

RECOMMANDATION UIT-R BT.1663

Méthodes d'évaluation, par visionnage d'experts de la qualité des systèmes d'imagerie numérique pour projection en salle sur grand écran¹

(Question UIT-R 15/6)

(2003)

L'Assemblée des radiocommunications de l'UIT,

considérant

- a) que l'UIT, ainsi que d'autres organismes de normalisation internationaux, étudient un nouveau service dénommé imagerie numérique sur grand écran (LSDI, *large screen digital imagery*);
- b) que l'on trouvera probablement plusieurs applications à la LSDI;
- c) qu'il sera nécessaire de spécifier des critères de qualité de fonctionnement et de vérifier que les solutions techniques considérées pour chaque application sont compatibles avec les critères de qualité de fonctionnement correspondant à l'application considérée;
- d) que ce type de vérification imposera nécessairement des évaluations subjectives effectuées dans des conditions scientifiques rigoureuses;
- e) que la capacité de détection de tel ou tel type de dégradation peut différer d'une méthode d'essai subjectif à l'autre;
- f) que les méthodes d'évaluation subjective spécifiées dans la Recommandation UIT-R BT.500 font intervenir des observateurs non experts et sont coûteuses en temps et en argent, compte tenu du grand nombre de notes d'évaluation dont il faut disposer pour distinguer les niveaux relatifs de qualité de fonctionnement des différents systèmes;
- g) qu'une nouvelle méthode d'évaluation subjective faisant intervenir un petit nombre de visionneurs experts est désormais proposée, qui permettra de différencier la qualité de fonctionnement des systèmes tout en offrant des procédures plus rapides et moins onéreuses,

recommande

1 d'utiliser la méthode d'évaluation décrite dans l'Annexe 1 de la présente Recommandation, méthode faisant intervenir des visionneurs experts, pour l'évaluation subjective des applications LSDI, lorsque, pour des raisons de temps ou de budget, il n'est pas possible de recourir aux méthodes d'évaluation par des observateurs non experts spécifiées dans la Recommandation UIT-R BT.500 et lorsque la sensibilité de la méthode exposée dans ladite Annexe est suffisante pour différencier les systèmes évalués,

¹ L'imagerie numérique sur grand écran (LSDI) est une famille de systèmes d'imagerie numérique qui peuvent être utilisés pour des programmes tels que séries télévisées, des pièces de théâtre, des manifestations sportives, des concerts, des événements culturels, etc., de la production à la présentation sur grand écran avec une qualité haute résolution dans des salles de cinéma, des salles de spectacle ou d'autres lieux convenablement équipés.

recommande en outre

- 1 dans le cadre de la procédure d'essai, de mener des études pour vérifier la sensibilité de la méthode d'essai décrite en Annexe 1;
- 2 afin d'améliorer la présente Recommandation, de mener d'autres études sur le traitement des résultats;
- 3 d'encourager les responsables des essais et les organisations qui procèdent aux essais de mettre à la disposition des administrations et des Membres de Secteur de l'UIT tous les matériaux et outils d'essais mis au point (tels que des logiciels pour présenter des éléments d'image côte à côte ou en miroirs) pour faciliter à l'avenir les essais effectués par d'autres organisations.

Annexe 1

Evaluation, par visionnage d'experts, de la qualité des systèmes d'imagerie numérique pour projection LSDI en salle

1 Introduction

Depuis quelques années, on recourt de plus en plus souvent à un visionnage d'experts pour vérifier rapidement la qualité de fonctionnement d'applications vidéo génériques.

La présente Annexe décrit une méthode d'essai qui fait intervenir des visionneurs experts et qui permettra d'obtenir des résultats cohérents d'un laboratoire à l'autre, avec un petit nombre d'observateurs experts.

2 Pourquoi une nouvelle méthode faisant intervenir un «visionnage d'experts»

Il est utile à ce stade de souligner les avantages découlant de l'application de la méthode proposée.

Tout d'abord, un essai d'évaluation subjective fait nécessairement intervenir le plus souvent au moins 15 observateurs choisis parmi des personnes «non expertes», ce qui allonge la durée des essais proprement dits et impose de rechercher constamment de nouveaux observateurs. Ce nombre élevé de visionneurs est nécessaire pour obtenir la sensibilité requise afin de différencier et de classer les différents systèmes ou de les déclarer comme étant équivalents, avec le degré de confiance voulu.

En deuxième lieu, les essais traditionnels, reposant sur des observateurs non experts, peuvent ne pas révéler des différences qui pourraient devenir manifestes, même pour des personnes non expertes, au bout d'un certain temps d'observation.

Troisièmement, les méthodes d'évaluation traditionnelles consistent généralement à mesurer la qualité (ou les différences de qualité), mais non pas à identifier directement les artefacts ou les autres phénomènes physiques qui sont à la base des particularités relevées.

La méthode proposée ci-après doit apporter une solution à ces trois problèmes.

3 Définition de l'expression «visionneur expert»

Dans la présente Recommandation, un «visionneur expert» est par définition une personne qui connaît les équipements utilisés dans l'évaluation, qui sait «ce qu'il faut rechercher» et qui peut avoir une connaissance approfondie de la structure détaillée de l'algorithme utilisé pour traiter les programmes vidéo considérés. En tout état de cause, un «visionneur expert» est une personne ayant une longue expérience des méthodes d'évaluation de la qualité, tout en étant un professionnel du domaine spécifique. Par exemple, lorsque l'on voudra évaluer des programmes LSDI, on sélectionnera comme «visionneurs experts» des spécialistes de la production ou de la postproduction de films cinématographiques ou de programmes vidéo de haute qualité (directeurs de la photographie, spécialistes de l'étalonnage colorimétrique, etc.), et l'on tiendra compte de la capacité des candidats à formuler des jugements subjectifs précis sur la qualité de l'image et les artefacts de compression.

4 Sélection des évaluateurs

Dans les évaluations faisant intervenir des visionneurs experts, il s'agit de demander à des spécialistes d'évaluer la qualité de l'image visionnée et/ou la visibilité des dégradations éventuelles.

Un groupe de visionneurs experts est composé de cinq à six personnes: avec un petit nombre d'observateurs, il est plus facile de procéder à la sélection et de décider rapidement.

En fonction des besoins de l'expérience, on pourra utiliser plusieurs groupes de visionneurs, rassemblés en un pool (représentant par exemple plusieurs laboratoires).

Etant entendu que les visionneurs experts peuvent avoir tendance à «orienter» leur jugement lorsqu'ils évaluent leurs propres technologies, on évitera d'inclure dans les groupes des personnes ayant directement participé à la mise au point du système étudié.

Il conviendra de vérifier l'acuité visuelle, normale ou après correction, de tous les observateurs (Test de Snellen) ainsi que leur vision des couleurs (Test d'Ishihara).

5 Séquences d'essai

Les séquences d'essai seront choisies en fonction de la gamme des paramètres de production et des niveaux de difficulté caractéristiques des conditions réelles dans lesquelles le système considéré sera utilisé. On choisira des séquences relativement difficiles à reproduire, mais sans excès. Dans l'idéal, on utilisera entre cinq et sept séquences d'essai.

La méthode de sélection des séquences d'essai pourra aussi dépendre de l'application en fonction de laquelle le système considéré a été mis au point.

A cet égard, les lignes qui suivent ne comporteront aucune autre indication quant à la procédure de choix des séquences d'essai: le responsable des essais choisira en fonction des éléments précités.

6 Conditions d'observation

Les conditions d'observation, qui seront décrites en détail dans le rapport d'essai, respecteront les éléments indiqués dans le Tableau 1 et seront constantes pendant toute la durée de l'essai.

TABLEAU 1

Condition d'observation	Valeur(s)	
	Minimum	Maximum
Dimension de l'écran (m)	6	16
Distance observateur-écran ⁽¹⁾	1,5 H	2 H
Luminance du projecteur (au centre de l'écran, blanc max.)	10 ftL	14 ftL
Luminance de l'écran (projecteur éteint)		<1/1 000 de la luminance du projecteur

⁽¹⁾ On utilisera la présentation «symétrie verticale» pour une distance observateur-écran voisine de 1,5 H. Lorsque la présentation utilisée sera de type «côte à côte», la distance observateur-écran devra être de l'ordre de 2 H.

7 Méthodologie

7.1 Sessions d'évaluation

Une session d'évaluation (c'est-à-dire l'ensemble des séances tenues par un groupe d'observateurs) comportera deux phases (Phase I et Phase II).

7.1.1 Phase I

La Phase I sera un essai subjectif formel dans des conditions contrôlées (voir le § 6), qui permettra d'obtenir des résultats valides, sensibles et répétitifs. Les visionneurs experts attribuent une note d'observation aux séquences qu'ils visionnent en utilisant l'échelle décrite ci-après. Les membres du groupe ne sont pas autorisés à commenter entre eux ce qu'ils voient, et n'interviennent pas dans la présentation. Pendant cette phase, les experts N'ONT PAS connaissance du système de codage évalué ou de l'ordre de présentation des séquences. Les séquences d'essai sont présentées en ordre aléatoire: ainsi, les observateurs ne peuvent pas être influencés dans leur jugement.

7.1.1.1 Présentation des séquences

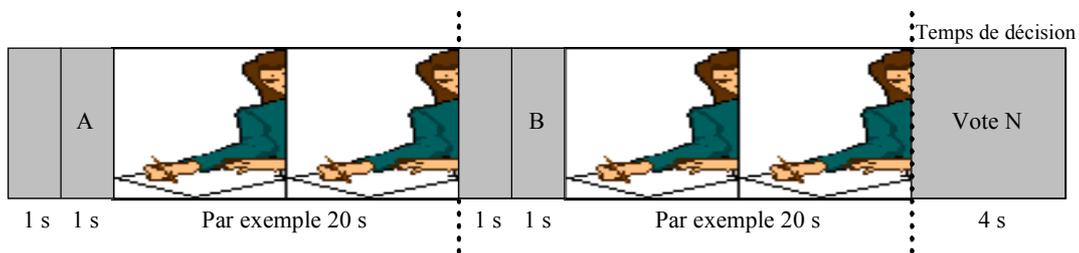
La méthode de présentation combine des éléments de la méthode d'évaluation continue à double stimulus simultané (SDSCE) décrite dans la Recommandation UIT-R BT.500, § 6.4) et des éléments de la méthode de double stimulus à échelle de qualité continue (DSCQS) décrite dans la Recommandation UIT-R BT.500, § 5). A des fins de référence, on pourra dénommer cette méthode «méthode SDS» (double stimulus simultané).

Comme avec la méthode SDSCE, on présentera chaque fois deux images sur un écran partagé. Dans la plupart des cas, l'une des sources sera la référence (image source), l'autre étant la séquence d'essai; dans d'autres cas, les deux images seront issues de la référence. La référence sera la source présentée de façon transparente (c'est-à-dire sans compression autre que la compression inhérente au support d'enregistrement). La séquence d'essai sera obtenue à partir de la source, mais traitée par l'un des systèmes évalués. Le débit et/ou le niveau de qualité correspondront aux valeurs spécifiées pour l'essai. Mais contrairement aux conditions d'application de la méthode SDSCE, les observateurs n'auront pas connaissance de la nature des deux séquences.

La présentation en écran partagé se fera soit de façon traditionnelle (sans effet miroir) soit en «symétrie verticale», par retournement horizontal de l'image occupant la partie droite de l'écran. Les images de départ étant des images plein écran, on ne pourra afficher que la moitié de chaque image, et l'on veillera donc à montrer la même moitié sur chaque demi-écran à chaque présentation.

Comme avec la méthode DSCQS, on présentera chaque paire d'images deux fois de suite, la première fois pour permettre aux visionneurs de l'observer et de l'examiner soigneusement, la seconde pour la confirmation et la notation. Chaque séquence durera entre 15 et 30 s. Les séquences pourront être précédées d'une «étiquette», ce qui facilitera le travail des visionneurs (voir la Fig.1, représentant une paire d'images présentées sans effet miroir).

FIGURE 1

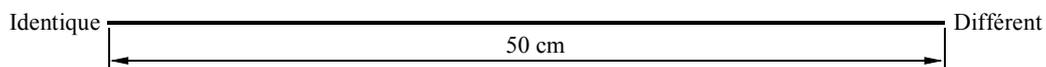


1663-01

7.1.1.2 Echelle d'évaluation

Le critère d'acceptation d'une application LSDI est que la séquence d'essai (image comprimée) ne puisse pas être distinguée de la référence. L'évaluation des systèmes à l'essai peut se faire au moyen des nombreuses méthodes d'évaluation traditionnellement utilisées. On propose par exemple les échelles de comparaison de stimulus de la Recommandation UIT-R BT.500, § 6.2. Comme exemple spécifique d'échelle d'évaluation, on propose l'échelle continue (sans catégorie) IDENTIQUE/DIFFERENT décrite au § 6.2.4.2 de la Recommandation UIT-R BT.500.

FIGURE 2



1663-02

7.1.1.3 Session d'évaluation

La session, qui pourra comprendre plusieurs séances selon le nombre de conditions d'essai, portera sur deux types d'observations: observations de test et observations de vérification. Dans une observation de test, une moitié de l'écran affiche la référence et l'autre la séquence d'essai. Dans une observation de vérification, les deux moitiés de l'écran affichent la référence. La vérification permet de déterminer si le jugement de l'observateur est biaisé.

Pour chaque système à l'essai, les observations suivantes devront être faites sur chaque séquence:

TABLEAU 2

Partie gauche de l'affichage	Partie droite de l'affichage
Moitié gauche de la référence	Moitié gauche de la séquence d'essai
Moitié droite de la référence	Moitié droite de la séquence d'essai
Moitié gauche de la séquence d'essai	Moitié gauche de la référence
Moitié droite de la séquence d'essai	Moitié droite de la référence

De préférence, chaque configuration sera présentée au moins deux fois. Pour chaque système, et pour chaque séquence d'essai, les observations de vérification suivantes devront être faites:

TABLEAU 3

Panneau gauche	Panneau droit
Moitié gauche de la référence	Moitié gauche de la référence
Moitié droite de la référence	Moitié droite de la référence

Ici encore, chaque configuration sera présentée au moins deux fois de préférence.

Les séances de la session d'essai ne devront pas durer plus d'une heure, et une pause de 15 min sera prévue entre deux séances consécutives. Les observations d'essai et les observations de vérification faites sur les combinaisons sortie codec/séquence d'essai seront réparties entre les séances de façon pseudo-aléatoire. Malgré le surcroît de complexité qui en résultera, il sera préférable de prévoir certaines restrictions dans la procédure. Par exemple, dans une session comportant quatre séances, on pourrait affecter de façon aléatoire chacune des quatre observations d'essai correspondant à une combinaison signal codec/signal d'essai à une position choisie de façon aléatoire dans l'une des séances. Cette méthode présente l'avantage de garantir une bonne répartition, sur la totalité de la session, de toutes les observations d'essai.

7.1.1.4 Traitement des notes d'évaluation

Pour une observation donnée, la note est la distance entre la position «IDENTIQUE» de l'échelle et la marque faite par l'observateur sur une échelle de 0-100. Les résultats sont convertis en notes moyennes d'opinion qui permettent de classer les systèmes testés. Selon le nombre d'observations par système (observateurs \times séquences d'essai \times répétitions), on pourra procéder à une analyse de variance (ANOVA)². Les résultats des observations de vérification pourront servir à déterminer un «seuil de hasard».

² Un total de 10 à 20 observations concernant un élément de moindre importance suffit à justifier un traitement statistique déductif, une analyse de variance par exemple.

7.1.2 Phase II

La Phase II a notamment pour objet de donner une meilleure précision dans le classement relatif des résultats de la Phase I, dont la précision et la fiabilité dépendent du nombre d'observateurs et/ou des séances d'évaluation. Autre objectif important, il s'agit de susciter des commentaires en ce qui concerne les éléments sur lesquels les images sont perçues comme différentes et sur lesquels les évaluations de la Phase I ont été faites.

Dans la Phase II, les experts examinent les séquences qui ont été visionnées et, dans ce cas, ils peuvent formuler des commentaires sur les séquences projetées, visionner telle ou telle partie d'une séquence ou la totalité d'une séquence autant de fois qu'ils le souhaitent, soit pour l'examiner soit pour démontrer un point, ce qui leur permet de parvenir à un consensus dans leur évaluation et de bien décrire ce qu'ils voient. Le recours aux fonctions «ralenti», «décomposition du mouvement» et «arrêt sur image» est même autorisé si le visionneur le souhaite. Ici, les opérations appelleront parfois l'intervention du responsable de l'essai, qui pourra les coordonner.

7.1.2.1 Regroupement des séquences d'essai

Pour le bon déroulement de la Phase II, il faudra regrouper les séquences d'essai par contenu en différents jeux de séquences de base pour visionnage d'experts (BES, *basic expert viewing set*) comportant les séquences codées dérivées d'une même source et ordonnées en fonction du classement établi pendant la Phase I.

Les séquences d'essai seront classées dans l'ordre croissant des notes moyennes d'opinion. Il y aura autant d'ensembles BES que de séquences différentes utilisées pendant l'essai.

7.1.2.2 Sous-groupe de visionnage d'experts des séquences de base

Pendant une séance de sous-groupe de visionnage d'experts des séquences de base (BEV, *basic expert viewing*), les experts examinent toutes les séquences d'un ensemble de visionnage de base pour experts donné pour confirmer ou modifier le classement établi à l'issue des essais formels de la Phase I: c'est donc l'occasion de confirmer ou de modifier la visibilité relative des différences.

7.1.2.3 Plan de Phase II

Pendant la Phase II, tous les visionnages d'experts des séquences de base (BEV) doivent être effectués. Les experts seront informés de ce que l'ordre de présentation correspond à l'ordre de classement établi pendant la Phase I, mais ils n'auront aucun élément d'information concernant le classement des systèmes proposés.

La Phase II sera un travail de groupe à l'issue duquel les experts formuleront des avis consensuels.

Avant le début de la Phase II, les visionneurs recevront, si possible par écrit, des instructions leur demandant:

- d'examiner toutes les séquences couvertes par le BEV considéré;
- de commenter l'ordre de classement des séquences couvertes par le BEV considéré; en cas de désaccord, un nouveau classement sera établi;
- de commenter chaque cas en formulant des observations détaillées sur la nature des différences éventuellement observées;
- de documenter leur classement, leurs commentaires et leurs observations.

Il appartiendra au responsable de l'essai de rassembler toutes les observations formulées par les groupes et de repérer les divergences éventuelles. Pendant les tests proprement dits, les résultats obtenus par les différents groupes à l'issue de la Phase I et de la Phase II resteront confidentiels, de telle sorte que les groupes suivants ne puissent pas être influencés.

Lorsque cela sera possible, le responsable de l'essai pourra signaler les divergences et rechercher une solution au problème en demandant d'autres évaluations sur les points controversés. Cette dernière disposition a pour objet d'assurer un consensus général.

8 Rapport

Le rapport final de la série d'essais, établi par le responsable, comprendra les informations suivantes:

- résultats de la Phase I (y compris les tableaux de notes moyennes d'opinion, et résultats, selon le cas, des analyses statistiques effectuées);
 - observations formulées par les experts pendant la Phase II;
 - observations relatives à toute réévaluation d'un classement;
 - toutes informations utiles sur les conditions d'observation, les caractéristiques des signaux source, le traitement du signal, les caractéristiques du projecteur, les réglages du projecteur, les données de réglage chromatique, la sélection des visionneurs et les conditions d'essais proprement dites;
 - description détaillée des caractéristiques du système d'affichage (courbes de moyenne des temps entre défaillances, etc.);
 - résumé et conclusions.
-