

Union internationale des télécommunications

UIT-T

SECTEUR DE LA NORMALISATION
DES TÉLÉCOMMUNICATIONS
DE L'UIT

P.910

(04/2008)

SÉRIE P: QUALITÉ DE TRANSMISSION
TÉLÉPHONIQUE, INSTALLATIONS TÉLÉPHONIQUES
ET RÉSEAUX LOCAUX

Qualité audiovisuelle dans les services multimédias

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

Recommandation UIT-T P.910

UIT-T



RECOMMANDATIONS UIT-T DE LA SÉRIE P
**QUALITÉ DE TRANSMISSION TÉLÉPHONIQUE, INSTALLATIONS TÉLÉPHONIQUES
 ET RÉSEAUX LOCAUX**

Vocabulaire et effets des paramètres de transmission sur l'opinion des usagers	Série	P.10
Caractéristiques des terminaux vocaux	Série	P.30
		P.300
Systèmes de référence	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 vocale	Série	P.80
Méthodes d'évaluation objective et subjective de la qualité vocale et de vidéo	Série	P.800
Qualité audiovisuelle dans les services multimédias	Série	P.900
Aspects relatifs à la qualité de transmission et à la qualité de service aux points de terminaison des réseaux à protocole Internet	Série	P.1000
Communications impliquant des véhicules	Série	P.1100
Modèles et outils d'évaluation de la qualité de flux médias transmis en continu	Série	P.1200
Évaluation des téléconférences	Série	P.1300
Lignes directrices pour l'analyse et l'évaluation statistiques des mesures de qualité et le compte rendu de ces mesures	Série	P.1400
Méthodes d'évaluation objective et subjective de la qualité des services autres que les services vocaux ou de vidéo	Série	P.1500

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

Recommandation UIT-T P.910

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

Résumé

La Recommandation UIT-T P.910 décrit les méthodes d'évaluation subjective, non interactive, utilisées pour évaluer la qualité vidéo globale unidirectionnelle pour des applications multimédia, telles que visioconférence, stockage et extraction, télémédecine, etc. Ces méthodes peuvent être utilisées à plusieurs fins différentes y compris la sélection des algorithmes, la hiérarchisation des systèmes vidéo selon leurs performances et l'évaluation du niveau de qualité au cours d'une connexion audiovisuelle. Cette Recommandation définit en outre les caractéristiques des séquences source à utiliser, comme par exemple la durée, le type de contenu, le nombre de séquences, etc.

Historique

Edition	Recommandation	Approbation	Commission d'études	ID unique*
1.0	ITU-T P.910	1996-08-30	12	11.1002/1000/3641
2.0	ITU-T P.910	1999-09-30	12	11.1002/1000/4751
3.0	ITU-T P.910	2008-04-06	9	11.1002/1000/9317

* Pour accéder à la Recommandation, reporter cet URL <http://handle.itu.int/> dans votre navigateur Web, suivi de l'identifiant unique, par exemple <http://handle.itu.int/11.1002/1000/11830-en>.

AVANT-PROPOS

L'Union internationale des télécommunications (UIT) est une institution spécialisée des Nations Unies dans le domaine des télécommunications et des technologies de l'information et de la communication (ICT). Le Secteur de la normalisation des télécommunications (UIT-T) est un organe permanent de l'UIT. Il est chargé de l'étude des questions techniques, d'exploitation et de tarification, et émet à ce sujet des Recommandations en vue de la normalisation des télécommunications à l'échelle mondiale.

L'Assemblée mondiale de normalisation des télécommunications (AMNT), qui se réunit tous les quatre ans, détermine les thèmes d'étude à traiter par les Commissions d'études de l'UIT-T, lesquelles élaborent en retour des Recommandations sur ces thèmes.

L'approbation des Recommandations par les Membres de l'UIT-T s'effectue selon la procédure définie dans la Résolution 1 de l'AMNT.

Dans certains secteurs des technologies 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.

Le respect de cette Recommandation se fait à titre volontaire. Cependant, il se peut que la Recommandation contienne certaines dispositions obligatoires (pour assurer, par exemple, l'interopérabilité et l'applicabilité) et considère que la Recommandation est respectée lorsque toutes ces dispositions sont observées. Le futur d'obligation et les autres moyens d'expression de l'obligation comme le verbe "devoir" ainsi que leurs formes négatives servent à énoncer des prescriptions. L'utilisation de ces formes ne signifie pas qu'il est obligatoire de respecter la Recommandation.

DROITS DE PROPRIÉTÉ INTELLECTUELLE

L'UIT attire l'attention sur la possibilité que l'application ou la mise en œuvre de la présente Recommandation puisse donner lieu à l'utilisation d'un droit de propriété intellectuelle. L'UIT ne prend pas position en ce qui concerne l'existence, la validité ou l'applicabilité des droits de propriété intellectuelle, qu'ils soient revendiqués par un membre de l'UIT ou par une tierce partie étrangère à la procédure d'élaboration des Recommandations.

A la date d'approbation de la présente Recommandation, l'UIT n'avait pas été avisée de l'existence d'une propriété intellectuelle protégée par des brevets à acquérir pour mettre en œuvre la présente Recommandation. Toutefois, comme il ne s'agit peut-être pas de renseignements les plus récents, il est vivement recommandé aux développeurs de consulter la base de données des brevets du TSB sous <http://www.itu.int/ITU-T/ipr/>.

© UIT 2017

Tous droits réservés. Aucune partie de cette publication ne peut être reproduite, par quelque procédé que ce soit, sans l'accord écrit préalable de l'UIT.

TABLE DES MATIÈRES

		Page
1	Domaine d'application	1
2	Références.....	1
3	Termes et définitions	1
4	Abréviations et acronymes	3
5	Signal de source.....	3
	5.1 Environnement d'enregistrement	4
	5.2 Système d'enregistrement	4
	5.3 Caractéristiques des scènes.....	5
6	Méthodes d'essai et modèles expérimentaux	6
	6.1 Evaluation par catégories absolues (ACR).....	6
	6.2 Evaluation par catégories absolues avec référence cachée (ACR-HR).....	7
	6.3 Evaluation par catégories de dégradation (DCR).....	8
	6.4 Méthode de comparaison par paires (PC)	9
	6.5 Comparaison des méthodes d'essai.....	9
	6.6 Conditions de référence	10
	6.7 Modèles expérimentaux.....	10
7	Procédures d'évaluation	11
	7.1 Conditions de visualisation.....	11
	7.2 Système de traitement et de reproduction	12
	7.3 Visualisateurs	12
	7.4 Instructions données aux visualisateurs et session d'apprentissage.....	13
8	Analyse statistique et compte rendu des résultats.....	13
Annexe A – Détails relatifs à la caractérisation des séquences d'essai.....		15
	A.1 Filtre de Sobel.....	15
	A.2 Mode d'utilisation des informations SI et TI pour la sélection des séquences d'essai	16
	A.3 Exemples	16
Annexe B – Echelles d'évaluation supplémentaires.....		18
	B.1 Echelles d'évaluation	18
	B.2 Dimensions d'évaluation additionnelles	19
Annexe C – Présentation simultanée de paires de séquences		21
	C.1 Introduction	21
	C.2 Synchronisation	21
	C.3 Conditions de visualisation.....	21
	C.4 Présentations.....	21
Annexe D – Classes vidéo et audio et leurs attributs		22
Appendice I – Séquences d'essai.....		24

	Page
Appendice II – Instructions pour les essais de visualisation.....	25
II.1 Méthodes ACR et ACR-HR	25
II.2 Méthode DCR.....	25
II.3 Méthode PC	25
Appendice III – Le double stimulus pour une évaluation continue	27
III.1 Procédure d'essai	27
III.2 Phase préparatoire	27
III.3 Caractéristiques du protocole d'essai.....	27
III.4 Traitement des données	28
III.5 Fidélité des observateurs	31
Appendice IV – L'évaluation axée sur l'objet	33
Appendice V – Une échelle d'évaluation supplémentaire pour DCR	35
Bibliographie.....	36

Recommandation UIT-T P.910

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

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 aux débits spécifiés pour les classes TV3, MM4, MM5 et MM6 dans le Tableau D.2 pour des applications telles que la visiophonie, la visioconférence et les applications de stockage/extraction. Les méthodes pourront 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

La présente Recommandation se réfère à certaines dispositions des Recommandations UIT-T et textes suivants qui, de ce fait, en sont partie intégrante. Les versions indiquées étaient en vigueur au moment de la publication de la présente Recommandation. Toute Recommandation ou tout texte étant sujet à révision, les utilisateurs de la présente Recommandation sont invités à se reporter, si possible, aux versions les plus récentes des références normatives suivantes. La liste des Recommandations de l'UIT-T en vigueur est régulièrement publiée. La référence à un document figurant dans la présente Recommandation ne donne pas à ce document, en tant que tel, le statut d'une Recommandation.

- [UIT-T J.61] Recommandation UIT-T J.61 (1988), *Qualité de transmission des circuits de télévision destinés à être utilisés dans les communications internationales*.
- [UIT-T P.800] Recommandation UIT-T P.800 (1996), *Méthodes d'évaluation subjective de la qualité de transmission*.
- [UIT-T P.930] Recommandation UIT-T P.930 (1996), *Principes d'un système de dégradation de référence pour signaux vidéo*.
- [UIT-R BT.500-9] Recommandation UIT-R BT.500-9 (1998), *Méthodologie d'évaluation subjective de la qualité des images de télévision*.
- [UIT-R BT.601-4] Recommandation UIT-R BT.601-4 (1994), *Paramètres de codage de télévision numérique pour studios*.
- [UIT-R BT.814-1] Recommandation UIT-R BT.814-1 (1994), *Spécifications et méthodes de réglage de la brillance et du contraste des dispositifs de visualisation*.
- [CEI/TR 60268-13] CEI/TR 60268-13 (1998), *Equipements pour systèmes électroacoustiques – Partie 13: Essais d'écoute des haut-parleurs*
<<http://webstore.iec.ch/webstore/webstore.nsf/artnum/022890>>.

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 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. Voir le paragraphe 5.3.1 pour l'équation relative à cette information.

3.5 information de perception temporelle (TI, *temporal perceptual information*): 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. Voir le paragraphe 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.

3.8 fidélité 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 et acronymes

La présente Recommandation utilise les abréviations et acronymes suivantes:

ACR	évaluation par catégories absolues (<i>absolute category rating</i>)
ACR-HR	évaluation par catégories absolues avec référence cachée (<i>absolute category rating with hidden reference</i>)
CCD	dispositif à couplage de charge (<i>charge coupled device</i>)
CI	intervalle de confiance (<i>confidence interval</i>)
CIF	format intermédiaire commun (<i>common intermediate format</i>)
NOTE –	Un format d'image défini dans la Recommandation [b-UIT-T H.261] pour la vidéophonie: 352 lignes × 288 pixels.
CRT	tube à rayons cathodiques (<i>cathode ray tube</i>)
DCR	évaluation par catégories de dégradation (<i>degradation category rating</i>)
DV	note différentielle de visualisateur (<i>differential viewer score</i>)
%GOB	pourcentage de bons ou de meilleurs (proportions de bons et d'excellents) (<i>percent of good or better</i>)
LCD	affichage par cristaux liquides (<i>liquid crystal display</i>)
MOS	note moyenne d'opinion (<i>mean opinion score</i>)
PC	comparaison par paires (<i>pair comparison</i>)
%POW	pourcentage de médiocres ou de pires (proportion de notes médiocres et mauvaises) (<i>percent of poor or worse</i>)
PVS	séquence vidéo traitée (<i>processed video sequence</i>)
QCIF	quart de CIF
NOTE –	Un format d'image défini dans la Recommandation [b-UIT-T H.261] pour la vidéophonie: 176 lignes × 144 pixels.
S/N	rapport signal sur bruit (<i>signal-to-noise ratio</i>)
SI	information spatiale (<i>spatial information</i>)
SIF	format intermédiaire normalisé (<i>standard intermediate format</i>)
NOTE –	Un format d'image défini dans la norme [b-ISO/CEI 11172] (MPEG-1): 352 lignes × 288 pixels × 25 trames/s et 352 lignes × 240 pixels × 30 trames/s.
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 sources 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 YUV (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, *automatic gain control*).

On mesurera le rapport S/N pondéré conformément au paragraphe 3.2.1 de la Partie C de la Recommandation [UIT-T J.61]. La valeur de ce rapport sera supérieure à 45 dB (eff.).

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 la Recommandation [UIT-R BT.601-4]. Afin d'éviter la distorsion du signal de source, celui-ci sera stocké sous forme numérique, par exemple sur ordinateur ou sur magnétoscope D1 en format 4:2:2.

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 sous-paragraphes ci-dessous présentent des méthodes permettant de quantifier les informations spatiales et temporelles des scènes d'essai étant applicables aux essais de qualité vidéo, aussi bien actuels que futurs. La position de la scène vidéo dans la matrice spatio-temporelle est importante 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.

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^{e} ligne et dans la j^{e} colonne de la n^{e} 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 aux débits spécifiés pour les classes TV3, MM4, MM5 et MM6 dans le Tableau D.2 ci-après. D'autres méthodes d'essai sont décrites dans les Appendices III et IV.

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.

Le chronogramme pour la présentation du stimulus est décrit sur la Figure 1. Si le temps imparti aux votes est constant (par exemple dans le cas où plusieurs visualisateurs assistent simultanément à la présentation d'une même bande), alors il sera inférieur ou égal à 10 s. La durée de la présentation pourra être réduite ou augmentée, selon le contenu du matériel expérimental.

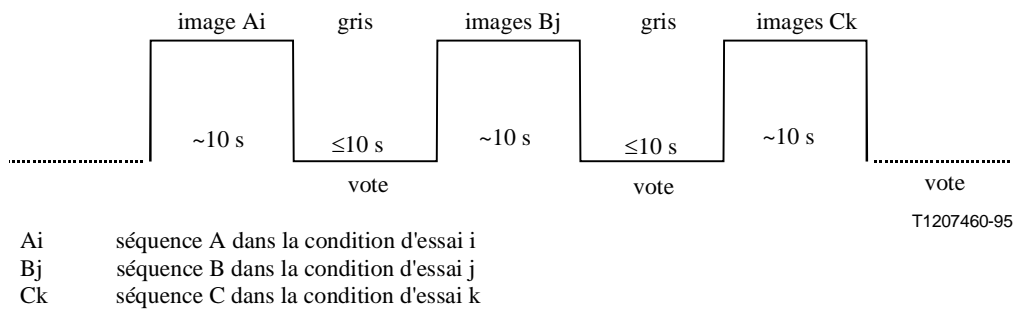


Figure 1 – Présentation du stimulus dans la méthode ACR

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 répétitions nécessaires en reprenant les mêmes conditions d'essai à différents instants de l'essai.

6.2 Evaluation par catégories absolues avec référence cachée (ACR-HR)

La méthode d'évaluation par catégories absolues avec référence cachée consiste à porter un jugement pour ranger indépendamment, dans une échelle de catégories, des séquences d'essai présentées les unes après les autres. Cette procédure d'essai doit contenir une version de référence de chaque séquence d'essai, présentée comme n'importe quel autre stimulus d'essai. On parle de référence cachée. Lors de l'analyse des données, on calculera une note différentielle de qualité (DMOS) entre chaque séquence d'essai et la référence (cachée) correspondante. Cette procédure est connue sous le terme de "référence cachée".

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

Le chronogramme pour la présentation du stimulus est décrit sur la Figure 1. Si le temps imparti aux votes est constant (par exemple dans le cas où plusieurs visualisateurs assistent simultanément à la présentation d'une même bande), alors il sera inférieur ou égal à 10 s. La durée de la présentation pourra être réduite ou augmentée, selon le contenu du matériel expérimental.

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

Les notes différentielles de visualisateur (DV) sont calculées pour chaque sujet et pour chaque séquence vidéo traitée (PVS). La référence cachée correspondante (REF) est utilisée afin de calculer la note DV au moyen de la formule suivante:

$$DV(PVS) = V(PVS) - V(REF) + 5$$

où V est l'évaluation ACR du visualisateur. Lorsque l'on utilise cette formule, une note DV de 5 correspond à une qualité "excellente" et une note de 1 à une "mauvaise" qualité. Une note DV supérieure à 5 (c'est-à-dire lorsque la séquence traitée est jugée de meilleure qualité que la séquence de référence cachée associée) sera généralement considérée comme valide. On peut aussi appliquer à ces notes une fonction d'écrasement pour éviter que ces notes de visualisateur ACR-HR (DV) n'influent outre mesure sur la note moyenne d'opinion globale:

$$DV_écrasée = (7*DV)/(2+DV) \text{ lorsque } DV > 5.$$

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-HR, on obtient le nombre de réplifications nécessaires en reprenant les mêmes conditions d'essai à différents instants de l'essai.

La méthode ACR-HR doit uniquement être utilisée lorsque la vidéo de référence est jugée par un expert dans le domaine comme étant de qualité "excellente" ou "bonne" sur l'échelle à cinq niveaux ci-dessus.

La méthode ACR-HR peut ne pas convenir pour l'analyse de dégradations inhabituelles qui se produisent au cours de la première ou de la dernière seconde de la séquence vidéo. Le visualisateur n'étant pas assez familier avec la séquence vidéo de référence, une dégradation qui serait évidente dans d'autres circonstances peut ne pas être relevée (par exemple, si la séquence s'interrompt juste avant la fin, un visualisateur peut ne pas pouvoir déterminer s'il s'agit du contenu normal ou d'une erreur liée au réseau).

6.3 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. Si le temps imparti aux votes est constant (par exemple dans le cas où plusieurs visualisateurs assistent simultanément à la présentation d'une même bande), alors il sera inférieur ou égal à 10 s. La durée de la présentation pourra être réduite ou augmentée, selon le contenu du matériel expérimental.

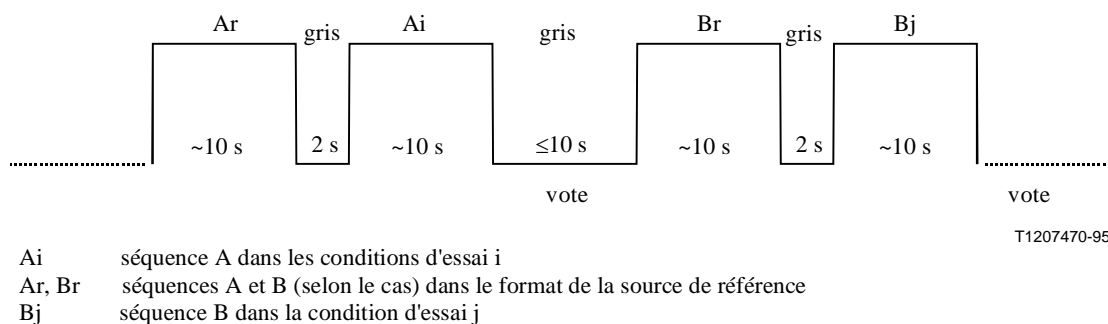


Figure 2 – Présentation du stimulus dans la méthode DCR

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 DCR, on obtient le nombre de répétitions nécessaires en reprenant les mêmes conditions d'essai à différents instants de l'essai.

6.4 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. Si le temps imparti aux votes est constant (par exemple dans le cas où plusieurs visualisateurs assistent simultanément à la présentation d'une même bande), alors il sera inférieur ou égal à 10 s. La durée de la présentation pourra être réduite ou augmentée, selon le contenu du matériel expérimental; mais elle doit être d'environ 10 s.

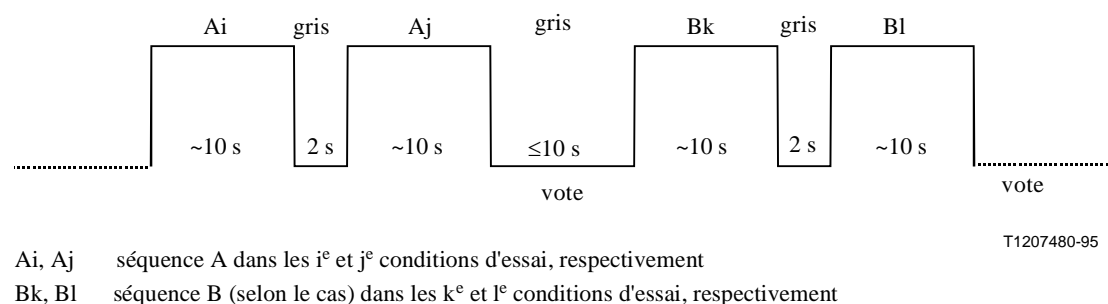


Figure 3 – Présentation du stimulus dans la méthode PC

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 réplifications 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 Recommandations [UIT-R BT.500-9] et [UIT-T P.800].

6.5 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, ACR-HR 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 la Recommandation [UIT-R BT.500-9], 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.

La méthode ACR-HR présente tous les avantages de la méthode ACR en termes de présentation et de rapidité. L'avantage principal de la méthode ACR-HR par rapport à la méthode ACR est que les incidences liées à la perception de la vidéo de référence peuvent être soustraites des notes subjectives. L'influence des préférences liées aux scènes présentées (par exemple, des visualisateurs à qui la vidéo de référence plaît ou non), de la qualité de la vidéo de référence (par exemple, de légères différences de qualité liées à la caméra) et du moniteur (par exemple, de qualité professionnelle ou destiné au grand public) sur les notes finales est ainsi réduite. La méthode ACR-HR convient bien aux expériences de grande ampleur, sous réserve que toutes les vidéos de référence soient au moins de "bonne" qualité. Toutefois, la méthode ACR-HR peut ne pas rendre compte de certaines dégradations qui sont facilement détectées dans le cadre d'une méthode différentielle directe (par exemple, la méthode DCR). Par exemple, une réduction systématique du gain de couleur (par exemple, des couleurs ternes) peut ne pas être détectée avec la méthode ACR-HR.

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.

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.

6.6 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 [UIT-T P.930] 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.

6.7 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., [b-Kirk], 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) répliques (c'est-à-dire répétitions de conditions identiques) dans l'expérience. Il y a plusieurs raisons pour utiliser des répliques, la plus importante étant que les données répliquées permettent de mesurer la "fidélité 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 fidélité intra-individuelle.

Les répliques permettent de calculer la fidélité 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

Le Tableau 1 énumère les conditions de visualisation habituellement utilisées pour évaluer la qualité des images de télévision; il y a lieu de spécifier les valeurs de paramètre effectivement utilisées. Pour la comparaison des résultats, toutes les conditions doivent être identiques et constantes dans tous les laboratoires effectuant le même type d'essais.

Le récepteur de contrôle doit être du type et de la taille appropriés pour l'application en question. Lorsque les séquences sont visualisées sur un écran d'ordinateur PC, il faut spécifier les caractéristiques d'affichage telles que la définition de l'écran, le type de carte vidéo, etc.

En ce qui concerne le format de visualisation, il est préférable d'utiliser le plein écran mais si, pour une raison ou une autre, les séquences doivent être affichées dans une fenêtre de l'écran, le fond de l'écran doit être un gris à 50% correspondant aux coordonnées $Y=U=V=128$ (U étant des valeurs/sans signe).

7.1 Conditions de visualisation

L'essai doit avoir lieu dans les conditions suivantes:

Tableau 1 – Conditions de visualisation

Paramètre	Valeur
Distance de visualisation (Note 1)	1-8 H (Note 2)
Luminance (valeur crête) de l'écran	100-200 cd/m (Note 2)
Rapport de la luminance du fond à la luminance maximale de l'écran	$\leq 0,05$
Rapport de la luminance de l'écran affichant du noir dans l'obscurité complète à celle correspondant au blanc (valeur crête)	$\leq 0,1$
Rapport de la luminance du fond (derrière le moniteur) à la luminance maximale de l'image (Note 3)	$\leq 0,2$
Chromaticité du fond (Note 4)	D ₆₅
Eclairage du fond du local (Note 3)	≤ 20 lux

NOTE 1 – 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. A ce propos, il convient de déterminer à l'avance la distance de visualisation préférée pour les essais de qualification, qui dépend généralement de l'application.

NOTE 2 – La lettre H correspond à la hauteur de l'image. Il faut 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.

NOTE 3 – Cette valeur correspond à un réglage se prêtant le mieux à la détection des distorsions; selon l'application, on peut utiliser des valeurs plus élevées ou moins élevées.

NOTE 4 – Dans le cas d'un moniteur d'ordinateur PC, on peut adapter la chromaticité du fond à celle de l'écran.

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, à cristaux liquides, plasma, de projection ou d'un autre type, en tenant compte du type d'application et du but de l'expérience. Les dimensions et le type du moniteur utilisé devront convenir à l'application étudiée.

Les moniteurs devront être alignés conformément aux procédures définies dans la Recommandation [UIT-R BT.814-1].

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.

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^e d'un optotype normal [b-Snellen]. 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 manquer plus de 2 planches sur 12 [b-Beck].

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 fidélité des sujets à partir des données et la méthode utilisée à cette fin sera indiquée. Certains critères de fidélité subjective sont donnés dans la Recommandation [UIT-R BT.500-9] et le Rapport [CEI/TR 60268-13].

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 2 pour la méthode ACR.

Tableau 2 – 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 [b-Manuel UIT-T], 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.

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 œuvre le filtre de Sobel en convoluant deux noyaux de 3×3 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^{e} ligne et à la j^{e} colonne. Le terme $Gv(i, j)$ sera le résultat de la première convolution et sera calculé comme suit:

$$\begin{aligned} Gv(i, j) = & -1 \times x(i-1, j-1) - 2 \times x(i-1, j) - 1 \times x(i-1, j+1) + \\ & + 0 \times x(i, j-1) + 0 \times x(i, j) + 0 \times x(i, j+1) + \\ & + 1 \times x(i+1, j-1) + 2 \times x(i+1, j) + 1 \times x(i+1, j+1) \end{aligned}$$

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

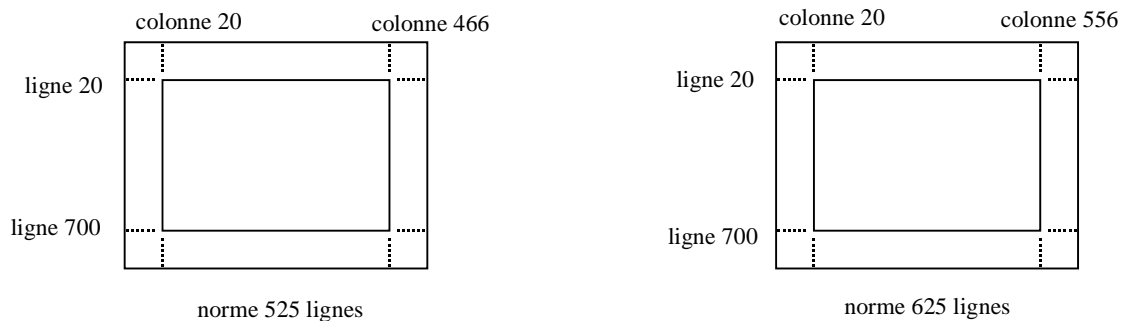
$$\begin{aligned} Gh(i, j) = & -1 \times x(i-1, j-1) + 0 \times x(i-1, j) + 1 \times x(i-1, j+1) + \\ & - 2 \times x(i, j-1) + 0 \times x(i, j) + 2 \times x(i, j+1) + \\ & - 1 \times x(i+1, j-1) + 0 \times x(i+1, j) + 1 \times x(i+1, j+1) \end{aligned}$$

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

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

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 la Recommandation [UIT-R BT.601-4].



T1207490-95

Figure A.1 – Sous-images à utiliser pour calculer SI et TI pour formats 525 et 625 lignes de la Recommandation [UIT-R BT.601-4]

On pourra trouver de plus amples renseignements sur le filtre de Sobel dans [b-Gonzalez].

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 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]. Le Tableau A.1 énumère, selon leurs catégories de contenu, les scènes d'essai de cet exemple.

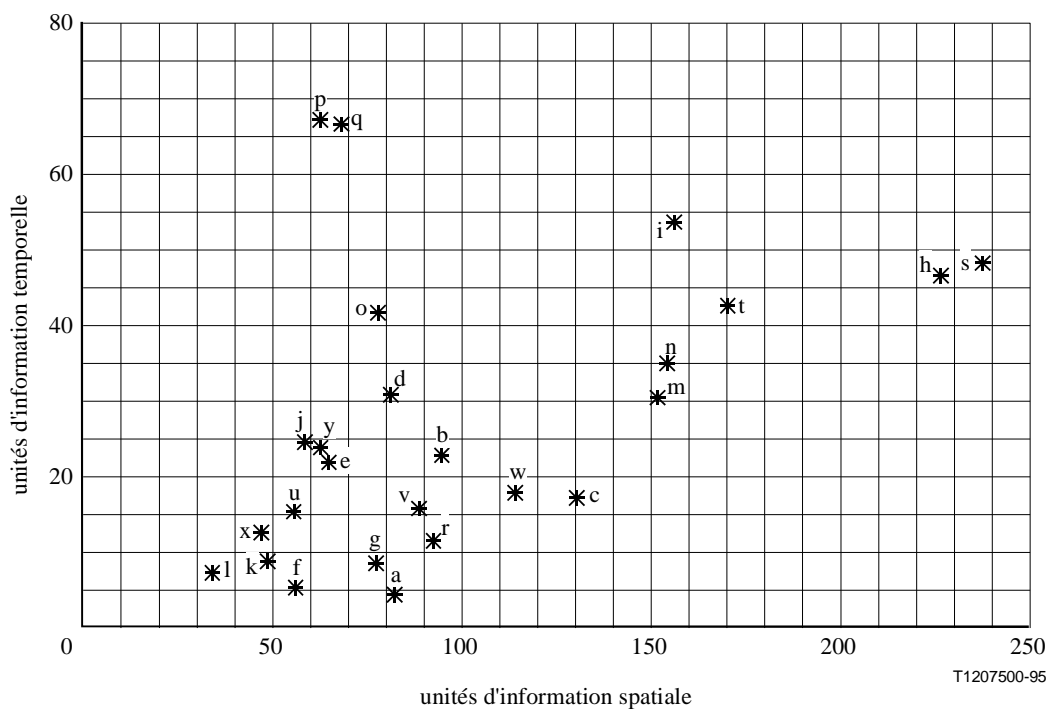


Figure A.2 – Exemple de tracé spatio-temporel d'un ensemble de scènes d'essai

Tableau A.1 – 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 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 ou de caméra (exemples de TV radiodiffusée)	flogar(h), fball(i), fedas(y)

Annexe B

Echelles d'évaluation supplémentaires

(Cette Annexe fait partie intégrante de la présente Recommandation.)

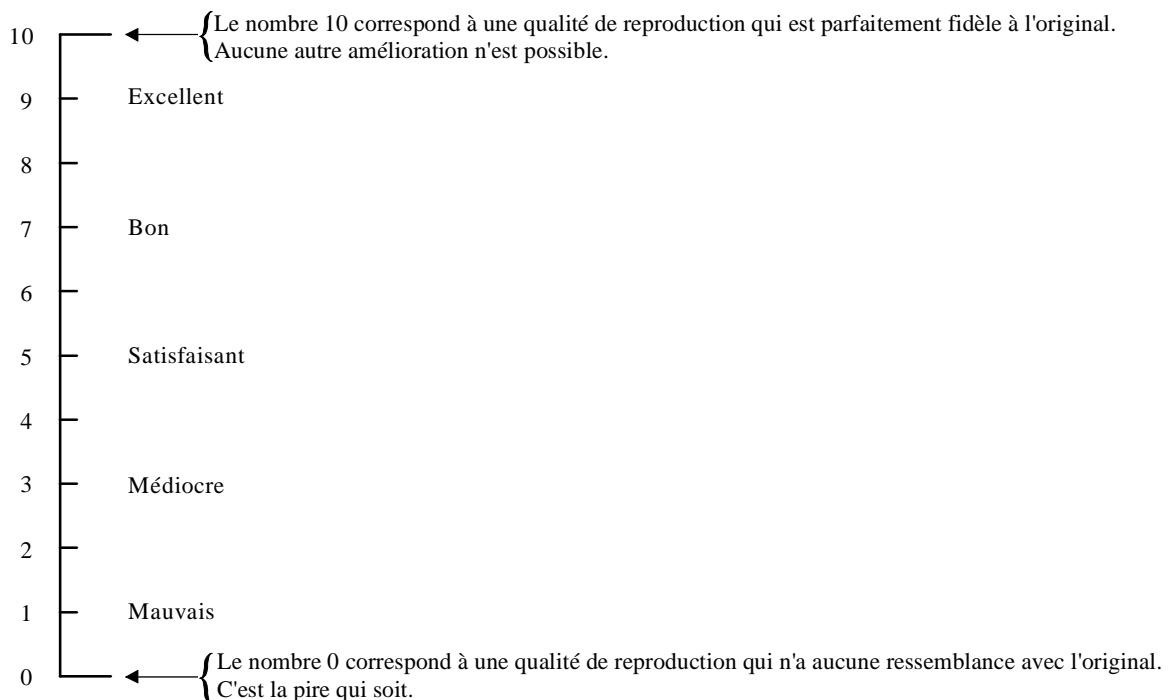
B.1 Echelles d'évaluation

Il est souvent nécessaire de faire appel à des échelles d'évaluation comportant plus de cinq degrés, surtout pour l'évaluation de codecs vidéo à faible débit. 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 paragraphe 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 – 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 clairement décrite au cours de la phase d'apprentissage. Voir également le Rapport [CEI/TR 60268-13] et [b-Manuel UIT-T], section 2.6, échelle a).



T0407280-96

Figure B.2 – 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 [b-Virtanen].

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.

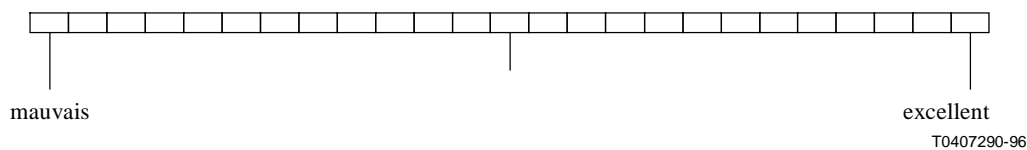


Figure B.3 – Echelle quasi-continue pour l'évaluation de la qualité

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);
- sautellement (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 [b-RACE].

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.

Questionnaire

Auriez-vous l'amabilité de répondre aux questions suivantes, concernant la dernière séquence présentée?

Vous pouvez exprimer votre opinion en cochant les échelles ci-dessous.

1) Comment évalueriez-vous les couleurs des images?

2) Comment évalueriez-vous le contraste des images?

3) Comment évalueriez-vous les bords des images?

4) Comment évalueriez-vous la continuité des mouvements?

5) Avez-vous remarqué un quelconque papillotement dans la séquence? Oui Non

Si vous avez remarqué du papillotement, veuillez l'évaluer sur l'échelle ci-dessous.

6) Avez-vous remarqué un quelconque maculage dans la séquence? Oui Non

Si vous avez remarqué du maculage, veuillez l'évaluer sur l'échelle ci-dessous.

NOTE – Lorsque ces échelles sont utilisées, toutes les catégories de qualité ou de dégradation prises en compte (comme la continuité des mouvements, le papillotement, le maculage, etc.) doivent être expliquées clairement au cours des sessions d'apprentissage.

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.

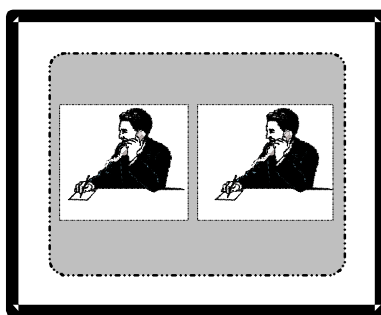
La méthode de la présentation simultanée exige des précautions particulières afin que les sujets évitent tout biais dû au type de présentation.

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 paragraphe 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 pouces.



T1207510-95

Figure C.1 – Position relative des deux séquences en présentation simultanée

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.

Annexe D

Classes vidéo et audio et leurs attributs

(Cette Annexe fait partie intégrante de la présente Recommandation.)

Dans la présente Recommandation, la qualité vidéo la plus élevée qui soit prise en considération est le format vidéo 4:2:2 à composantes Y, C_R, C_B codées en MIC linéaire à 8 bit/pixel selon la Recommandation [UIT-R BT.601].

Tableau D.1 – Définition des classes vidéo

TV 0	Classe vidéo sans pertes selon la Recommandation [UIT-R BT.601], 8 bits par pixel, utilisée pour des applications sans compression.
TV 1	Classe vidéo utilisée pour la post-production complète avec de nombreuses couches de correction et de traitement, en transmission par réseau interne de studio. Classe également utilisée pour la transmission de site distant à studio. Perçue comme transparente en comparaison de la classe TV 0.
TV 2	Classe vidéo utilisée pour des modifications simples, des corrections peu nombreuses, des incrustations de caractères/logos, l'insertion de programmes et la transmission entre maillons. En radiodiffusion, il s'agira par exemple d'une transmission de réseau à filiale. Autres exemples: liaison descendante régionale d'un système câblé vers une tête de réseau locale; système de vidéoconférence de haute qualité. Perçue comme étant presque transparente en comparaison de la classe TV 0.
TV 3	Classe vidéo utilisée pour l'acheminement vers les foyers privés/consommateurs (sans modification). Autres exemples: système câblé entre tête de réseau locale et terminal privé; système de vidéoconférence de qualité moyenne à élevée. Légers défauts présents par rapport à la classe TV 2.
MM 4	Toutes les trames sont codées. Légers défauts par rapport à la classe TV 3. Visioconférence de qualité moyenne. Généralement ≥ 30 images/s.
MM 5	Trames pouvant être perdues dans le codeur. Apparition possible de défauts perceptibles mais niveau de qualité utile pour des tâches précises, comme la visioconférence de qualité inférieure.
MM 6	Série d'images fixes. Classe non destinée à la cinématographie totale (exemples d'utilisation: surveillance, affichages graphiques).

Tableau D.2 – Attributs des classes vidéo

Classe vidéo	Format spatial	Fréquence des trames reçues (Note 1)	Variation typique du délai dû à la latence (Note 2)	Débit vidéo nominal (Mbit/s)
TV 0	[UIT-R BT.601]	Max FR	(Note 2)	270
TV 1	[UIT-R BT.601]	Max FR	(Note 2)	18 à 50
TV 2	[UIT-R BT.601]	Max FR	(Note 2)	10 à 25
TV 3	[UIT-R BT.601]	Max FR occasionnel Répétition de trames	(Note 2)	1,5 à 8
MM 4a	[UIT-R BT.601]	~30 ou ~25 fps	Délai ≈ 150 ms Variation ≈ 50 ms	~1,5

Tableau D.2 – Attributs des classes vidéo

Classe vidéo	Format spatial	Fréquence des trames reçues (Note 1)	Variation typique du délai dû à la latence (Note 2)	Débit vidéo nominal (Mbit/s)
MM 4b	CIF	~30 ou ~25 fps	Délai ≈ 150 ms Variation ≈ 50 ms	~0,7
MM 5a	CIF	10-30 fps	Délai ≈ 1000 ms Variation ≈ 500 ms	~0,2
MM 5b	\leq CIF	1-15 fps	Délai ≈ 1000 ms Variation ≈ 500 ms	~0,05
MM 6	CIF-16CIF	Limite $\rightarrow 0$ fps	Pas de restrictions	$<0,05$, Limite $\rightarrow 0$ fps

NOTE 1 – Normalement, 30 fps (trames par seconde) pour les systèmes à 525 lignes et 25 fps pour les systèmes à 625 lignes.

NOTE 2 – Les systèmes de radiodiffusion possèdent tous une latence constante mais non nécessairement faible, dans un seul sens ainsi qu'une variation de délai (temps de propagation) constante. Dans la plupart des applications, la latence sera faible, c'est-à-dire comprise entre 50 ms et 500 ms. Pour la vidéoconférence de haute qualité et pour les types d'applications conversationnelles en général, il y a lieu que la latence soit de préférence inférieure à 150 ms (voir la Recommandation [b-UIT-T G.114]). Des variations de délai sont autorisées dans l'étendue indiquée mais ne devraient pas conduire à des effets de désynchronisation trop perceptibles.

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 paragraphes 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 [b-Rapport 1213 du CCIR].

D'autres séquences de la vidéothèque du CCIR 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 la Recommandation [UIT-R BT.601-4], en version 525 comme en 625 lignes.

Tableau I.1 – 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, ACR-HR, 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 œuvre spécifique.

II.1 Méthodes ACR et ACR-HR

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.

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 cochez la case 1 si vous préférez la première séquence ou la case 2 si vous préférez la deuxième séquence de la paire:

Observez avec attention la paire de séquences vidéo dans leur entier avant d'exprimer votre jugement.

Appendice III

Le double stimulus pour une évaluation continue

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

La méthode du double stimulus pour une évaluation continue (SDSCE, *simultaneous double stimulus for a continuous evaluation*) est indiquée pour évaluer les effets de dégradations éparses, telles que les erreurs de transmission, sur la fidélité de l'information visuelle. Elle est dérivée de la méthode d'évaluation continue de la qualité avec stimulus unique (SSCQE) décrite dans la Recommandation [UIT-R BT.500-9].

III.1 Procédure d'essai

Le groupe d'observateurs visionne simultanément deux séquences, la version de référence et la version soumise à l'essai. Si toutes deux sont au format SIF ou plus petit, elles peuvent être visualisées sur le même moniteur; sinon, il convient d'utiliser deux moniteurs côte à côte.

Les observateurs sont invités à juger les différences entre les deux séquences et à donner une évaluation de la fidélité de l'information vidéo en positionnant le curseur du dispositif de vote. Pour une fidélité totale, le curseur doit être mis en haut de l'échelle (valeur 100), pour une fidélité nulle, au bas de l'échelle (valeur 0).

Les observateurs, qui savent laquelle des deux versions est celle de référence, sont invités à se prononcer tout au long de l'essai.

III.2 Phase préparatoire

Il s'agit d'une phase décisive qui a pour but d'éviter tout malentendu au sujet de la procédure à suivre. Il conviendra de donner aux observateurs des instructions écrites afin qu'ils disposent tous des mêmes informations. Celles-ci comporteront des explications sur les séquences à visionner, sur ce qu'il y a lieu d'évaluer (c'est-à-dire les différences de qualité) et sur la manière d'exprimer leur appréciation. Le responsable de l'essai répondra à toutes les questions posées par les observateurs afin d'éviter autant que possible toute équivoque.

La préparation sera suivie d'une démonstration qui permettra aux observateurs de se familiariser avec la procédure de notation et la nature des dégradations.

Enfin, il conviendra de faire un essai simulé dans des conditions représentatives avec des séquences différentes de celles de l'essai qui seront visionnées les unes après les autres, sans interruption.

A la fin de l'essai simulé, le responsable devra s'assurer que pour une évaluation portant sur des séquences d'essai et de référence identiques, les notes attribuées sont proches de 100. Si ce n'est pas le cas, il devra recommencer les explications et l'essai simulé.

III.3 Caractéristiques du protocole d'essai

Les définitions suivantes sont utilisées dans la description du protocole d'essai:

- *Segment vidéo (VS)*: correspond à une séquence vidéo.
- *Condition d'essai (TC)*: une condition TC peut être un processus vidéo spécifique, une condition de transmission ou les deux. Chaque segment VS doit être traité en fonction d'une condition TC au moins. Il faudra ajouter des références à la liste des conditions TC pour établir les paires de séquences "référence/référence" qu'il y a lieu d'évaluer.
- *Session (S)*: une session est une série de combinaisons VS/TC différentes sans séparation, en ordre pseudo-aléatoire. Chaque session comporte au moins une fois chaque segment VS et chaque condition TC, mais pas nécessairement toutes les combinaisons des deux. Toutes les

combinaisons VS/TC doivent être évaluées par le même nombre d'observateurs (mais pas nécessairement les mêmes).

- *Présentation d'essai (TP)*: une série de sessions englobant toutes les combinaisons VS/TC.
- *Phase de notation*: chaque observateur est invité à se prononcer tout au long de la session.

III.4 Traitement des données

Dès qu'un essai a eu lieu, on dispose d'un ou de plusieurs fichiers de données contenant toutes les notes attribuées lors des différentes sessions, qui représentent l'ensemble des résultats de la présentation d'essai (TP). On peut faire une première vérification de la validité des données en s'assurant que chaque combinaison VS/TC a été prise en considération et qu'un nombre identique de votes a été attribué à chacune d'elles.

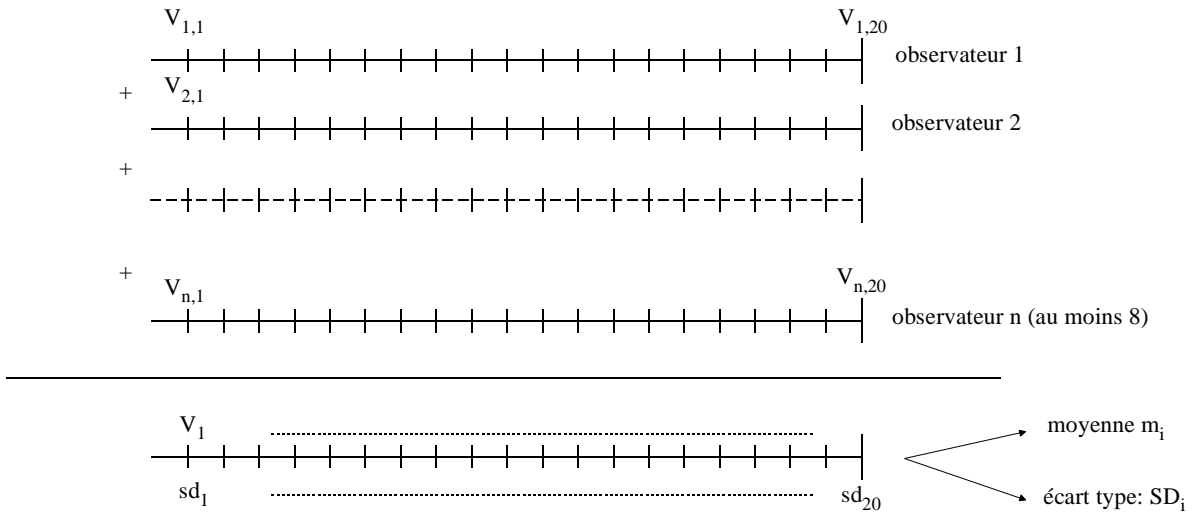
Les données des essais effectués conformément à ce protocole peuvent être traitées de trois manières différentes:

- par l'analyse statistique de chaque segment VS individuel;
- par l'analyse statistique de chaque condition TC individuelle;
- par l'analyse statistique globale de toutes les combinaisons VS/TC.

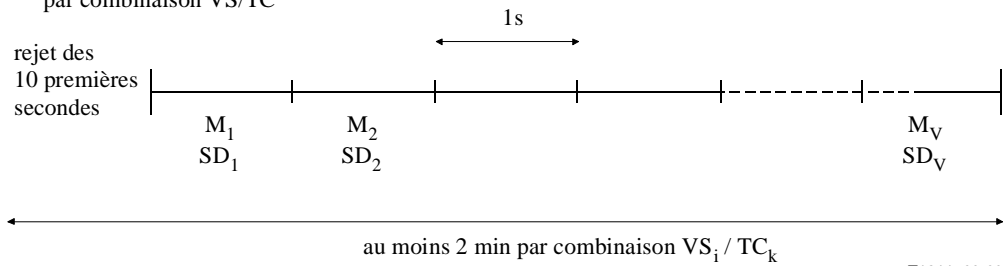
Dans chaque cas, l'analyse nécessite plusieurs étapes:

- les moyennes et les écarts types sont calculés sur la base du total de chaque série de notes attribuées à un instant donné de l'évaluation (Figure III.1);
- chaque segment vidéo est ensuite considéré comme une compilation de segments d'évaluation d'une durée maximale de 10 secondes chacun. Du fait qu'il n'y a aucun effet de mémoire récente (ou effet d'indulgence) dans l'évaluation de ces séquences de 10 secondes à peine, on calcule pour chaque segment la moyenne et l'écart type des moyennes calculées à l'étape précédente (Figure III.1). Lorsqu'on souhaite des informations détaillées sur la variabilité de la qualité, il faut raccourcir la durée des segments d'évaluation (à une seconde environ). Le résultat de cette étape peut être présenté dans un diagramme temporel analogue à celui de la Figure III.2;
- on analyse ensuite la distribution statistique des moyennes calculées à l'étape précédente (c'est-à-dire celles qui correspondent à chaque segment d'évaluation) et leur fréquence d'apparition. Pour éviter tout effet de mémoire récente se rapportant aux combinaisons VS/TC antérieures, les 10 premières secondes d'évaluation de chaque échantillon VS/TC sont refusées. Un exemple est donné à la Figure III.3;
- on calcule les caractéristiques de gêne globale en totalisant les fréquences d'apparition. Ce faisant, il faut tenir compte des intervalles de confiance, comme décrit dans la Figure III.4. A cette fonction de distribution statistique cumulative correspond une caractéristique de gêne qui montre la relation entre les moyennes pour chaque segment d'évaluation et la fréquence cumulative d'apparition.

1) calcul de la note moyenne (V) et de l'écart type (sd) par série de notes attribuées à un instant donné lors de chaque séquence d'évaluation de chaque combinaison VS/TC



2) calcul de la moyenne (M) et de l'écart type (SD) par séquence d'évaluation d'une seconde par combinaison VS/TC



T1211700-99

Figure III.1 – Traitement des données

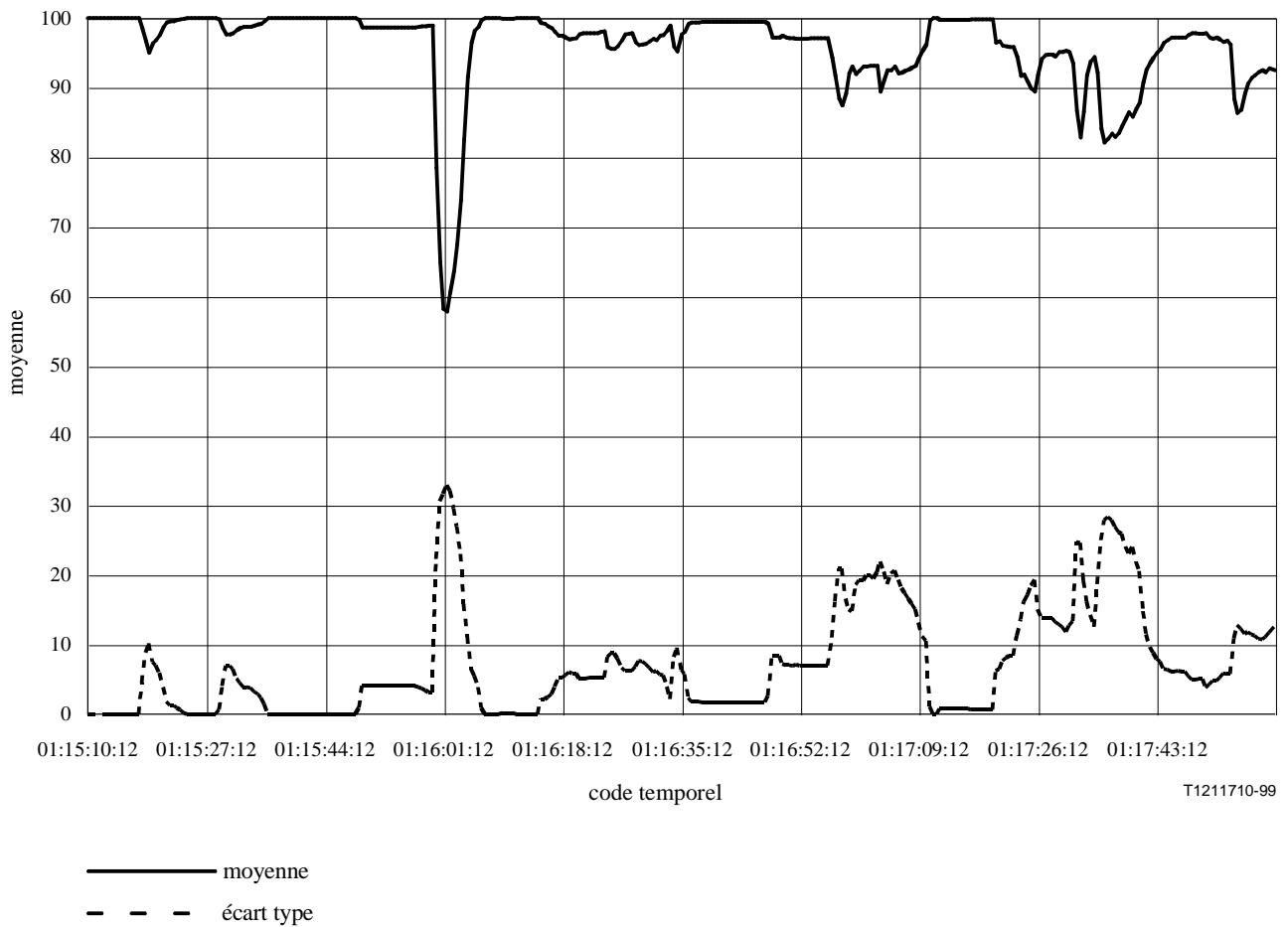


Figure III.2 – Diagramme temporel brut

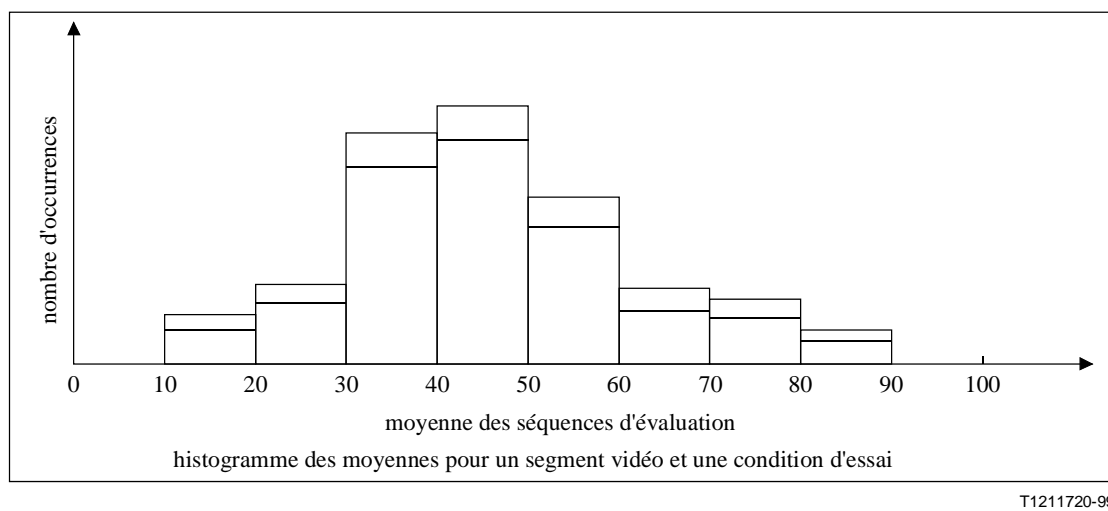


Figure III.3 – Relation entre les caractéristiques de dégradation et leur nombre d'occurrences

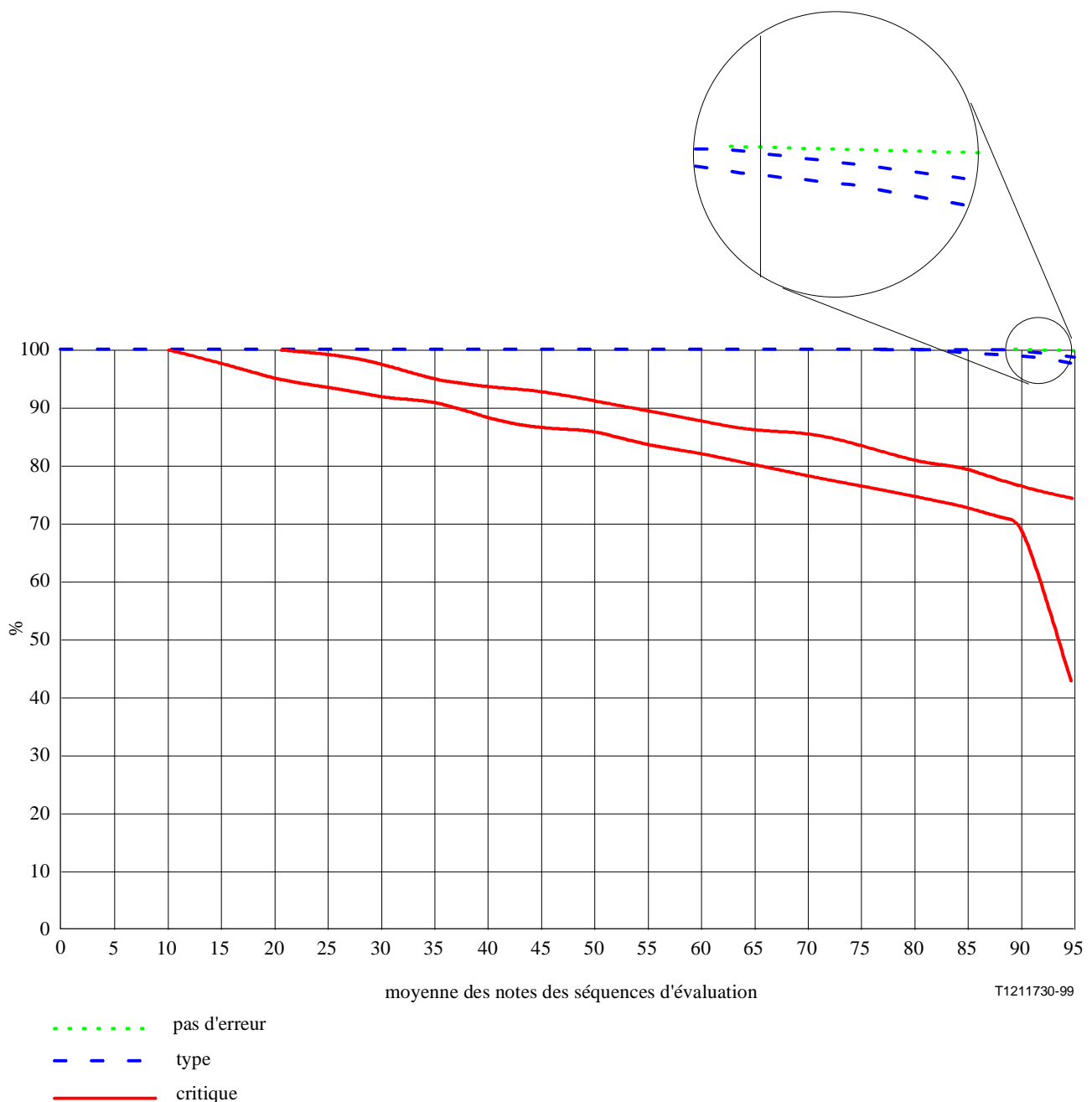


Figure III.4 – Caractéristiques de gêne globale calculées à partir des distributions statistiques en tenant compte de l'intervalle de confiance

III.5 Fidélité des observateurs

On peut évaluer qualitativement la fidélité des observateurs en notant leur comportement pendant la visualisation de la combinaison "séquence de référence/séquence de référence". Dans ces conditions, les observateurs devraient donner des notes très proches de 100, qui prouveraient que la procédure a été comprise et que les notes ne sont pas données au hasard.

Par ailleurs, on peut vérifier la fidélité des observateurs au moyen de procédures proches de celles décrites dans la Recommandation [UIT-R BT.500-9] pour la méthode SSCQE.

Dans la procédure SDSCE, la fidélité des notes dépend des paramètres suivants:

Décalages systématiques – Au cours d'un essai, un observateur est éventuellement trop optimiste ou trop pessimiste, ou peut avoir mal compris la procédure d'évaluation (la signification de l'échelle de notation). Cela peut se traduire par une série de notes systématiquement décalées vers le haut ou vers le bas par rapport à la moyenne, si elles ne sont totalement hors normes.

Anomalies ponctuelles – Comme c'est le cas pour d'autres procédures d'essai réputées, les observateurs attribuent parfois des notes sans porter une grande attention à l'observation et à l'examen de la qualité de la séquence visionnée. Dans ce cas, la courbe d'évaluation globale peut se situer "relativement" dans la moyenne, mais présenter néanmoins des anomalies ponctuelles.

Ces deux effets indésirables (comportement atypique et anomalie) peuvent être évités. La préparation des participants revêt bien sûr une importance considérable, mais il faudrait disposer d'un moyen permettant de détecter et, au besoin, de refuser des observateurs manquant de cohérence.

Appendice IV

L'évaluation axée sur l'objet

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

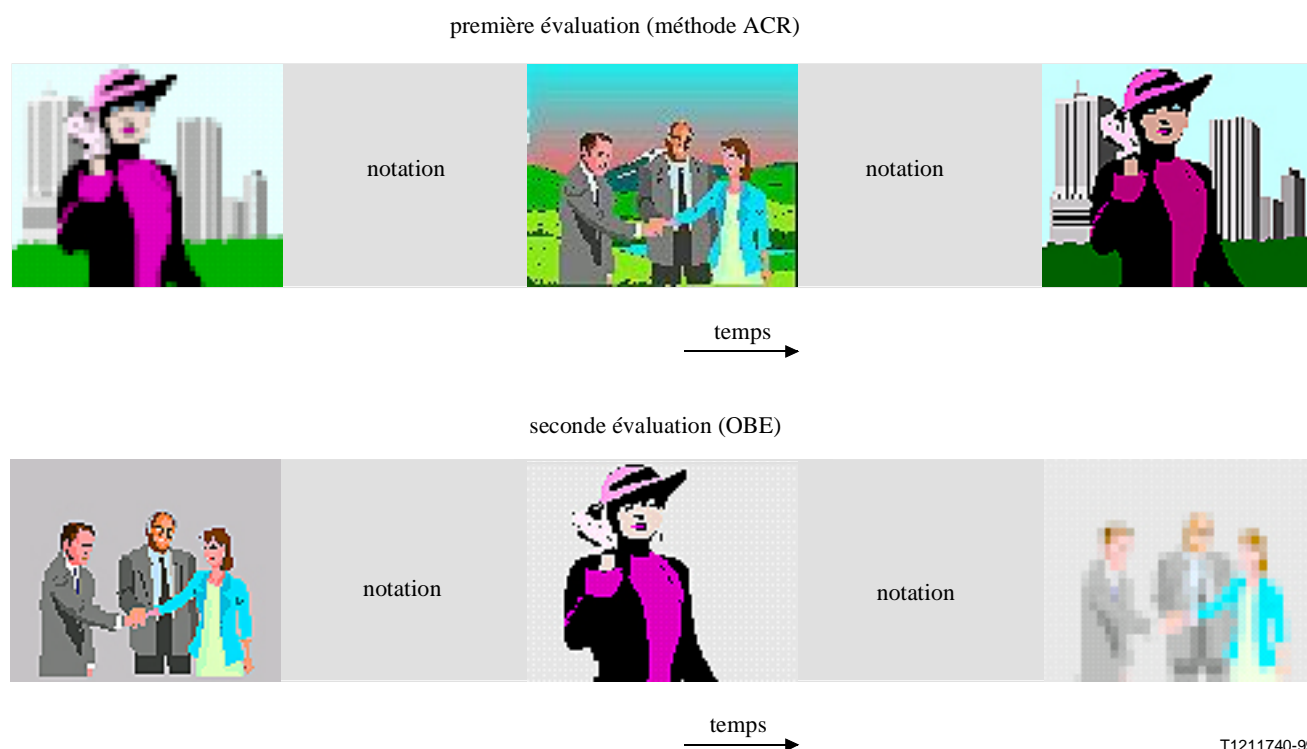
Il convient d'évaluer les fonctionnalités axées sur l'objet pour l'ensemble de la scène et pour les objets individuels. Généralement, en effet, une scène composée d'objets codés indépendamment peut être utilisée telle qu'elle a été produite, mais dans certains cas elle peut aussi être manipulée et chaque objet individuel peut être utilisé dans un contexte totalement différent. Pour ces raisons il est important d'établir un équilibre entre la qualité globale de l'ensemble de la scène et la qualité de la texture et des contours de chaque objet individuel.

Il convient en conséquence d'évaluer les fonctionnalités axées sur l'objet (qui sont l'échelonnabilité des objets et l'échelonnabilité de la qualité axée sur l'objet) en procédant par deux étapes:

Evaluation de l'image complète – il s'agit d'un test classique portant sur toute la séquence, contenant l'ensemble des objets VO. Comme méthode d'évaluation on peut utiliser l'ACR (voir le paragraphe 6.1) ou la DCR (voir le paragraphe 6.3), le choix dépendant de la gamme de débits et du caractère déterminant des séquences sources.

Evaluation axée sur l'objet (OBE, object-based evaluation) – dans cet essai, un seul objet VO est visualisé sur fond gris et les observateurs sont invités à en évaluer la qualité/dégradation (au moyen de la méthode utilisée pour évaluer l'image complète). Il est nécessaire de spécifier le pourcentage de bande qu'il y a lieu de consacrer à l'objet VO. Ce dernier sera extrait de la séquence utilisée pour évaluer l'image complète.

La Figure IV.1 illustre les deux essais nécessaires à l'évaluation de l'échelonnabilité des objets.



T1211740-99

Figure IV.1 – Essais pour évaluer l'échelonnabilité des objets

Dans le cas de l'échelonnabilité de la qualité axée sur l'objet, il convient de faire des essais pour évaluer les échelonnabilités spatiale et temporelle, cela uniquement par la méthode OBE.

Tant pour l'échelonnabilité spatiale que temporelle, il convient d'appliquer l'OBE qui permet d'évaluer en une seule opération les objets VO codés au débit "de base" et les mêmes objets VO codés aux débits améliorés spécifiés.

Généralement, l'évaluation des fonctionnalités axées sur l'objet doit tenir compte tant de la qualité de l'ensemble de l'image que de celle des objets individuels. On fera la première par des méthodes standard, la seconde par la méthode OBE.

Pour établir une comparaison entre les différents systèmes sur la base du codage axé sur l'objet, le responsable doit spécifier à l'avance la pondération relative qu'il convient d'accorder à la qualité globale et à la qualité des objets individuels.

Dans certains cas particuliers, il sera utile d'utiliser des critères d'évaluation axés sur la fonction. Dans l'évaluation d'un système de télésurveillance destiné à un garage, par exemple, l'échelonnabilité de la qualité doit être évaluée en termes de lisibilité des plaques minéralogiques des voitures. L'observateur portera son jugement sur chaque cas individuel, selon l'objectif de l'essai et la nature de l'application.

Enfin, on peut utiliser l'évaluation de la qualité axée sur l'objet pour déterminer l'effet de la qualité des objets individuels sur la qualité d'ensemble. Les résultats d'une telle étude pourraient servir à optimiser les systèmes de codage axés sur l'objet.

Appendice V

Une échelle d'évaluation supplémentaire pour DCR

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

On peut utiliser une échelle de dégradation à neuf degrés représentée sur la Figure V.1. Sur celle-ci, le degré 8 correspond au seuil de perceptibilité de la dégradation, à savoir le degré auquel l'observateur n'est pas absolument certain de percevoir la dégradation.

9	Imperceptible
8	
7	Perceptible mais non gênante
6	
5	Légèrement gênante
4	
3	Gênante
2	
1	Très gênante

Figure V.1 – Echelle de dégradation à neuf degrés numériques

Bibliographie

- [b-UIT-T G.114] Recommandation UIT-T G.114 (2003), *Temps de transmission dans un sens*.
- [b-UIT-T H.261] Recommandation UIT-T H.261 (1993), *Codec vidéo pour services audiovisuels à p x 64 kbit/s*.
- [b-UIT-T P.920] Recommandation UIT-T P.920 (1996), *Méthodes d'essai interactives pour communications audiovisuelles*.
- [b-UIT-R BT.812] Recommandation UIT-R BT.812 (1992), *Evaluation subjective de la qualité des images alphanumériques et graphiques en télétexte et dans des services similaires*.
- [b-UIT-R BT.815-1] Recommandation UIT-R BT.815-1 (1994), *Spécification d'un signal de mesure du contraste des dispositifs de visualisation*.
- [b-ISO/CEI 11172] ISO/CEI 11172:1993, *Codage de l'image animée et du son associé pour les supports de stockage numérique jusqu'à environ 1,5 Mbit/s*.
- [b-Beck] *Pseudo Isochromatic Plates* (1940), engraved and printed by The Beck Engraving Co., Inc., Philadelphia and New York, United States.
- [b-Gonzalez] Gonzalez, R.C. and Wintz, P. (1987), *Digital Image Processing*, 2nd Edition, Addison-Wesley Publishing Co., Reading, Massachusetts.
- [b-Guilford] Guilford, P. (1954), *Psychometric methods*, McGraw-Hill, New York.
- [b-Kirk] Kirk, R.E. (1982), *Experimental Design – Procedures for the Behavioural Sciences*, 2nd Edition, Brooks/Cole Publishing Co., California.
- [b-Manuel UIT-T] Manuel UIT-T (1993), *Manuel de téléphonométrie*, UIT, Genève.
- [b-RACE] RACE Industrial Consortium Project 1018 HIVITS, WP B5, *Picture Quality Measurement*, 1988.
- [b-Rapport 1213 du CCIR] Rapport 1213 du CCIR (1990), *Images et séquences d'essai pour l'évaluation subjective des codes numériques*, Annexe du Volume XI, Partie 1.
- [b-Snellen] Echelle de Snellen.
- [b-Virtanen] Virtanen, M.T., Gleiss, N. and Goldstein, M. (1995), *On the use of Evaluative Category Scales in Telecommunications*, Human Factors in Telecommunication Conference, Melbourne.

SÉRIES DES RECOMMANDATIONS UIT-T

Série A	Organisation du travail de l'UIT-T
Série D	Principes de tarification et de comptabilité et questions de politique générale et d'économie relatives aux télécommunications internationales/TIC
Série E	Exploitation générale du réseau, service téléphonique, exploitation des services et facteurs humains
Série F	Services de télécommunication non téléphoniques
Série G	Systèmes et supports de transmission, systèmes et réseaux numériques
Série H	Systèmes audiovisuels et multimédias
Série I	Réseau numérique à intégration de services
Série J	Réseaux câblés et transmission des signaux radiophoniques, télévisuels et autres signaux multimédias
Série K	Protection contre les perturbations
Série L	Environnement et TIC, changement climatique, déchets d'équipements électriques et électroniques, efficacité énergétique; construction, installation et protection des câbles et autres éléments des installations extérieures
Série M	Gestion des télécommunications y compris le RGT et maintenance des réseaux
Série N	Maintenance: circuits internationaux de transmission radiophonique et télévisuelle
Série O	Spécifications des appareils de mesure
Série P	Qualité de transmission téléphonique, installations téléphoniques et réseaux locaux
Série Q	Commutation et signalisation et mesures et tests associés
Série R	Transmission télégraphique
Série S	Equipements terminaux de télégraphie
Série T	Terminaux des services télématiques
Série U	Commutation télégraphique
Série V	Communications de données sur le réseau téléphonique
Série X	Réseaux de données, communication entre systèmes ouverts et sécurité
Série Y	Infrastructure mondiale de l'information, protocole Internet, réseaux de prochaine génération, Internet des objets et villes intelligentes
Série Z	Langages et aspects généraux logiciels des systèmes de télécommunication