: accès aux informations par le biais de récepteurs fixes et/ou mobiles; débit d'images pouvant être fixe ou variable; grande variété de tailles d'image possibles (à savoir, de SQCIF à HDTV); signaux audio, texte et/ou signaux sonores généralement imbriqués dans l'image vidéo; possibilité de traitement de l'image vidéo au moyen de codecs vidéo évolués; et forte dépendance de la distance d'observation préférée vis-à-vis de l'application.

Il convient d'appliquer dans ce nouveau contexte les méthodes d'évaluation subjective spécifiées dans la Recommandation UIT-R BT.500. En outre, de nouvelles méthodes pourraient être envisagées pour les systèmes multimédias afin de répondre aux besoins des utilisateurs en ce qui concerne les caractéristiques du domaine multimédia.

La présente Recommandation décrit des méthodes non interactives d'évaluation subjective de la qualité vidéo dans les applications multimédias. Ces méthodes peuvent notamment être utilisées pour choisir des algorithmes, déterminer la catégorie d'un système audiovisuel en fonction de ses performances ou évaluer le niveau de qualité vidéo pendant une connexion audiovisuelle.

Les termes et définitions applicables dans la présente Recommandation figurent dans l'Appendice 3 à l'Annexe 1.

2 Description générale

2.1 Conditions d'observation

Les conditions d'observation recommandées sont données dans le Tableau 1. La taille et le type de l'affichage utilisé devraient être choisis en fonction de l'application considérée. Comme plusieurs technologies d'affichage doivent être utilisées dans les applications multimédias, il convient d'indiquer toutes les informations utiles concernant l'affichage (par exemple, fabricant, modèle et spécifications) utilisé pour l'évaluation.

Lorsque des systèmes utilisant des PC sont utilisés pour présenter les séquences, il convient d'indiquer aussi les caractéristiques de ces systèmes (par exemple, carte d'affichage vidéo).

Le Tableau 2 donne un exemple de paramètres de configuration du système multimédia à évaluer.

Si les images d'essai sont obtenues à partir d'un décodeur-lecteur combiné particulier, il faut ôter l'enveloppe propriétaire pour obtenir un affichage anonyme. Cette opération est nécessaire pour garantir que l'évaluation de la qualité n'est pas influencée par la connaissance de l'environnement d'origine.

Lorsque les systèmes évalués dans un essai utilisent un format d'image réduit (par exemple CIF, SIF ou QCIF, etc.), les séquences devraient être affichées dans une fenêtre de l'écran. L'arrière-plan de l'écran devrait être de couleur grise 50%.

TABLEAU 1

Conditions d'observation recommandées à utiliser pour l'évaluation de la qualité dans les applications multimédias

Paramètre	Valeur
Distance d'observation ⁽¹⁾	Valeur imposée: 1-8 H Pas de valeur imposée: suivant la préférence de l'observateur
Luminance de crête de l'écran	70-250 cd/m ²
Rapport de la luminance de l'écran inactif à la luminance de crête	≤ 0,05
Rapport de la luminance de l'écran, quand on ne reproduit que le niveau du noir dans une salle complètement obscure, à celle qui correspond au blanc maximal	≤ 0,1
Rapport de la luminance de l'arrière-plan, derrière le moniteur, à la luminance de crête de l'image ⁽²⁾	≤ 0,2
Chromaticité de l'arrière-plan ⁽³⁾	D ₆₅
Eclairement d'ambiance de la salle ⁽²⁾	≤ 20 lux

⁽¹⁾ La distance d'observation dépend généralement de l'application.

⁽²⁾ La valeur indiquée permet une détectabilité maximale des distorsions; pour certaines applications, des valeurs supérieures sont autorisées ou sont déterminées par l'application.

⁽³⁾ Pour les moniteurs de PC, la chromaticité de l'arrière-plan devrait être la plus proche possible de la chromaticité du «point blanc» de l'affichage.

TABLEAU 2
Configuration du système multimédia à évaluer

Paramètre	Spécification
Type d'affichage	
Taille de l'affichage	
Carte d'affichage vidéo	
Fabricant	
Modèle	
Informations concernant l'image	

2.2 Signaux source

Le signal source fournit directement l'image de référence et les signaux d'entrée pour le système à évaluer. La qualité des séquences source devrait être aussi élevée que possible. Le signal vidéo devrait en principe être enregistré dans des fichiers multimédias utilisant yuv (formats 4:2:2, 4:4:4) ou RGB (24 ou 32 bits). Lorsque l'expérimentateur souhaite comparer les résultats obtenus par différents laboratoires, il est nécessaire d'utiliser un ensemble commun de séquences source afin d'éliminer une autre source de variation.

2.3 Choix du matériel d'essai

Le nombre et le type des scènes d'essai sont très importants pour l'interprétation des résultats de l'évaluation subjective. Pour certains processus, les dégradations observées sur la plupart des séquences peuvent être plus ou moins identiques. Dans ces conditions, les résultats obtenus à partir d'un petit nombre de séquences (par exemple, deux) devraient rester significatifs. Toutefois, l'impact des nouveaux systèmes dépend souvent pour beaucoup de la scène et du contenu de la séquence. En pareil cas, le nombre et le type des scènes d'essai devraient être choisis de manière à pouvoir procéder à une généralisation raisonnable pour les programmes normaux. En outre, il convient de choisir un matériel «critique mais sans excès» pour le système à évaluer. Par «sans excès» on entend que la scène pourra très bien faire partie de programmes de télévision normaux. Les caractéristiques spatiales et temporelles perçues pour une scène pourraient donner une indication utile de la complexité de cette scène. Des mesures des caractéristiques spatiales et temporelles perçues sont présentées plus en détail à l'Appendice 1 à l'Annexe 1.

2.4 Gamme de conditions et ancrage

Etant donné que la plupart des méthodes d'évaluation 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). Il convient, si possible, d'utiliser une large gamme de qualité.

2.5 Observateurs

Il convient de sélectionner 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, ni des observateurs expérimentés. Avant chaque séance, les observateurs seront sélectionnés à l'aide de mires de Snellen ou de Landolt pour leur acuité visuelle normale ou rendue normale par correction et leur vision normale des couleurs, cela à l'aide de mires choisies à cet effet (d'Ishihara, par exemple).

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.

Les expérimentateurs devraient donner le plus de détails possibles sur les caractéristiques des groupes d'évaluation qu'ils ont retenus afin d'étudier plus avant ce facteur. Ils pourraient donner des précisions sur l'activité professionnelle (par exemple, fonctionnaire d'un organisme de radiodiffusion, étudiant d'une université, personnel de bureau), le sexe et l'âge.

2.6 Instructions pour les évaluations

La méthode d'évaluation, les types de dégradation ou de facteur de qualité auxquels il faut s'attendre, l'échelle d'évaluation, le séquençement, etc., seront présentés avec soin aux observateurs. Les séquences d'entraînement qui montrent la gamme et le type de dégradation à évaluer doivent présenter d'autres scènes que celles qui sont utilisées dans les essais mais avoir une sensibilité comparable.

2.7 Modèle expérimental

Le soin est laissé à l'expérimentateur de choisir un modèle expérimental en fonction d'objectifs spécifiques en termes de coût et de précision. Il est préférable d'inclure au moins deux reproductions (c'est-à-dire répétitions de conditions identiques) dans l'expérience. Les reproductions permettent de calculer la fiabilité individuelle et, si nécessaire, d'ignorer des résultats non fiables issus de certains observateurs. Elles permettent aussi de compenser dans une certaine mesure les effets d'apprentissage dans le cadre d'un essai donné. On obtiendra une autre amélioration du traitement des effets d'apprentissage en prévoyant quelques «présentations fictives» au début de chaque séance d'essai. Il conviendra de choisir des conditions représentatives des présentations qui seront utilisées plus tard pendant la séance. Les présentations préliminaires ne sont pas à prendre en compte dans l'analyse statistique des résultats d'essai.

Une séance, c'est-à-dire une série de présentations, ne devrait pas durer plus d'une demi-heure.

Lorsque plusieurs scènes ou algorithmes sont évalués, ils devraient être présentés dans un ordre aléatoire. Cet ordre aléatoire pourrait être modifié pour faire en sorte que les mêmes scènes ou les mêmes algorithmes ne soient pas présentés à des moments proches (c'est-à-dire consécutivement).

3 Méthodes d'évaluation

La qualité vidéo des systèmes multimédias peut être examinée au moyen des méthodes décrites dans la Recommandation UIT-R BT.500. Les méthodes retenues sont énumérées au § 3.1.

Le § 3.2 décrit une méthode additionnelle appelée méthode d'évaluation subjective de la qualité vidéo multimédia SAMVIQ, qui tire parti des caractéristiques du domaine multimédia et qui peut être utilisée pour évaluer les performances des systèmes multimédias.

3.1 Méthodes décrites dans la Recommandation UIT-R BT.500

Il convient d'utiliser les méthodes suivantes décrites dans la Recommandation UIT-R BT.500 pour évaluer la qualité vidéo des systèmes multimédias:

- Méthode à double stimulus utilisant une échelle de dégradation (DSIS), décrite au § 4 à l'Annexe 1 de la Recommandation UIT-R BT.500.

- Méthode à double stimulus utilisant une échelle de qualité continue (DSCQS), décrite au § 5 à l'Annexe 1 de la Recommandation UIT-R BT.500.
- Méthodes à un seul stimulus (SS), décrites au § 6.1 à l'Annexe 1 de la Recommandation UIT-R BT.500.
- Méthodes de comparaison de stimulus (SC), décrites au § 6.2 à l'Annexe 1 de la Recommandation UIT-R BT.500.
- Méthode d'évaluation continue de la qualité avec stimulus unique (SSCQE), décrite au § 6.3 à l'Annexe 1 de la Recommandation UIT-R BT.500.

3.2 Evaluation subjective de la qualité vidéo multimédia (SAMVIQ)

Dans cette méthode, l'observateur a accès à plusieurs versions d'une même séquence. Lorsqu'il a évalué toutes les versions, il peut ensuite accéder au contenu de la séquence suivante.

L'observateur peut choisir aléatoirement chacune des différentes versions par le biais d'une interface graphique informatique. Il a la possibilité d'arrêter chaque version d'une séquence, de la revoir et d'en modifier la note. Cette méthode comporte une séquence de référence (c'est-à-dire non traitée) explicite ainsi que plusieurs versions de la même séquence qui incluent à la fois des séquences traitées et une séquence non traitée (c'est-à-dire une référence cachée). Chaque version d'une séquence est affichée isolément et évaluée sur une échelle de qualité continue analogue à celle utilisée dans la méthode DSCQS. Du point de vue fonctionnel, la méthode ressemble donc beaucoup à une méthode à un seul stimulus avec accès aléatoire, mais un observateur peut visualiser la référence explicite chaque fois qu'il le souhaite, ce qui rend cette méthode analogue à une méthode utilisant une référence.

La méthode SAMVIQ utilise une échelle de qualité continue pour l'évaluation de la qualité intrinsèque des séquences vidéo. Chaque observateur déplace un curseur sur une échelle continue graduée de 0 à 100 avec 5 plages de qualité (excellente, bonne, assez bonne, médiocre, mauvaise).

L'évaluation de la qualité est réalisée *scène par scène* (voir la Fig. 1), avec *une référence explicite, une référence cachée et divers algorithmes*.

Pour rendre la méthode plus facilement compréhensible, on définit les termes spécifiques qui suivent:

<i>Scène:</i>	contenu audiovisuel
<i>Séquence:</i>	scène avec traitement combiné ou sans traitement
<i>Algorithme:</i>	une ou plusieurs techniques de traitement d'image.

3.2.1 Référence explicite, référence cachée et algorithmes

Une méthode d'évaluation comporte généralement des ancres de qualité afin de stabiliser les résultats. Deux ancres de qualité élevée sont utilisées dans la méthode SAMVIQ pour les raisons précisées ci-après. La réalisation de plusieurs essais a montré que l'écart type des notes est minimalisé si on utilise une *référence explicite* plutôt qu'une référence cachée ou qu'aucune référence. En particulier, pour évaluer les performances d'un codec, il vaut mieux utiliser une référence explicite pour obtenir des résultats les plus fiables possibles. Une *référence cachée* est également utilisée afin d'évaluer la qualité intrinsèque de la référence; on n'utilise pas la référence explicite car la présentation est anonyme de même que les séquences traitées. La désignation explicite de «référence» a une influence sur environ 30% des observateurs, qui donnent la note la plus élevée (100) à la référence explicite. Cette note est totalement différente de la note correspondante donnée pour la référence cachée. Il est à noter que lorsque aucune référence n'est disponible, l'essai reste possible mais l'écart type est nettement plus grand.

La méthode SAMVIQ convient bien dans un contexte multimédia car il est possible de combiner différents éléments du traitement d'image (par exemple type de codec, format d'image, débit, mise à jour temporelle, zoomage, etc.). Un *algorithme* désigne l'un de ces éléments ou une combinaison de ces éléments.

3.2.2 Conditions d'essai

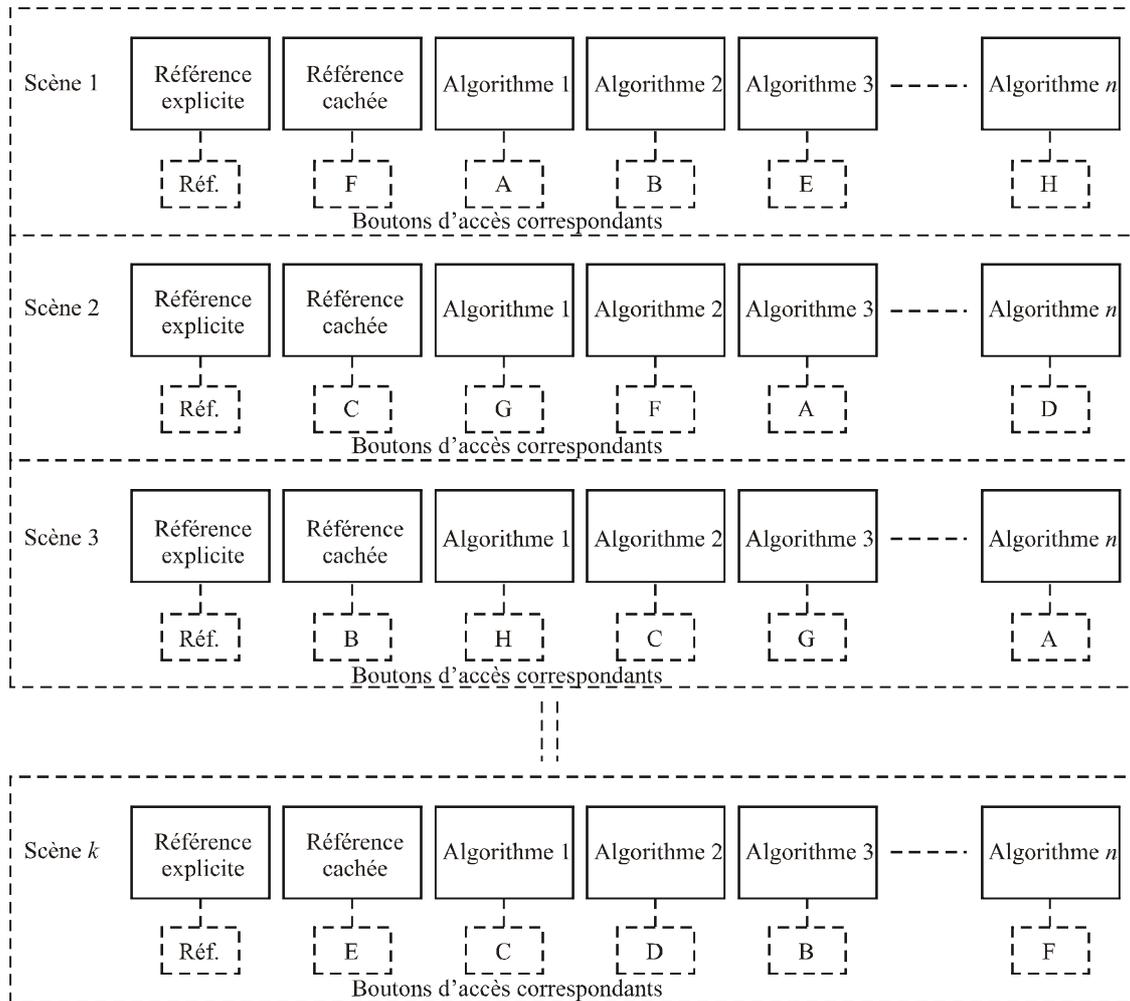
La variation de la criticité pendant une scène est limitée car on choisit des contenus homogènes en suivant les mêmes règles que celles qui sont implicitement utilisées par les autres méthodes donnant une note globale (par exemple, les méthodes à un seul stimulus). Une durée maximale de visualisation de séquence de 10 ou 15 s est alors suffisante pour obtenir une note de qualité stabilisée et fiable. Il convient d'utiliser les décodeurs-lecteurs propriétaires, ou une copie d'écran de leur sortie, afin de maintenir la qualité d'affichage voulue.

3.2.3 Organisation d'un essai

- a) L'essai est réalisé scène par scène, comme décrit sur la Fig.1.
- b) Pour une scène donnée, il est possible de visualiser et de noter n'importe quelle séquence dans n'importe quel ordre. Chaque séquence peut être visualisée et notée plusieurs fois.
- c) D'une scène à l'autre, l'accès aux séquences est randomisé, ce qui empêche les observateurs de tenter de voter de manière identique selon un ordre établi. En réalité, pour un essai donné, l'ordre des algorithmes reste le même afin de simplifier l'analyse et la présentation des résultats. Seul l'accès correspondant à partir d'un bouton identique est randomisé.
- d) Lors de la première visualisation, la séquence considérée doit être visualisée en totalité avant d'être notée, afin d'éviter que la séquence soit notée et arrêtée immédiatement.
- e) Pour pouvoir passer à la scène suivante, toutes les séquences de la scène considérée doivent avoir été notées.
- f) L'essai est terminé quand toutes les séquences de toutes les scènes ont été notées.

FIGURE 1

Exemple d'organisation d'un essai pour la méthode SAMVIQ



1788-01

La méthode SAMVIQ est mise en œuvre sous forme logicielle. En plus des boutons d'accès montrés sur la Fig. 1, des boutons «lecture», «arrêt», «scène suivante» et «scène précédente» sont nécessaires pour permettre à l'observateur de gérer la présentation des différentes scènes (voir par exemple l'Appendice 2 à l'Annexe 1). Lorsqu'une note a été donnée par l'observateur, elle devrait apparaître au-dessous du bouton d'accès correspondant à la scène considérée. Lorsque toutes les versions différentes d'une séquence ont été évaluées, l'observateur a toujours la possibilité de comparer les notes et, si nécessaire, de les modifier. Il n'est pas nécessaire de revoir la totalité de la séquence considérée car les grandes différences ont déjà été mises en avant au cours de la première visualisation.

Annexe 2

Présentation et analyse des données

1 Informations relatives à l'essai

Des informations précises relatives à l'environnement de l'essai sont nécessaires pour pouvoir reproduire un essai ou comparer les résultats de différents essais. Il est donc suggéré d'indiquer les informations relatives à l'environnement d'essai décrites dans le Tableau 3.

TABLEAU 3

Informations relatives à l'essai

Nom de la méthode	
Technologie d'affichage	
Nom de référence de l'affichage	
Niveau de luminance de crête (cd/m ²)	
Niveau de luminance du noir (cd/m ²)	
Niveau du noir: PLUGE (seuil de perception de la distance entre le niveau du noir et le niveau du noir supra = 8). Sinon, indiquer la valeur seuil	
Niveau de luminance de l'arrière-plan (cd/m ²)	
Eclairement (lux)	
Distance d'observation: – Pas de valeur imposée: devant l'affichage – Valeur imposée: nH	
Taille de l'affichage (diagonale en pouces)	
Rapport largeur/hauteur de l'affichage	
Format de l'affichage (nombre de colonnes et de lignes)	
Format d'entrée de l'image (nombre de colonnes et de lignes)	
Format de sortie de l'image ⁽¹⁾ (nombre de colonnes et de lignes)	
Température du blanc: D65 sinon Coordonnées du blanc (x, y)	
Nombre d'observateurs effectifs	

⁽¹⁾ Cette information est nécessaire lorsque l'image d'entrée est traitée, par exemple si son échelle est changée, au moment de son affichage.

Les caractéristiques de l'affichage peuvent avoir une incidence sur les résultats de l'essai. D'autres informations (par exemple, la réponse en luminance (fidélité du gamma) et les couleurs primaires) devraient être nécessaires pour les affichages sur écran plat.

Les caractéristiques des séquences vidéo sont importantes pour concevoir un essai ou pour expliquer ses résultats. Il est suggéré d'indiquer les caractéristiques spatio-temporelles décrites à l'Appendice 1 de l'Annexe 1. Il convient de tenir compte de ces informations pour la collecte de séquences d'essai dans la bibliothèque de matériel vidéo à utiliser pour l'évaluation subjective de la qualité vidéo dans les applications multimédias.

2 Méthodes d'analyse

Les méthodes d'analyse sont celles qui sont décrites au § 2 à l'Annexe 2 de la Recommandation UIT-R BT.500.

3 Sélection des observateurs

Pour les méthodes énumérées au § 3.1 de l'Annexe 1, les procédures de sélection sont décrites au § 2.3 à l'Annexe 2 de la Recommandation UIT-R BT.500.

Pour la méthode SAMVIQ, la procédure de sélection est décrite dans le paragraphe qui suit. Toutefois, cette procédure pourrait être utilisée pour les méthodes à un seul stimulus, DSIS et DSCQS. Elle est plus simple à mettre en œuvre que la procédure correspondante utilisée dans la Recommandation UIT-R BT.500 pour ces méthodes.

3.1 Procédure de sélection pour la méthode SAMVIQ

Chaque observateur doit disposer d'une méthode stable et cohérente pour évaluer correctement la dégradation de qualité pour chaque scène et algorithme. Le critère de rejet permet de confirmer le niveau de cohérence des notes d'un observateur par rapport à la note moyenne pour tous les observateurs ayant participé à une séance d'essai donnée. Dans la méthode SAMVIQ, comme dans la méthode DSQCS, tous les algorithmes (référence cachée ou implicite, ancre faible, séquences codées) peuvent être pris en considération. Le critère de décision est fondé sur une corrélation des notes individuelles avec les notes moyennes correspondantes pour tous les observateurs ayant participé à l'essai.

3.2 Corrélation de Pearson

La relation entre l'échelle de qualité et la plage de notes données par les observateurs est supposée être linéaire pour pouvoir appliquer la corrélation de Pearson.

Le principal objectif est de vérifier par une méthode simple si les notes données par un observateur sont cohérentes avec les notes moyennes pour tous les observateurs et pour l'ensemble de la séance d'essai. La référence cachée est considérée comme une ancre de qualité élevée. Si les ancres faible et élevée sont incluses, elles augmentent la note de corrélation; inversement, les décalages de corrélation entre les observateurs sont réduits.

$$r(x, y) = \frac{\sum_{i=1}^n x_i y_i - \frac{\left(\sum_{i=1}^n x_i\right)\left(\sum_{i=1}^n y_i\right)}{n}}{\sqrt{\left(\sum_{i=1}^n x_i^2 - \frac{\left(\sum_{i=1}^n x_i\right)^2}{n}\right)\left(\sum_{i=1}^n y_i^2 - \frac{\left(\sum_{i=1}^n y_i\right)^2}{n}\right)}}$$

x_i : note moyenne pour tous les observateurs pour le triplet (algorithme, débit, scène)

y_i : note donnée par un seul observateur pour le même triplet

n : (nombre d'algorithmes) \times (nombre de scènes)

i : {nombre de codec, nombre de débit, nombre de scène}.

3.3 Corrélation des rangs de Spearman

La corrélation des rangs de Spearman peut être appliquée même si la relation entre l'échelle de qualité et la plage de notes données par les observateurs n'est pas supposée être linéaire¹.

$$r(x, y) = \left[1 - \frac{6 \times \sum_{i=1}^n [R(x_i) - R(y_i)]^2}{n^3 - n} \right]$$

x_i : note moyenne pour tous les observateurs pour le triplet (algorithme, débit, scène)

y_i : note donnée par un seul observateur pour le même triplet

n : (nombre d'algorithmes) \times (nombre de scènes)

$R(x_i$ ou $y_i)$: rang

i : {nombre de codec, nombre de débit, nombre de scène}.

3.4 Critère de rejet final pour éliminer un observateur ayant participé à un essai donné

La corrélation des rangs de Spearman et la corrélation de Pearson sont utilisées pour éliminer un ou plusieurs observateurs conformément aux conditions suivantes:

SI [moyenne(r) – sdt(r)] > Seuil de corrélation maximal (MCT).

Seuil de rejet = Seuil de corrélation maximal (MCT).

SINON Seuil de rejet = [moyenne(r) – écart type(r)].

SI [r (Observateur_i)] > Seuil de rejet.

ALORS l'observateur «i» ayant participé à l'essai n'est pas éliminé.

SINON l'observateur «i» ayant participé à l'essai est éliminé.

où:

$r = \min$ (corrélation de Pearson, corrélation des rangs de Spearman)

moyenne(r): moyenne des corrélations pour tous les observateurs ayant participé à un essai donné

sdt(r): écart type des corrélations pour tous les observateurs ayant participé à un essai donné

Seuil de corrélation maximal (MCT) = 0,85.

La valeur de MCT de 0,85 est valable pour les méthodes SAMVIQ et DSCQS; une valeur de MCT de 0,7 doit être utilisée pour les méthodes à un seul stimulus et DSIS.

¹ Les résultats de corrélation de Pearson sont généralement très proches des résultats de corrélation de Spearman.

Appendice 1 à l'Annexe 1

Informations spatiales et spatio-temporelles

Les informations spatio-temporelles sont indiquées ci-après sous la forme de valeurs uniques pour chaque trame dans une séquence d'essai complète. Cela se traduit par une série temporelle de valeurs qui varieront généralement un peu. Les valeurs d'information de perception indiquées ci-dessous éliminent cette variabilité au moyen d'une fonction à maximum (valeur maximale pour la séquence). La variabilité proprement dite peut servir à d'utiles études, par exemple pour des tracés d'informations spatio-temporelles trame par trame. Le recours à des distributions des informations pour l'ensemble d'une séquence d'essai permet aussi de mieux évaluer les scènes comportant des coupes.

Information de perception spatiale (SI): grandeur représentative du degré de détail spatial d'une image. Cette valeur augmente généralement avec la complexité spatiale des scènes. Cette grandeur n'est pas censée mesurer l'entropie ni être associée aux informations définies en théorie de la communication. L'information de perception spatiale, *SI*, est fondée sur le filtre de Sobel. Chaque trame vidéo (plan de la luminance) à l'instant n (F_n) est d'abord filtrée par le filtre de Sobel [$\text{Sobel}(F_n)$]. On calcule ensuite, pour chaque trame passant par le filtre de Sobel, l'écart type pour l'ensemble des pixels (std_{space}). Cette opération est répétée pour chaque trame de la séquence vidéo et permet d'obtenir une série temporelle d'informations spatiales sur la scène. La valeur maximale contenue dans la série temporelle (max_{time}) est choisie pour représenter le contenu de la scène en informations spatiales. Ce processus peut être représenté sous forme d'équation, comme suit:

$$SI = \max_{time} \{std_{space} [\text{Sobel}(F_n)]\}$$

Information de perception temporelle (TI): grandeur représentative du degré de changements temporels d'une séquence vidéo. Cette valeur augmente généralement avec l'animation des séquences. Cette grandeur n'est pas censée mesurer l'entropie ni être associée aux informations définies en théorie de la communication.

On calcule la grandeur d'information temporelle, *TI*, en tant que valeur maximale dans le temps (max_{time}) de l'écart type pour l'ensemble de l'espace (std_{space}) de $M_n(i, j)$ pour tous les i et tous les j , soit:

$$TI = \max_{time} \{std_{space} [M_n(i, j)]\}$$

où $M_n(i, j)$ est la différence entre les pixels au même point dans la trame mais appartenant à deux trames successives, soit:

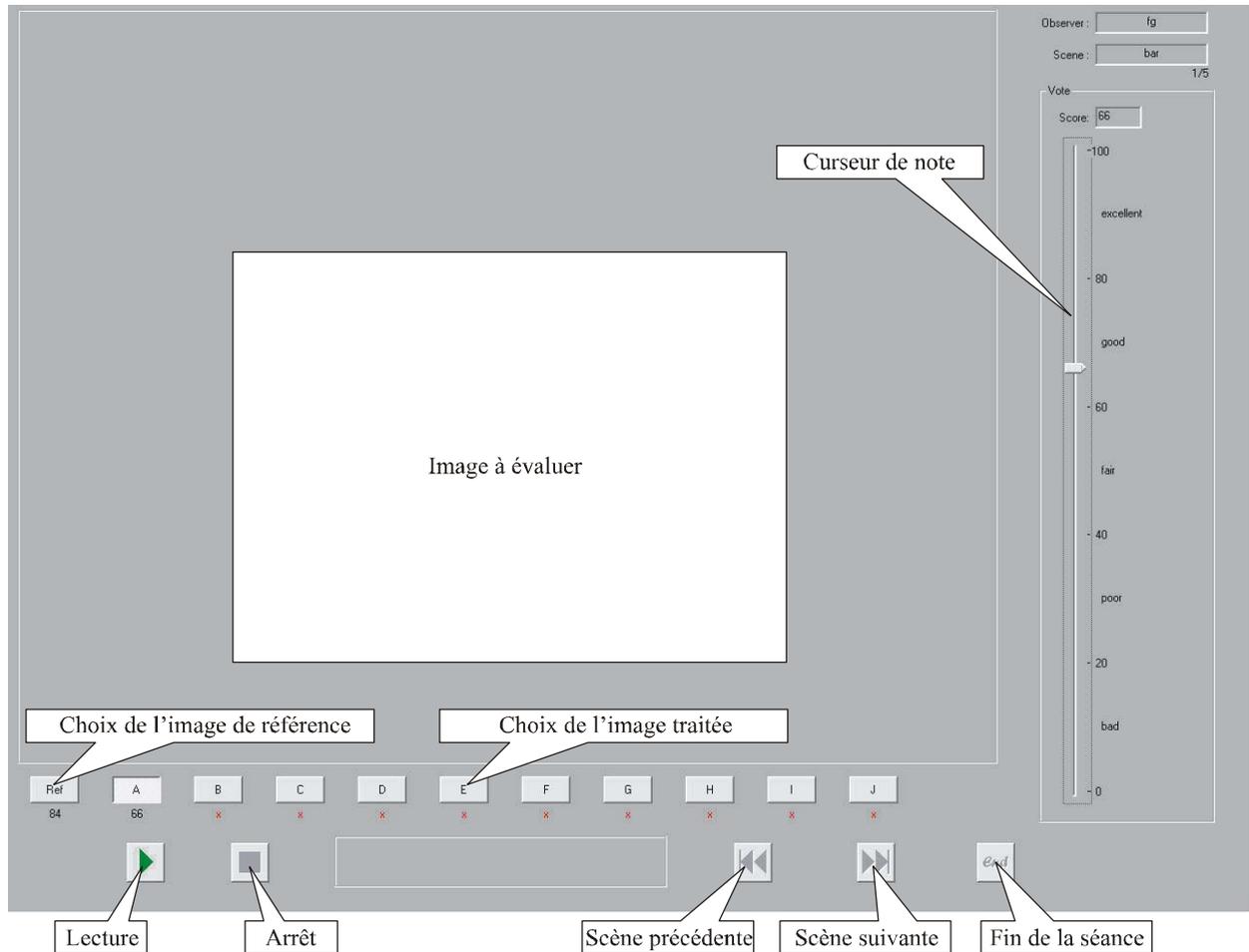
$$M_n(i, j) = F_n(i, j) - F_{n-1}(i, j)$$

où $F_n(i, j)$ est le pixel situé dans la i ème ligne et dans la j ème colonne de la n ème trame dans le temps.

NOTE 1 – Pour les scènes comportant des coupes, deux valeurs peuvent être indiquées: l'une où la coupe est incluse dans la grandeur d'information temporelle, l'autre où elle en est exclue.

Appendice 2 à l'Annexe 1

Exemple d'interface pour la méthode SAMVIQ



1788-02

Appendice 3 à l'Annexe 1

Termes et définitions

Algorithme	Une ou plusieurs opérations de traitement d'image
AVI	Entrelacement audio vidéo (<i>audio video interleaved</i>)
CCD	Dispositif à transfert de charge (<i>charge coupled device</i>)
CI	Intervalle de confiance (<i>confidence interval</i>)

CIF	Format intermédiaire commun (<i>common intermediate format</i>) (format d'image défini dans la Recommandation H.261 pour des visiophones: 352 lignes × 288 pixels)
CRT	Tube cathodique (<i>cathode ray tube</i>)
DSCQS	Méthode à double stimulus utilisant une échelle de qualité continue (<i>double stimulus using a continuous quality scale method</i>)
DSIS	Méthode à double stimulus utilisant une échelle de dégradation (<i>double stimulus using an impairment scale method</i>)
LCD	Affichage à cristaux liquides (<i>liquid crystal display</i>)
MOS	Note moyenne d'opinion (<i>mean opinion score</i>)
PDP	Ecran à plasma (<i>plasma display panel</i>)
PS	Segment de programme (<i>programme segment</i>)
QCIF	Quart de CIF (<i>quarter CIF</i>) (format d'image défini dans la Recommandation H.261 pour des visiophones: 176 lignes × 144 pixels)
SAMVIQ	Évaluation subjective de la qualité vidéo multimédia (<i>subjective assessment of multimedia video quality</i>)
SC	Méthode de comparaison de stimulus (<i>stimulus comparison method</i>)
Scène	Contenu audiovisuel
sdt	Ecart type (<i>standard deviation</i>)
Séquence	Scène avec traitement combiné ou sans traitement
SI	Information spatiale (<i>spatial information</i>)
SIF	Format intermédiaire standard (<i>standard intermediate format</i>) [formats d'image définis dans la norme ISO 11172 (MPEG-1): 352 lignes × 288 pixels × 25 images/s et 352 lignes × 240 pixels × 30 images/s]
S/N	Rapport signal sur bruit (<i>signal-to-noise ratio</i>)
SP	Présentation simultanée (<i>simultaneous presentation</i>)
SQCIF	Sub-QCIF
SS	Méthode à un seul stimulus (<i>single stimulus method</i>)
SSCQE	Méthode d'évaluation continue de la qualité avec stimulus unique (<i>single stimulus using a continuous quality evaluation method</i>)
TI	Information temporelle (<i>temporal information</i>)
TP	Présentation de l'essai (<i>test presentation</i>)
TS	Séance d'essai (<i>test session</i>)
VTR	Magnétoscope (<i>video tape recorder</i>)
