

Remplacée par une version plus récente



UNION INTERNATIONALE DES TÉLÉCOMMUNICATIONS

UIT-T

SECTEUR DE LA NORMALISATION
DES TÉLÉCOMMUNICATIONS
DE L'UIT

P.910

(08/96)

SÉRIE P: QUALITÉ DE TRANSMISSION
TÉLÉPHONIQUE

Qualité audiovisuelle dans les services multimédias

**Méthodes subjectives d'évaluation de la qualité
vidéographique pour les applications
multimédias**

Recommandation UIT-T P.910
Remplacée par une version plus récente

(Antérieurement Recommandation du CCITT)

Remplacée par une version plus récente

RECOMMANDATIONS UIT-T DE LA SÉRIE P QUALITÉ DE TRANSMISSION TÉLÉPHONIQUE

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

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

Remplacée par une version plus récente

RECOMMANDATION UIT-T P.910

MÉTHODES SUBJECTIVES D'ÉVALUATION DE LA QUALITÉ VIDÉOGRAPHIQUE POUR LES APPLICATIONS MULTIMÉDIAS

Résumé

La présente Recommandation est destinée à définir des méthodes subjectives et non interactives permettant d'évaluer la qualité d'images vidéo numériques codées à des débits faibles et moyens (inférieurs ou égaux à 2 Mbit/s) pour des applications telles que la vidéophonie, la vidéoconférence et les applications de stockage/extraction.

Les points ci-après seront pris en considération dans la présente Recommandation:

- montage en laboratoire pour l'établissement des séquences d'essai;
- montage en laboratoire pour la réalisation des évaluations subjectives;
- caractéristiques des séquences d'essai;
- méthodes d'essai et modèles expérimentaux;
- analyse des données.

La présente Recommandation ne traite pas de certains sujets qui sont déjà examinés dans d'autres Recommandations, à savoir:

- les conditions vidéo de référence définies dans la Recommandation P.930;
- les méthodes de réglage des récepteurs de contrôle décrites dans le Rapport 1221 du CCIR;
- les méthodes d'essai interactives définies dans la Recommandation P.920.

Source

La Recommandation UIT-T P.910, élaborée par la Commission d'études 15 (1993-1996) de l'UIT-T, a été approuvée le 30 août 1996 selon la procédure définie dans la Résolution n° 1 de la CMNT.

Remplacée par une version plus récente

AVANT-PROPOS

L'UIT (Union internationale des télécommunications) est une institution spécialisée des Nations Unies dans le domaine des télécommunications. L'UIT-T (Secteur de la normalisation des télécommunications) est un organe permanent de l'UIT responsable de l'établissement de normes mondiales (Recommandations) en matière de télécommunications, et auquel participent quelque 179 pays membres, 84 exploitations de télécommunications reconnues, 145 organisations scientifiques et industrielles et 38 organisations internationales.

L'approbation des Recommandations par les membres de l'UIT-T s'effectue selon la procédure définie dans la Résolution n° 1 de la Conférence mondiale de normalisation des télécommunications (CMNT) (Helsinki, 1993) De plus, la CMNT, qui se réunit tous les quatre ans, approuve les Recommandations qui lui sont soumises et établit le programme d'études pour la période suivante.

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

NOTE

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

© UIT 1996

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

Remplacée par une version plus récente

TABLE DES MATIÈRES

	Page
1	Domaine d'application..... 1
2	Références..... 1
3	Termes et définitions 2
4	Abréviations..... 3
5	Signal de source 4
5.1	Environnement d'enregistrement 4
5.2	Système d'enregistrement 4
5.2.1	Caméra..... 4
5.2.2	Echantillonnage et stockage du signal vidéo..... 5
5.3	Caractéristiques des scènes..... 5
5.3.1	Mesure des informations de perception spatiale 6
5.3.2	Mesure des informations de perception temporelle..... 6
6	Méthodes d'essai et modèles expérimentaux 6
6.1	Evaluation par catégories absolues (ACR)..... 6
6.2	Evaluation par catégories de dégradation (DCR)..... 7
6.3	Méthode de comparaison par paires (PC)..... 8
6.4	Comparaison des méthodes d'essai..... 9
6.5	Conditions de référence 10
6.6	Modèles expérimentaux..... 10
7	Procédures d'évaluation 10
7.1	Conditions de visualisation..... 10
7.2	Système de traitement et de reproduction..... 11
7.3	Visualisateurs..... 11
7.4	Instructions données aux visualisateurs et session d'apprentissage..... 12
8	Analyse statistique et compte rendu des résultats..... 12
Annexe A	Détails relatifs à la caractérisation des séquences d'essai 13
A.1	Filtre de Sobel..... 13
A.2	Mode d'utilisation des informations SI et TI pour la sélection des séquences d'essai. 14
A.3	Exemples..... 15
Annexe B	Echelles d'évaluation supplémentaires 16
B.1	Echelles d'évaluation 16
B.2	Dimensions d'évaluation additionnelles 18

Remplacée par une version plus récente

	Page
Annexe C - Présentation simultanée de paires de séquences.....	20
C.1 Introduction.....	20
C.2 Synchronisation	20
C.3 Conditions de visualisation.....	20
C.4 Présentations	21
Appendice I - Séquences d'essai	21
Appendice II - Instructions pour les essais de visualisation	22
II.1 Méthode ACR.....	22
II.2 Méthode DCR.....	23
II.3 Méthode PC	23

Remplacée par une version plus récente

Recommandation P.910

MÉTHODES SUBJECTIVES D'ÉVALUATION DE LA QUALITÉ VIDÉOGRAPHIQUE POUR LES APPLICATIONS MULTIMÉDIAS

(Genève, 1996)

1 Domaine d'application

La présente Recommandation est destinée à définir des méthodes subjectives et non interactives permettant d'évaluer la qualité d'images vidéo numériques codées à des débits faibles et moyens (inférieurs ou égaux à 2 Mbit/s) pour des applications telles que la vidéophonie, la vidéoconférence et les applications de stockage/extraction. Les méthodes décrites dans les articles qui suivent conviennent également pour l'évaluation de l'influence d'erreurs de transmission sur de tels systèmes vidéo. Elles pourront donc servir plusieurs objectifs différents, y compris (par exemple) la sélection des algorithmes, la hiérarchisation des systèmes vidéo selon leurs performances ainsi que l'évaluation du niveau de qualité au cours d'une connexion vidéo.

2 Références

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

- [1] Recommandation UIT-T P.930 (1996), *Principes d'un système de dégradation de référence pour signaux vidéo.*
- [2] Recommandation UIT-T P.920 (1996), *Méthodes d'essai interactives pour communications audiovisuelles.*
- [3] Recommandation UIT-R BT.601-4 (1994), *Paramètres de codage de télévision numérique pour studios.*
- [4] Recommandation UIT-R BT.500-6 (1994), *Méthodologie d'évaluation subjective de la qualité des images de télévision.*
- [5] Publication 268-13 de la CEI: 1985, *Equipements pour systèmes électroacoustiques - Partie 13: essais d'écoute des haut-parleurs.*
- [6] CCITT: *Manuel de téléphonométrie*, Genève, 1993.
- [7] Recommandation UIT-T P.800 (1996), *Méthodes d'évaluation subjective de la qualité de transmission.*
- [8] Recommandation UIT-R BT.814-1 (1994), *Spécifications et méthodes de réglage de la brillance et du contraste des dispositifs de visualisation.*
- [9] Recommandation UIT-R BT.815-1 (1994), *Spécification d'un signal de mesure du contraste des dispositifs de visualisation.*

Remplacée par une version plus récente

- [10] Rapport BT.1213 du CCIR: 1990, *Images et séquences d'essai pour l'évaluation subjective des codecs numériques - Annexe au Volume XI - Partie 1.*
- [11] Recommandation 567-3 du CCIR: 1990, *Qualité de transmission des circuits de télévision destinés à être utilisés dans les communications internationales - Volume XII.*
- [12] Recommandation UIT-R BT.812 (1994), *Evaluation subjective de la qualité des images alphanumériques et graphiques en télétexte et dans des services similaires.*

3 Termes et définitions

La présente Recommandation définit les termes suivants.

3.1 gamma: paramètre qui décrit la séparation entre niveaux de gris sur un écran d'affichage. La relation entre la luminance et la tension du signal d'entrée n'est pas linéaire, la tension étant portée à une puissance gamma. Afin de compenser cette non-linéarité, on applique généralement dans la caméra un facteur de correction qui est une fonction inverse du gamma. Le gamma possède également une influence sur le rendu des couleurs.

3.2 essais d'optimisation: essais subjectifs qui sont normalement effectués soit au cours du développement soit au cours de la normalisation d'un algorithme ou système nouveau. Le but de ces essais est d'évaluer la performance de nouveaux outils afin d'optimiser les algorithmes ou les systèmes étudiés.

3.3 essais de qualification: essais subjectifs qui sont normalement effectués afin de comparer les performances de systèmes ou équipements du commerce. Ces essais doivent être exécutés dans les conditions expérimentales qui sont les plus représentatives que possible des conditions réelles d'emploi.

3.4 information de perception spatiale (SI) (spatial perceptual information): grandeur qui indique généralement le degré de détail spatial d'une image. Cette valeur est habituellement plus élevée pour des scènes à coordonnées spatiales plus complexes. Cette grandeur n'est pas censée mesurer l'entropie ni être associée aux informations définies en théorie de la communication. Voir 5.3.1 pour l'équation relative à cette information.

3.5 information de perception temporelle (TI) (temporal perceptual information): grandeur qui indique généralement le degré de changements temporels d'une séquence vidéo. Cette valeur est habituellement plus élevée pour des séquences très animées. Cette grandeur n'est pas censée mesurer l'entropie ni être associée aux informations définies en théorie de la communication. Voir 5.3.2 pour l'équation relative à cette information.

3.6 transparence (fidélité): concept décrivant la performance d'un codec ou d'un système par rapport à un système de transmission idéal et sans dégradation.

Deux types de transparence peuvent être définis:

Le premier type décrit le degré de conformité du signal traité avec le signal d'entrée, ou avec le signal idéal, au moyen d'un critère mathématique. S'il n'y a aucune différence, le système est totalement transparent. Le deuxième type décrit le degré de conformité du signal traité avec le signal d'entrée, ou avec le signal idéal, pour un observateur humain. Si aucune différence ne peut être perçue dans une quelconque situation expérimentale, le système est transparent à la perception. Le qualificatif "transparent" sans référence explicite à un critère sera utilisé pour les systèmes qui sont transparents à la perception.

3.7 réplication: répétition du même état de circuit (avec le même matériel de source) pour le même sujet.

Remplacée par une version plus récente

3.8 fiabilité d'un essai subjectif:

- a) la fiabilité intra-individuelle ("à l'intérieur d'un même sujet") représente la constance d'un certain nombre d'évaluations répétées sur le même sujet dans la même condition d'essai;
- b) la fiabilité inter-individuelle ("entre sujets différents") représente la constance d'un certain nombre d'évaluations de sujets différents dans la même condition d'essai.

3.9 validité d'un essai subjectif: concordance entre la valeur moyenne d'évaluations obtenues lors d'un essai et la valeur vraie que cet essai vise à mesurer.

3.10 conditions de référence: conditions fictives ajoutées aux conditions d'essai afin d'offrir une base aux évaluations issues d'expériences différentes.

3.11 référence explicite (référence de source): condition utilisée par les évaluateurs comme référence pour exprimer leur opinion, lorsqu'ils utilisent la méthode d'évaluation DCR. Cette référence est d'abord affichée dans chaque paire de séquences. Le format de la référence explicite est habituellement celui qui est utilisé à l'entrée des codecs en essai (par exemple UIT-R BT.601-4, CIF, QCIF, SIF, etc.). Dans le corps de la présente Recommandation, les mots "explicite" et "de source" seront omis chaque fois que le contexte précise le sens de la "référence".

3.12 référence implicite: condition utilisée par les évaluateurs comme référence pour exprimer leur opinion au sujet du matériel mis à l'essai, lorsqu'ils utilisent la méthode d'évaluation ACR. Si la référence implicite est suggérée par l'expérimentateur, elle doit être bien connue de tous les évaluateurs (par exemple: systèmes de télévision conventionnels, réalité).

4 Abréviations

La présente Recommandation utilise les abréviations suivantes.

ACR évaluation par catégories absolues (*absolute category rating*)

CCD dispositif à couplage de charge (*charge coupled device*)

CI intervalle de confiance (*confidence interval*)

CIF format intermédiaire commun (format d'image défini dans la Recommandation H.261 pour la vidéophonie: 325 lignes x 288 pixels) (*common intermediate format*)

CRT tube à rayons cathodiques (*cathode ray tube*)

DCR évaluation par catégories de dégradation (*degradation category rating*)

%GOB pourcentage de bons ou de meilleurs (proportions de bons et d'excellents) (*percent of good or better*)

LCD affichage par cristaux liquides (*liquid crystal display*)

MOS note moyenne d'opinion (*mean opinion score*)

PC comparaison par paires (*pair comparison*)

%POW pourcentage de médiocres ou de pires (proportion de notes médiocres et mauvaises) (*percent of poor or worse*)

QCIF quart de CIF (format d'image défini dans la Recommandation H.261 pour la vidéophonie: 176 lignes x 144 pixels)

S/N rapport signal sur bruit (*signal to noise ratio*)

SI information spatiale (*spatial information*)

Remplacée par une version plus récente

SIF	format intermédiaire normalisé (format d'image défini dans l'ISO-11172 (MPEG-1): 352 lignes x 288 pixels x 25 trames/s et 352 lignes x 240 pixels x 30 trames/s) (<i>standard intermediate format</i>)
SP	présentation simultanée (<i>simultaneous presentation</i>)
std	écart type (<i>standard deviation</i>)
TI	information temporelle (<i>temporal information</i>)
VTR	magnétoscope (<i>video tape recorder</i>)

5 Signal de source

Pour commander les caractéristiques du signal de source, il convient de définir les séquences d'essai en fonction de l'objectif de celui-ci puis de les enregistrer sur un système de stockage numérique. Lorsque l'expérimentateur souhaite comparer les résultats issus de différents laboratoires, il est nécessaire d'utiliser un ensemble commun de séquences d'essai pour éliminer toute autre source de variation.

5.1 Environnement d'enregistrement

On peut placer une (ou des) source(s) d'éclairage (lampes à incandescence ou à fluorescence) au-dessus ou à côté de la caméra. Lors de la mise en place des lampes, tenir compte du fait que l'éclairage par le haut est plus représentatif des bureaux et qu'il convient d'y faire appel pour la mise en scène d'environnements bureautiques. Il y a lieu d'éviter les lampes de studio et autres sources non typiques.

Les conditions d'éclairage lumineux de la salle, dans le champ de vision, pourront varier de 100 lux à environ 10 000 lux pour l'usage à l'intérieur des bâtiments. Il faut tenir compte de la variation (fréquence du courant alternatif) de l'éclairage (par fluorescence) car cela peut provoquer un scintillement dans la séquence vidéo enregistrée.

Il y a lieu de vérifier et d'indiquer avec soin les conditions d'éclairage, les couleurs des murs, le facteur de réflexion de la surface, etc.

5.2 Système d'enregistrement

5.2.1 Caméra

Il convient d'enregistrer les séquences d'images avec une caméra à capteur CCD de haute qualité.

Le rapport S/N du signal d'entrée vidéo peut avoir une incidence importante sur la performance du codec.

Pour définir le signal d'entrée vidéo, il convient de spécifier les points suivants:

- la dynamique des signaux Y U V (luminance + différence de couleur);
- le facteur de correction gamma (qui doit être égal à 0,45);
- la bande passante/les pentes des filtres;
- la sensibilité de la caméra en conditions de très faible luminosité ainsi que les caractéristiques de l'éventuelle commande automatique de gain (AGC).

On mesurera le rapport S/N pondéré conformément à la Recommandation 567-3 du CCIR, Partie C, paragraphe 3.2.1 [11]. La valeur de ce rapport sera supérieure à 45 dB (eff.).

Remplacée par une version plus récente

L'instabilité ou la gigue des signaux d'horloge pourrait provoquer des effets perturbateurs. Une stabilité minimale de $0,5 \times 10^{-6}$ est requise pour la base de temps de la caméra.

On pourra faire appel à des systèmes à distance focale fixe ou variable. Pour les terminaux de table, une focale comprise entre 30 cm et 120 cm est raisonnable, tandis que pour les systèmes multi-utilisateurs une focale comprise entre 50 cm et l'infini serait sans doute mieux appropriée. Pour tenir compte de la variation d'éclairage lumineux dans la salle d'enregistrement, il convient d'utiliser soit un diaphragme réglable soit des filtres neutres (gris). La caméra devra posséder un équilibrage automatique des blancs de façon à réaliser l'adaptation à la température de couleur de la source lumineuse. La correction de la température du blanc peut aller de 2700° K (utilisation à l'intérieur des bâtiments avec ampoule à incandescence) à 6500° K (température de la lumière du jour avec ciel couvert).

5.2.2 Echantillonnage et stockage du signal vidéo

Les signaux de source vidéo fournis par la caméra seront échantillonnés conformément à la Partie A de [3]. Afin d'éviter la distorsion du signal de source, celui-ci sera stocké sous forme numérique, par exemple en format 4:2:2 sur ordinateur ou sur magnétoscope D-1.

5.3 Caractéristiques des scènes

La sélection des scènes d'essai est un point important. Les informations de perception spatiale et de perception temporelle sont en particulier des paramètres critiques des scènes. Ces paramètres jouent un rôle crucial pour déterminer le degré de compression vidéo possible et donc le niveau de dégradation qui est subi lorsque la scène est transmise sur une voie de service de transmission numérique à débit fixe. Les scènes d'essai vidéo représentant des expositions ou marchés analogues doivent être choisies de manière que leurs informations de perception spatiale et temporelle soient compatibles avec les services vidéo que la voie de service de transmission est appelée à acheminer. La série de scènes d'essai couvrira la gamme complète des informations spatiales et temporelles intéressant les utilisateurs des dispositifs en essai.

L'Annexe A et les Appendices I et II donnent des détails sur la caractérisation des séquences d'essai ainsi que des exemples de scènes d'essai appropriées.

Le nombre de séquences sera défini en fonction du modèle expérimental. Pour éviter de lasser les observateurs et d'obtenir une fiabilité minimale des résultats, au moins quatre types différents de scènes (c'est-à-dire portant sur des sujets différents) seront choisis pour les séquences.

Les paragraphes ci-dessous présentent des méthodes permettant de quantifier les informations spatiales et temporelles des scènes d'essai sont applicables aux essais de qualité vidéo, aussi bien actuels que futurs. La position de la scène vidéo dans la matrice spatio-temporelle est important parce que la qualité d'une scène vidéo transmise (surtout après son passage par un codec à faible débit) dépend souvent beaucoup de ce type de position. Les valeurs d'information spatiale ou temporelle présentées ici peuvent être utilisées pour assurer une couverture appropriée du plan spatio-temporel.

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 perceptive 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 courbes de répartition des informations dans une séquence d'essai permet aussi de mieux évaluer les scènes comportant des coupes.

Remplacée par une version plus récente

5.3.1 Mesure des informations de perception spatiale

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 [$Sobel(F_n)$]. On calcule ensuite, pour chaque trame passant par le filtre de Sobel, l'écart type de la répartition 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 (temps maximal) est choisie pour représenter le contenu de la scène en informations spatiales. Ce processus peut être représenté sous forme d'équation, comme suit:

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

5.3.2 Mesure des informations de perception temporelle

L'information de perception temporelle (TI) est fondée sur la caractéristique de différence cinétique, $M_n(i,j)$, qui est la différence entre les valeurs des pixels (du plan de luminance) au même point dans l'espace mais à des instants ou trames différents. La différence $M_n(i,j)$ en fonction du temps (n) est définie comme suit:

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

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

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

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

De plus grandes différences cinétiques dans des trames adjacentes produiront des valeurs plus grandes de TI.

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

6 Méthodes d'essai et modèles expérimentaux

Le mesurage de la qualité des images perçues nécessite l'emploi de méthodes de notation subjectives. Pour que de telles mesures soient représentatives, il faut qu'il existe une relation entre les caractéristiques physiques du "stimulus" (qui est en l'occurrence la séquence vidéo présentée aux sujets lors d'un test) et l'amplitude de la sensation provoquée par le stimulus, ainsi que sa nature.

Un certain nombre de méthodes expérimentales ont été validées pour différents objets. Trois méthodes sont recommandées ci-dessous pour des applications utilisant des connexions jusqu'à 2 Mbit/s.

Le choix final d'une de ces méthodes pour une application particulière dépend de plusieurs facteurs, comme le contexte, le but recherché et le point du processus de développement auquel le test doit être exécuté.

6.1 Evaluation par catégories absolues (ACR)

La méthode d'évaluation par catégories absolues consiste à porter un jugement pour ranger indépendamment, dans une échelle de catégories, des séquences d'essai présentées les unes après les autres. (Cette méthode est également appelée "méthode du stimulus unique".)

La méthode spécifie qu'après chaque présentation d'une séquence, les sujets sont invités à évaluer la qualité de celle-ci.

Remplacée par une version plus récente

Le chronogramme pour la présentation du stimulus est décrit sur la Figure 1. Le temps imparti aux votes sera inférieur ou égal à 10 s, selon le mode de scrutin utilisé. La durée de la présentation pourra être réduite ou augmentée, selon le contenu du matériel expérimental.

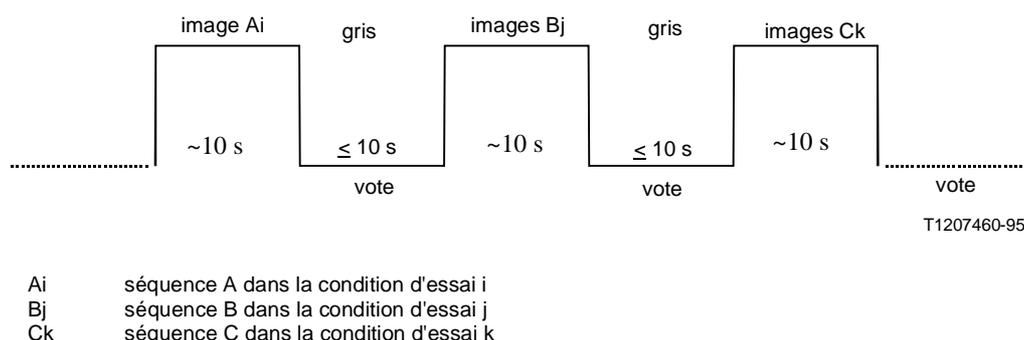


FIGURE 1/P.910

Présentation du stimulus dans la méthode ACR

Pour l'évaluation de la qualité globale, il y a lieu d'utiliser l'échelle à cinq niveaux suivante:

- 5 Excellent
- 4 Bon
- 3 Satisfaisant
- 2 Médiocre
- 1 Mauvais

Si une plus grande puissance de discrimination est requise, on peut faire appel à une échelle à neuf niveaux. L'Annexe B donne des exemples d'échelles numériques ou continues appropriées, ainsi que de dimensions d'évaluation autres que la qualité globale. De telles dimensions peuvent servir à obtenir plus de renseignements sur différents facteurs de qualité perçue lorsque l'évaluation de la qualité globale est presque la même pour certains systèmes en essai, bien que ces systèmes soient clairement perçus comme étant différents.

Dans la méthode ACR, on obtient le nombre de réplifications nécessaires en reprenant les mêmes conditions d'essai à différents instants de l'essai.

6.2 Evaluation par catégories de dégradation (DCR)

L'évaluation par catégories de dégradation implique que les séquences d'essai soient présentées par paires: le premier stimulus présenté dans chaque paire est toujours la référence de source, tandis que le second stimulus est la même source présentée au moyen de l'un des systèmes en essai (cette méthode est également appelée "méthode à double stimulus utilisant une échelle de dégradation").

Lorsque l'on utilise des formats d'image réduits (par exemple CIF, QCIF, SIF), il peut être utile d'afficher simultanément, sur le même moniteur, la référence de source et la séquence d'essai. L'Annexe C donne des directives au sujet de ce mode de présentation.

Le chronogramme pour la présentation du stimulus est décrit sur la Figure 2. Le temps imparti aux votes sera inférieur ou égal à 10 s, selon le mode de scrutin utilisé. La durée de la présentation pourra être réduite ou augmentée, selon le contenu du matériel expérimental.

Remplacée par une version plus récente

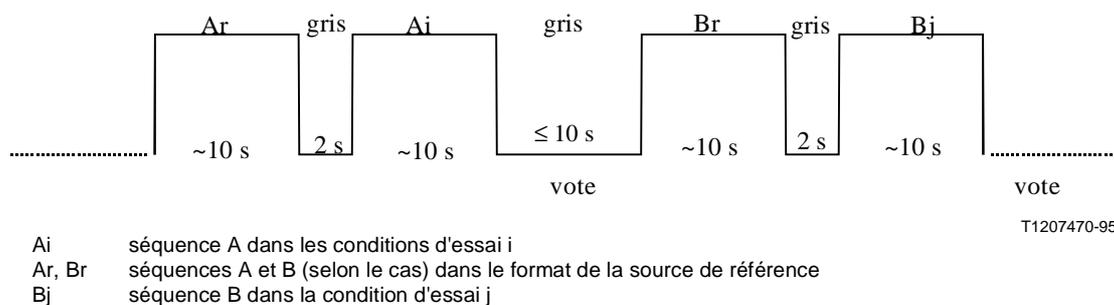


FIGURE 2/P.910

Présentation du stimulus dans la méthode DCR

Dans ce cas, les sujets sont invités à évaluer la dégradation du second stimulus par rapport à la référence.

Pour l'évaluation de la qualité globale, il y a lieu d'utiliser l'échelle à cinq niveaux suivante:

- 5 Imperceptible
- 4 Perceptible mais non gênante
- 3 Légèrement gênante
- 2 Gênante
- 1 Très gênante

Dans la méthode ACR, on obtient le nombre de réplifications nécessaires en reprenant les mêmes conditions d'essai à différents instants de l'essai.

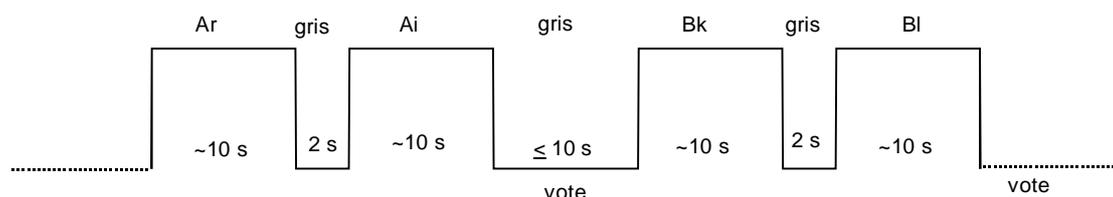
6.3 Méthode de comparaison par paires (PC)

La méthode des comparaisons par paires implique que les séquences d'essai soient présentées en paires. Chaque paire est formée de la même séquence, présentée d'abord au moyen d'un système à l'essai puis au moyen d'un autre système.

Les systèmes en essai (A, B, C, etc.) sont généralement associés selon toutes les $n(n-1)$ combinaisons possibles: AB, BA, CA, etc. Toutes les paires de séquences devront donc être présentées dans les deux ordres possibles (par exemple AB, BA). Après chaque paire, un jugement est émis sur l'élément d'une paire qui est préféré dans le contexte du scénario d'essai.

Le chronogramme pour la présentation du stimulus est décrit sur la Figure 3. Le temps imparti aux votes sera inférieur ou égal à 10 s, selon le mode de scrutin utilisé. La durée de la présentation pourra être réduite ou augmentée, selon le contenu du matériel expérimental; mais elle doit être d'environ 10 s.

Remplacée par une version plus récente



T1207480-95

Ai, Aj séquence A dans les $j^{\text{ème}}$ et $j^{\text{ème}}$ conditions d'essai, respectivement
Bk, Bl séquence B (selon le cas) dans les $k^{\text{ème}}$ et $l^{\text{ème}}$ conditions d'essai, respectivement

FIGURE 3/P.910

Présentation du stimulus dans la méthode PC

Lorsque des résolutions réduites sont utilisées (par exemple CIF, QCIF, SIF), il peut être utile d'afficher chaque paire de séquences simultanément sur un même moniteur. L'Annexe C donne des directives sur ce procédé de présentation.

Pour la méthode PC, le nombre de répétitions n'a généralement pas besoin d'être pris en considération car la méthode implique par elle-même une présentation répétitive des mêmes conditions, mais sur des paires différentes.

Une variante de la méthode PC fait appel à une échelle catégorielle pour mesurer plus en détail les différences entre paires de séquences. Voir les références [4] et [7].

6.4 Comparaison des méthodes d'essai

Lors du choix d'une méthode d'essai, un critère important est la différence fondamentale entre méthodes faisant appel à des références explicites (par exemple DCR) et méthodes ne faisant pas appel à des références explicites (par exemple ACR et PC). Cette deuxième classe de méthodes ne contrôle ni la transparence ni la fidélité.

Il convient d'utiliser la méthode DCR lorsque l'on contrôle la fidélité de transmission par rapport au signal de source. Ce facteur présente souvent de l'importance pour l'évaluation de systèmes de haute qualité. La méthode DCR est depuis longtemps spécifiée en tant que méthode de base dans [4], pour l'évaluation des images de télévision dont la qualité normale représente les niveaux très élevés de la vidéophonie et de la vidéoconférence. D'autres méthodes peuvent être utilisées pour évaluer les systèmes de haute qualité. Les observations spécifiques de l'échelle DCR (dégradation "imperceptible/perceptible") sont précieuses lorsque la détection d'une dégradation par l'observateur est un facteur important.

Lorsqu'il importe de vérifier la fidélité par rapport au signal de source, il convient donc d'utiliser la méthode DCR.

La méthode DCR sera également appliquée pour l'évaluation de systèmes de haute qualité, dans le contexte des communications multimédias, cela grâce à la discrimination entre dégradation imperceptible/perceptible sur l'échelle DCR ainsi que grâce à la comparaison avec la qualité de référence.

La méthode ACR est facile et d'application rapide. Sa présentation des stimuli est semblable à celle de l'usage courant des systèmes. La méthode ACR convient donc bien pour des essais de qualification.

Le principal mérite de la méthode PC est son haut pouvoir discriminatoire, qui est particulièrement précieux lorsque plusieurs objets d'essai sont de qualité presque égale.

Remplacée par une version plus récente

Lorsqu'il faut évaluer un grand nombre d'objets au cours du même essai, la procédure fondée sur la méthode PC tend à être longue. Dans ce cas, un essai ACR ou DCR peut d'abord être effectué avec un nombre limité d'observateurs, suivi d'un essai PC effectué seulement sur les objets qui ont reçu à peu près la même note d'évaluation.

6.5 Conditions de référence

Les résultats des évaluations de qualité dépendent souvent, non seulement de la qualité vidéo proprement dite, mais aussi d'autres facteurs tels que l'échelle de qualité totale des conditions d'essais, l'expérience et les attentes des évaluateurs, etc. Pour tenir compte de certains de ces effets, on peut ajouter et utiliser comme référence un certain nombre de conditions d'essai fictives. Ces conditions sont décrites dans la Recommandation P.930 [1] avec les procédures permettant de les obtenir. L'introduction du signal de source comme condition de référence dans un essai PC est spécialement recommandée lorsque les dégradations introduites par les objets d'essai sont petites.

Le niveau de qualité des conditions de référence devra couvrir au moins l'échelle de qualité des objets d'essai.

6.6 Modèles expérimentaux

Différents modèles expérimentaux, tels que le modèle totalement randomisé, le modèle à carré latin, le modèle à carré gréco-latin ou les modèles à carré de Youden, les modèles à réplication de bloc, etc. (Bibliographie, 5) peuvent être utilisés, leur sélection devant être déterminée par le but de l'expérience.

Le soin est laissé à l'expérimentateur de choisir un modèle méthodologique en fonction d'objectifs spécifiques en termes de coût et de précision. Le modèle peut aussi dépendre des conditions propres à chaque essai.

Il est recommandé d'inclure au moins deux (si possible trois ou quatre) répétitions (c'est-à-dire répétitions de conditions identiques) dans l'expérience. Il y a plusieurs raisons pour utiliser des répétitions, la plus importante étant que les données répliquées permettent de mesurer la "fiabilité intra-individuelle": à cette fin, on peut utiliser le même ordre de présentation dans des conditions identiques. Si un ordre de présentation différent est utilisé, la variation qui en résultera pour les données expérimentales sera composée de l'effet séquentiel et de la fiabilité intra-individuelle.

Les répétitions permettent de calculer la fiabilité intra-individuelle et, si nécessaire, d'écarter des résultats non fiables issus de certains sujets. Une estimation de l'écart type des lois aussi bien intra-individuelle qu'inter-individuelle est par ailleurs une condition préalable pour effectuer une analyse correcte de la variance et pour généraliser les résultats à une plus vaste population. De plus, les effets d'apprentissage dans le cadre d'un essai donné sont compensés dans une certaine mesure.

On obtiendra une autre amélioration du traitement des effets d'apprentissage en prévoyant une session d'apprentissage au cours de laquelle au moins cinq conditions seront présentées au début de chaque session d'essai. Il conviendra de choisir ces conditions de façon qu'elles soient représentatives des séquences qui seront présentées plus tard en session. Les présentations préliminaires ne sont pas à prendre en compte dans l'analyse statistique des résultats d'essai.

7 Procédures d'évaluation

7.1 Conditions de visualisation

Il y a lieu d'effectuer l'essai dans les conditions de visualisation suivante, conformément à [12]:
distance de visualisation de 4H à 8H (voir Note)

Remplacée par une version plus récente

luminance (valeur de crête)	de 70 cd/m ² à 200 cd/m ²
rapport de contraste de l'écran sans éclairage du fond	de 30 à 50
rapport luminance du fond sur luminance maximale de l'écran	~0,25
éclairage lumineux	environ 500 lux
chromaticité générale	blanc

NOTE - La lettre H correspond à la hauteur de l'image. Il y a lieu de définir la distance de visualisation en tenant compte non seulement des dimensions de l'écran mais aussi du type d'écran, du type d'application et du but de l'expérience.

Pour une hauteur d'écran donnée, il est probable que la distance de visualisation préférée par les sujets augmentera lorsque la qualité visuelle se dégradera. Concernant ce point, il convient de prédéterminer la distance de visualisation préférée pour les essais de qualification. Pour les essais d'optimisation, il faudra utiliser des distances plus critiques (c'est-à-dire 4H-6H).

Il est préférable d'utiliser toute la surface de l'écran pour afficher les séquences. Si toutefois les séquences doivent, pour une raison ou pour une autre, être affichées dans une fenêtre de l'écran, la couleur du fond d'écran devra être le gris à 50 % correspondant à la valeur $Y=U=V=128$ (les valeurs U et V n'étant pas signées).

Pour la comparaison des résultats d'essai, toutes les conditions de visualisation doivent être fixes et égales dans tous les laboratoires pour les mêmes types d'essais.

7.2 Système de traitement et de reproduction

Il existe deux méthodes pour obtenir des images d'essai à partir des enregistrements de source:

- par transmission ou relecture des enregistrements vidéo en temps réel au moyen des systèmes en essai, pendant que les sujets sont en train de regarder et de donner leurs impressions;
- par traitement en différé des enregistrements de source au moyen de l'appareil en essai et par enregistrement du signal de sortie afin d'obtenir un nouvel ensemble d'enregistrements.

Dans le deuxième cas, il y a lieu d'utiliser un magnétoscope afin de minimiser les dégradations pouvant être produites par le processus d'enregistrement. De toute façon, compte tenu du fait que les dégradations introduites par les procédés de codage à faible débit sont habituellement plus évidentes que les dégradations introduites par la modulation, on peut faire appel à des magnétoscopes de qualité professionnelle, tels que D2, MII et BetacamSP.

On peut utiliser un moniteur cathodique ou à cristaux liquides. Les dimensions et le type du moniteur utilisé devront convenir à l'application étudiée.

Les moniteurs devront être alignés conformément aux procédures définies en [8].

7.3 Visualisateurs

Le nombre de sujets pouvant participer à un essai de visualisation (ainsi qu'à des essais d'utilisabilité portant sur des terminaux ou sur des services) est compris entre 4 et 40. Quatre est le minimum absolu pour des raisons d'ordre statistique, tandis que quarante est un nombre qu'il y a rarement intérêt à dépasser.

Le nombre réel de participants à un essai spécifique dépendra en fait de la validité requise et de la nécessité de généraliser à une grande population à partir d'un échantillon.

Remplacée par une version plus récente

En général, au moins 15 observateurs doivent participer à l'expérience. Ils ne doivent pas, sur le plan professionnel, être directement impliqués dans l'évaluation de la qualité des images ni être des évaluateurs expérimentés.

Dans les phases initiales de la mise au point de systèmes de communications vidéo et lors d'expériences pilotes effectuées avant un vaste essai, de petits groupes d'experts (4 à 8) ou d'autres sujets spécialisés peuvent toutefois fournir des résultats indicatifs.

Avant une session, les observateurs doivent habituellement subir un examen d'acuité visuelle normale ou d'acuité visuelle corrigée à la normale, ainsi que de vision chromatique normale. Concernant l'acuité, aucune erreur ne doit être faite sur la ligne 20/30° d'un optotype normal (Bibliographie, 3). Celui-ci doit être échelonné en fonction de la distance de visualisation et l'essai d'acuité visuelle doit être effectué à l'endroit exact où les images vidéo seront visualisées par les sujets assis (c'est-à-dire que l'optotype sera placé en appui sur le moniteur). Concernant la couleur, il ne faut pas que 2 plans sur 12 (Bibliographie, 4) soient manqués.

7.4 Instructions données aux visualisateurs et session d'apprentissage

Avant de commencer l'expérience, il y a lieu de donner aux sujets le scénario de l'application prévue du système en essai. Ils recevront également, sous forme écrite, une description du type d'évaluation, l'échelle d'appréciation subjective et la présentation des stimuli. L'étendue et le type des dégradations seront présentés lors d'essais préliminaires, qui pourront contenir des séquences vidéo autres que celles des essais proprement dits.

Il ne faut pas déduire que la plus mauvaise qualité vue lors de la série d'apprentissage correspond au plus bas degré d'appréciation subjective de l'échelle.

On répondra avec soin - et seulement avant le début de la session - aux questions posées au sujet de la procédure ou de la signification des instructions, afin d'éviter toute distorsion.

L'Appendice II propose un texte pouvant être donné aux évaluateurs.

8 Analyse statistique et compte rendu des résultats

Les résultats devront être consignés avec les détails sur le montage expérimental. Pour chaque combinaison de variables d'essai, on donnera la valeur moyenne et l'écart type de la répartition statistique des notes d'évaluation.

On calculera la fiabilité des sujets à partir des données et la méthode utilisée à cette fin sera indiquée. Certains critères de fiabilité subjective sont donnés dans [4] et [5].

Il est intéressant d'analyser la répartition cumulée des notes. Comme ce type de loi est indépendant de la linéarité, ces courbes peuvent être particulièrement utiles pour des données dont la linéarité est sujette à caution, comme celles que l'on obtient par les méthodes ACR et DCR, en association avec des échelles catégorielles sans gradation (c'est-à-dire à jugement catégoriel).

Les données peuvent être organisées comme dans l'exemple donné par le Tableau 1 pour la méthode ACR.

Remplacée par une version plus récente

TABLEAU 1/P.910

Tableau informatif présentant la répartition cumulée de notes obtenues par la méthode ACR

Condition	Total des votes	Excellent	Bon	Satisfaisant	Médiocre	Mauvais	MOS	CI	Std	% GOB	% POW

Condition: critère indiquant une combinaison de variables expérimentales.

Total des votes: nombre de votes recueillis pour cette condition.

Excellent,, Satisfaisant,, Mauvais: nombre de votes dans ces catégories.

Les techniques classiques d'analyse de la variance devront être utilisées pour évaluer la portée des paramètres d'essai. Si l'évaluation vise à évaluer la qualité vidéo en fonction d'un paramètre, des techniques de lissage de courbe peuvent être utiles pour l'interprétation des données.

Dans le cas de comparaisons par paires, on trouvera dans le Manuel de téléphonométrie, paragraphe 2.6.2C [6], la méthode permettant de calculer la position de chaque stimulus sur une échelle d'intervalles. Dans cette méthode, la différence entre les stimuli correspond à la différence entre les préférences concernant ces stimuli.

Bibliographie

- [1] GONZALEZ (R.C.), WINTZ (P.): Digital Image Processing, 2nd Edition, Addison-Wesley Publishing Co., Reading, Massachusetts, 1987.
- [2] RACE Industrial Consortium Project 1018 HIVITS, WP B5, Picture quality measurement, 1988.
- [3] Grahm-Field Catalogue Number 13-1240.
- [4] Pseudo Isochromatic Plates, engraved and printed by, The Beck Engraving Co., Inc., Philadelphia and New York, USA.
- [5] KIRK (R.E.): Experimental Design – Procedures for the Behavioural Sciences, 2nd Edition, Brooks/Cole Publishing Co., California, 1982.
- [6] VIRTANEN (M.T.), GLEISS (N.), GOLDSTEIN (M.): On the use of Evaluative Category Scales in Telecommunications, HFT 1995, Human Factors in Telecommunication Conference, Melbourne, 1995.

Annexe A

Détails relatifs à la caractérisation des séquences d'essai

(Cette annexe fait partie intégrante de la présente Recommandation)

A.1 Filtre de Sobel

On met en oeuvre le filtre de Sobel en convoluant deux noyaux de 3x3 éléments dans chaque trame vidéo puis en prenant la somme quadratique des carrés des résultats de ces convolutions.

Pour $y = \text{Sobel}(x)$, on suppose que $x(i,j)$ représente le pixel de l'image d'entrée à la $i^{\text{ème}}$ ligne et à la $j^{\text{ème}}$ colonne. Le terme $Gv(i,j)$ sera le résultat de la première convolution et sera calculé comme suit:

Remplacée par une version plus récente

$$Gv(i,j) = -1*x(i-1,j-1) - 2*x(i-1,j) - 1*x(i-1,j+1) + \\ + 0*x(i,j-1) + 0*x(i,j) + 0*x(i,j+1) + \\ + 1*x(i+1,j-1) + 2*x(i+1,j) + 1*x(i+1,j+1)$$

De même, le terme Gh(i,j) sera le résultat de la deuxième convolution et sera calculé comme suit:

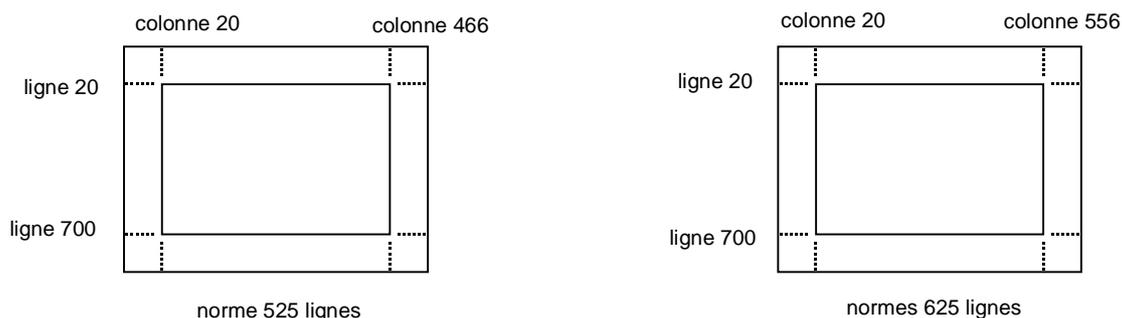
$$Gh(i,j) = -1*x(i-1,j-1) + 0*x(i-1,j) + 1*x(i-1,j+1) + \\ - 2*x(i,j-1) + 0*x(i,j) + 2*x(i,j+1) + \\ - 1*x(i+1,j-1) + 0*x(i+1,j) + 1*x(i+1,j+1)$$

A la sortie du filtre de Sobel, l'image sera donc, à la $i^{\text{ième}}$ ligne et à la $j^{\text{ième}}$ colonne, donnée par la relation suivante:

$$y(i,j) = \{ [Gv(i,j)]^2 + [Gh(i,j)]^2 \}^{0,5}$$

Les calculs sont effectués pour tous les $2 \leq i \leq N-1$ et tous les $2 \leq j \leq M-1$, où N est le nombre de lignes et M le nombre de colonnes.

Il est recommandé que les calculs soient effectués sur une sous-image de la trame vidéo afin d'éviter des effets de bord indésirables et parce que les bords extrêmes d'une trame vidéo sont habituellement invisibles aux utilisateurs d'écrans cathodiques. Pour cela, on pourra utiliser une sous-image appropriée, telle qu'illustrée à titre d'exemple dans la Figure A.1 pour les formats 625 et 525 lignes de l'UIT-R BT.601-4 [3].



T1207490-95

FIGURE A.1/P.910

Sous-images à utiliser pour calculer SI et TI pour formats UIT-R BT.601-4 [3] 525 et 625 lignes

On pourra trouver de plus amples renseignements sur le filtre de Sobel en (Bibliographie, 1).

A.2 Mode d'utilisation des informations SI et TI pour la sélection des séquences d'essai

Lors de la sélection des séquences d'essai, il peut être utile de comparer les informations spatiales relatives et les informations temporelles relatives se trouvant dans les diverses séquences. Généralement, la difficulté due à la compression est en relation directe avec les informations spatio-temporelles d'une séquence.

Si un petit nombre de séquences doivent être utilisées au cours d'un même essai, il est parfois important de choisir des séquences couvrant une vaste portion du plan des informations spatio-temporelles (voir la Figure A.2). Si quatre séquences d'essai doivent être utilisées dans un

Remplacée par une version plus récente

essai, on peut choisir une séquence dans chacun des quatre quadrants du plan des informations spatio-temporelles.

En variante, si l'on essaye de choisir des séquences d'essai qui sont équivalentes en difficulté de codage, il serait souhaitable de sélectionner des séquences possédant des valeurs similaires de SI et de TI.

A.3 Exemples

La Figure A.2 montre les proportions d'informations spatiales et temporelles pour quelques scènes représentatives. Elle indique également à quel endroit d'un plan ces informations spatio-temporelles peuvent être placées.

Sur l'axe $TI=0$ (en bas du tracé), on trouve les scènes non animées et celles dont le contenu cinétique est très réduit (comme les points l, f et a). Près du bord supérieur du tracé, on trouve les scènes à contenu cinétique important (comme les points p, q et i). Sur l'axe $SI=0$ (sur le bord gauche du tracé), on trouve les scènes contenant le moins de détails spatiaux (comme les points l, k, x, u et f). Près du bord droit du tracé, on trouve les scènes contenant le plus de détails spatiaux (comme les points h et s). Les valeurs de SI et de TI ont été obtenues au moyen des équations ci-dessus et au moyen d'images vidéo qui ont été échantillonnées spatialement conformément à la Recommandation UIT-R BT.601-4 [3]. Le Tableau A.1 énumère, selon leurs catégories de contenu, les scènes d'essai de cet exemple.

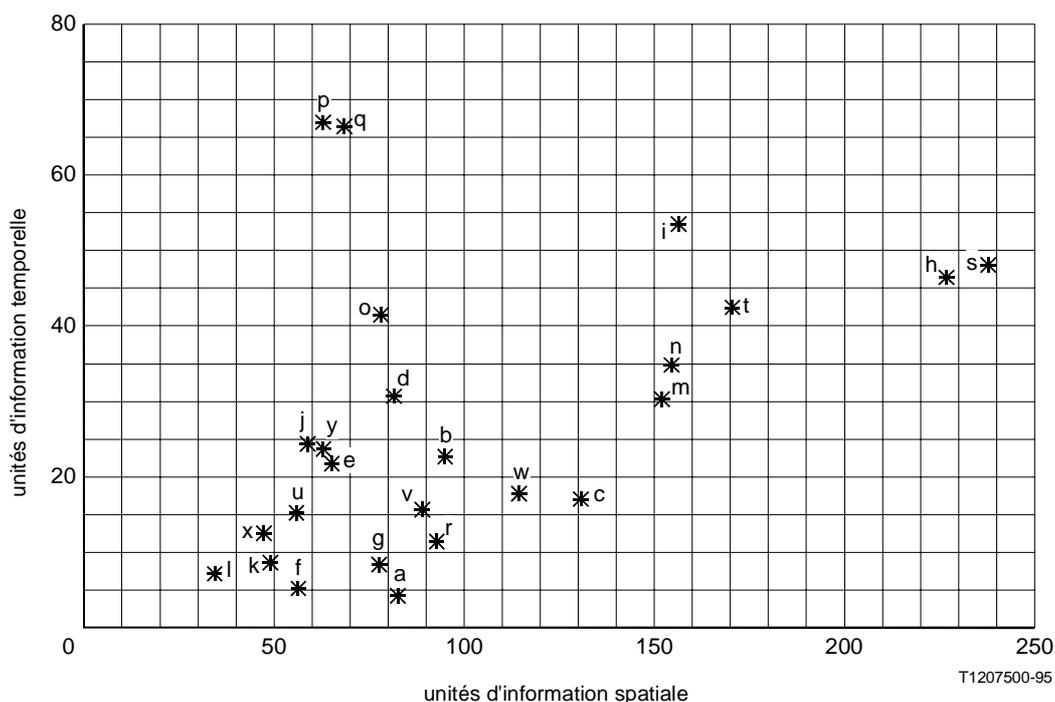


FIGURE A.2/P.910

Exemple de tracé spatio-temporel d'un ensemble de scènes d'essai

Remplacée par une version plus récente

TABLEAU A.1/P.910

Catégories de contenu des scènes d'essai

Catégorie	Description	Nom et lettre de la scène
A	une seule personne, tête et épaules surtout, détails et mouvements limités	vtc1nw(f), susie(j), disguy(k), disgal(l)
B	une seule personne avec données graphiques et/ou plus de détails	vtc2mp(a), vtc2zm(b), boblec(e), smity1(m), smity2(n), vowels(w), inspec(x)
C	plus d'une seule personne	3inrow(d), 5row1(g), intros(o), 3twos(p), 2wbord(q), split6(r)
D	graphiques avec pointage	washdc(c), cirkit(s), rodmap(t), filter(u), ysmite(v)
E	mouvements rapides d'objet et/ou de caméra (exemples de TV radiodiffusée)	flogar(h), ftball(i), fedas(y)

Annexe B

Echelles d'évaluation supplémentaires

(Cette annexe fait partie intégrante de la présente Recommandation)

B.1 Echelles d'évaluation

Surtout pour l'évaluation de codecs vidéo à faible débit, il est souvent nécessaire de faire appel à des échelles d'évaluation comportant plus de cinq degrés. Une échelle qui convient à cette fin est celle des neuf degrés, où l'on utilise, pour qualifier un degré de l'échelle sur deux, les cinq catégories qualitatives définies verbalement comme recommandé au 6.1 et décrit dans la Figure B.1.

9	Excellent
8	
7	Bon
6	
5	Satisfaisant
4	
3	Médiocre
2	
1	Mauvais

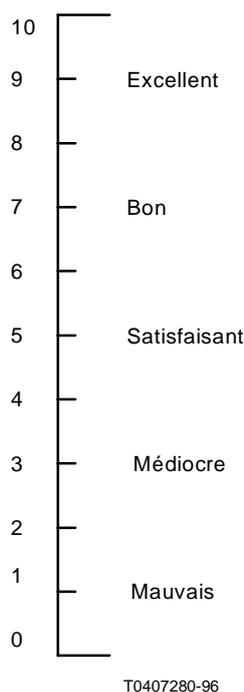
FIGURE B.1/P.910

Echelle qualitative à 9 degrés numériques

La Figure B.2 montre une autre extension de cette échelle, où les points extrêmes ont été définis verbalement comme étant des points de référence qui ne sont pas utilisés pour l'évaluation. Dans cette définition verbale, on fait appel à une sorte de référence (par exemple, dans la Figure B.2, c'est l'original qui est utilisé comme référence). Celle-ci peut être soit explicite soit implicite. Elle sera

Remplacée par une version plus récente

clairement décrite au cours de la phase d'apprentissage. Voir également [5] et [6], section 2.6, échelle a).



Le nombre 10 correspond à une qualité de reproduction qui est parfaitement fidèle à l'original. Aucune autre amélioration n'est possible.

Le nombre 0 correspond à une qualité de reproduction qui n'a aucune ressemblance avec l'original.

FIGURE B.2/P.910

Echelle qualitative à 11 degrés numériques

Pour ces deux types d'échelle, la réponse issue des sujets peut être enregistrée soit sous forme de nombres écrits sur une feuille de réponse, soit sous forme de marques portées sur l'échelle proprement dite (auquel cas il faut donner une échelle distincte sur la feuille de réponse pour chaque condition d'évaluation). Lorsque des réponses numériques sont requises, les sujets doivent être incités à utiliser des décimales (par exemple 2,2 au lieu de 2) tout en conservant la possibilité de n'utiliser que des entiers.

Il convient de noter qu'il est parfois difficile de traduire les noms des catégories d'échelle dans certaines langues. Ce faisant, la relation inter-catégorielle peut devenir différente de celle du langage d'origine (Bibliographie, 6).

Une autre possibilité consiste à utiliser des échelles continues.

Comme des données continues sont habituellement arrondies à un certain degré de précision, on peut utiliser une échelle de cotation analogue à celle qui est représentée sur la Figure B.3, afin de simplifier le recueil des données. Les étiquettes qualificatives ne sont utilisées qu'aux extrémités de l'échelle et une marque est inscrite au milieu de celle-ci. Ce procédé devrait réduire la distorsion due à l'interprétation des étiquettes. Chaque case peut correspondre à une valeur numérique précise et les données peuvent être recueillies sans ambiguïté.

Remplacée par une version plus récente

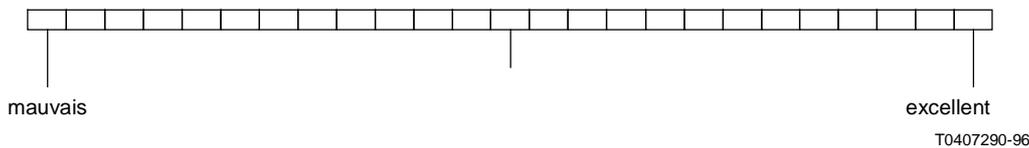


FIGURE B.3/P.910

Echelle quasi-continue pour l'évaluation de la qualité

B.2 Dimensions d'évaluation additionnelles

Si les systèmes évalués au cours d'un essai sont jugés comme ayant une qualité globale à peu près égale et obtiennent donc des notes très similaires, on peut avoir intérêt à évaluer les composantes qualitatives additionnelles sur des échelles distinctes pour chaque condition. Il est ainsi possible de recevoir des renseignements sur des caractéristiques spécifiques où des objets d'essai sont perçus comme étant nettement différents, même si la qualité globale est en fait presque identique. Les résultats issus de tels essais additionnels peuvent donner de précieuses informations de diagnostic au sujet des systèmes en essai.

Des exemples de dimensions d'évaluation sont énumérés ci-dessous. On peut admettre qu'ils définissent des facteurs contribuant à la qualité globale de l'image perçue. Ils indiquent également si un facteur apporte une contribution positive ou négative en termes de qualité:

brillance (positive)

contraste (positive)

rendu des couleurs (positive)

définition des contours (positive)

stabilité du fond (positive)

rapidité de réassemblage des images (positive)

sautillement (négative)

effets de maculage (négative)

effets de papillotement (négative)

images doubles/ombres (négative)

halo (négative)

La recherche a récemment montré que l'on pouvait combiner ces facteurs en leur attribuant des pondérations appropriées puis en les additionnant afin d'obtenir une prédiction de qualité globale. Voir plus loin (Bibliographie, 2).

Pour évaluer séparément les dimensions de la qualité vidéo globale, on peut faire appel à un questionnaire spécial. Le questionnaire ci-dessous donne des exemples de questions pouvant être posées après la présentation de chaque condition d'essai.

Remplacée par une version plus récente

Annexe C

Présentation simultanée de paires de séquences

(Cette annexe fait partie intégrante de la présente Recommandation)

C.1 Introduction

Lorsque les systèmes évalués lors d'un essai font appel à un format d'image réduit, comme les formats CIF, QCIF, SIF, etc., et que l'on utilise la méthode DCR ou PC, il peut être avantageux d'afficher simultanément les deux séquences de chaque paire sur le même moniteur.

Les avantages de l'utilisation de la présentation simultanée (SP) sont les suivants:

- 1) la présentation simultanée réduit considérablement la durée de l'essai;
- 2) si des dimensions d'image appropriées sont utilisées, il est plus facile pour les sujets d'évaluer les différences entre les stimuli;
- 3) comme le nombre de présentations est réduit de moitié dans les mêmes conditions d'essai, l'attention des sujets est habituellement meilleure lorsque la présentation simultanée est utilisée.

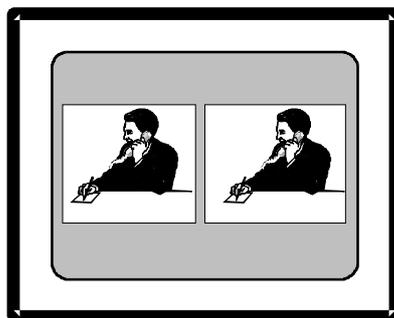
La méthode de la présentation simultanée exige des précautions particulières afin que les sujets évitent tout biais dû au type de présentation.

C.2 Synchronisation

Les deux séquences doivent être parfaitement synchronisées, c'est-à-dire qu'elles doivent toutes les deux commencer et finir à la même trame et que leur affichage doit être synchronisé. Cela n'empêche pas de comparer des séquences codées à des débits différents, à condition d'appliquer un suréchantillonnage temporel approprié.

C.3 Conditions de visualisation

Les séquences doivent être affichées dans deux fenêtres placées côte à côte sur un fond gris à 50% (spécifié au 5.1), comme indiqué sur la Figure C.1. Pour réduire le mouvement des yeux lors du passage du regard d'une fenêtre à l'autre, la distance de visualisation doit être de huit fois la hauteur de l'image (8H). La diagonale des moniteurs doit être d'au moins 14".



T1207510-95

FIGURE C.1/P.910

Position relative des deux séquences en présentation simultanée

Remplacée par une version plus récente

C.4 Présentations

Dans la méthode DCR, la référence doit toujours être placée du même côté (par exemple à gauche) et les sujets doivent toujours être informés des positions relatives de la référence ainsi que des conditions d'essai.

Dans la méthode PC, toutes les paires de séquences doivent être affichées dans les deux ordres possibles (par exemple AB, BA); c'est-à-dire que la séquence qui a été affichée à gauche sera ensuite affichée à droite et ainsi de suite.

Appendice I

Séquences d'essai

(Cet appendice ne fait pas partie intégrante de la présente Recommandation)

La sélection des séquences d'essai appropriées est un élément clé pour la planification des évaluations subjectives. Lorsqu'il faut corrélérer les résultats d'essai effectués avec différents groupes d'observateurs ou dans différents laboratoires, il importe de disposer d'une série commune de séquences.

Le Tableau I.1 décrit une première série de telles séquences, assorties chacune des informations suivantes:

- la catégorie (définie dans le Tableau A.1);
- une brève description de la scène;
- le format de la source (625 ou 525 lignes, format UIT-R BT.601-4 ou Betacam SP);
- les valeurs des informations spatiales et temporelles (définies respectivement aux 5.3.1 et 5.3.2).

Toutes les séquences énumérées dans le Tableau I.1 sont dans le domaine public et peuvent être librement utilisées pour les évaluations et les démonstrations. Certaines des séquences suggérées font partie de la vidéothèque du CCIR, décrite dans le Rapport 1213 du CCIR [10].

D'autres séquences de la vidéothèque de l'UIT-R pourront être utilisées pour des applications particulières, comme celles qui sont fondées sur le stockage et la consultation d'images vidéo.

La série de séquences d'essai est encore à l'étude. La série énumérée dans le Tableau I.1 peut être améliorée ou étendue d'au moins deux façons:

- 1) les séquences représentatives d'une assez large gamme d'applications doivent être incluses (par exemple la vidéophonie mobile, le téléenseignement, etc.);
- 2) le format source de chaque séquence doit être celui de l'UIT-R BT.601-4 [3], en version 525 comme en 625 lignes.

Remplacée par une version plus récente

TABLEAU I.1/P.910

Séquences d'essai pour évaluation de la qualité vidéo dans des applications multimédias

Séquence	Catégorie	Description	Format source	SI	TI
washdc	D	carte de Washington D.C. avec mouvement de la main et du crayon	Betacam SP (525 lignes)	130,5	17,0
3inrow	C	hommes à table, prise panoramique	Betacam SP (525 lignes)	81,7	30,8
vtc1nw	A	femme assise lisant une nouvelle	Betacam SP (525 lignes)	56,2	5,3
susie	A	jeune femme au téléphone	UIT-R BT.601-4 525/625 lignes	58,7	24,6
flower garden	E	paysage, prise panoramique	UIT-R BT.601-4 525/625 lignes	227,0	46,4
smity2	B	vendeur à son comptoir avec une revue	Betacam SP (525 lignes)	154,5	35,1

Appendice II

Instructions pour les essais de visualisation

(Cet appendice ne fait pas partie intégrante de la présente Recommandation)

Les instructions suivantes peuvent être données comme base aux évaluateurs participant à des expériences par la méthode ACR, DCR ou PC.

Les instructions doivent par ailleurs donner des renseignements sur la durée d'essai approximative, sur les pauses, sur les tests préliminaires et tous autres détails utiles aux évaluateurs. Ces renseignements ne sont pas inclus ici parce qu'ils dépendent de chaque mise en oeuvre spécifique.

II.1 Méthode ACR

Bonjour et merci de votre participation.

Dans cette expérience, vous verrez de brèves séquences vidéo sur l'écran qui est devant vous. Chaque fois que vous verrez une séquence, vous devrez en juger la qualité en cochant l'un des 5 niveaux de l'échelle suivante:

- 5 Excellent
- 4 Bon
- 3 Satisfaisant
- 2 Médiocre
- 1 Mauvais

Observez avec attention toute la séquence vidéo avant d'exprimer votre jugement.

Remplacée par une version plus récente

II.2 Méthode DCR

Bonjour et merci de votre participation.

Dans cette expérience, vous verrez de brèves séquences vidéo sur l'écran qui est devant vous. Chaque séquence sera présentée deux fois en succession rapide: dans chaque paire, seule la seconde séquence a subi un traitement de signal. A la fin de la présentation de chaque paire, vous devrez évaluer la dégradation de la seconde séquence par rapport à la première. Vous exprimerez votre jugement au moyen de l'échelle suivante:

- 5 Imperceptible
- 4 Perceptible mais pas gênant
- 3 Légèrement gênant
- 2 Gênant
- 1 Très gênant

Observez avec attention toute la paire de séquences vidéo avant d'exprimer votre jugement.

II.3 Méthode PC

Bonjour et merci de votre participation.

Dans cette expérience, vous verrez de brèves séquences vidéo sur l'écran qui est devant vous. Chaque séquence sera présentée deux fois en succession rapide, chaque fois avec un codec différent. L'ordre des séquences et la combinaison de codecs varient de manière aléatoire selon les paires. A la fin de chaque présentation de paire, vous devrez exprimer votre préférence en cochant une des cases indiquées ci-dessous. Vous cocherez la case 1 si vous préférez la première séquence ou la case 2 si vous préférez la deuxième séquence de la paire

1	2
---	---

Observez avec attention toute la paire de séquences vidéo avant d'exprimer votre jugement.

Remplacée par une version plus récente

SÉRIES DES RECOMMANDATIONS UIT-T

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