



UNION INTERNATIONALE DES TÉLÉCOMMUNICATIONS

UIT-T

SECTEUR DE LA NORMALISATION
DES TÉLÉCOMMUNICATIONS
DE L'UIT

P.911

(12/98)

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

Qualité audiovisuelle dans les services multimédias

**Méthodes d'évaluation subjective de la qualité
audiovisuelle pour applications multimédias**

Recommandation UIT-T P.911

(Antérieurement Recommandations du CCITT)

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
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.

RECOMMANDATION UIT-T P.911

METHODES D’EVALUATION SUBJECTIVE DE LA QUALITE AUDIOVISUELLE POUR APPLICATIONS MULTIMEDIAS

Résumé

La présente Recommandation décrit des méthodes non interactives d'évaluation subjective portant sur la qualité audiovisuelle globale perçue d'applications multimédias telles que la vidéoconférence, la consultation et l'archivage de données, la télémédecine, etc. Ces méthodes peuvent être utilisées à diverses fins telles que la sélection d'algorithmes, la catégorisation des performances audiovisuelles d'un système, l'évaluation du niveau de qualité lors d'une connexion audiovisuelle, etc. Lorsque des aspects interactifs doivent être évalués, il y a lieu d'utiliser les méthodes de test conversationnel décrites dans la Recommandation P.920. La présente Recommandation décrit également les caractéristiques des séquences de source à utiliser, comme la durée, le type de contenu, le nombre de séquences, etc. Finalement, elle donne des indications sur la relation entre qualité audio, qualité vidéo et qualité audiovisuelle, telles qu'elles ont été déduites des résultats d'essais conduits indépendamment dans différents laboratoires.

Source

La Recommandation UIT-T P.911, élaborée par la Commission d'études 12 (1997-2000) de l'UIT-T, a été approuvée le 3 décembre 1998 selon la procédure définie dans la Résolution n° 1 de la CMNT.

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. 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.

La Conférence mondiale de normalisation des télécommunications (CMNT), qui se réunit tous les quatre ans, détermine les thèmes d'études 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 n° 1 de la CMNT.

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, le terme *exploitation reconnue (ER)* désigne tout particulier, toute entreprise, toute société ou tout organisme public qui exploite un service de correspondance publique. Les termes *Administration*, *ER* et *correspondance publique* sont définis dans la *Constitution de l'UIT (Genève, 1992)*.

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 responsables de la mise en œuvre de consulter la base de données des brevets du TSB.

© UIT 1999

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.

TABLE DES MATIÈRES

	Page
1	1
2	1
3	2
4	3
5	4
6	5
6.1	5
6.2	7
6.3	7
6.4	8
6.5	9
6.6	9
6.7	9
7	10
7.1	10
7.2	11
7.3	12
7.4	12
8	12
Annexe A – Détails relatifs à la caractérisation des séquences d'essai.....	13
Annexe B – Classes vidéo et audio et leurs attributs.....	14
B.1 Classes vidéo et leurs attributs	14
B.2 Classes audio et leurs attributs	15
Annexe C – Analyse de la Relation entre qualité audio, qualité vidéo et qualité audiovisuelle	17
Appendice I – Séquences audiovisuelles d'essai	18
Appendice II – Instructions pour les essais audiovisuels.....	19
II.1 Méthode ACR	19
II.2 Méthode DCR	19
II.3 Méthode PC.....	20
Appendice III – Bibliographie.....	20

Recommandation P.911

MÉTHODES D'ÉVALUATION SUBJECTIVE DE LA QUALITÉ AUDIOVISUELLE POUR APPLICATIONS MULTIMÉDIAS

(Genève, 1998)

1 Domaine d'application

La présente Recommandation décrit des méthodes non interactives d'évaluation subjective portant sur la qualité audiovisuelle globale perçue d'applications multimédias telles que la vidéoconférence, la consultation et l'archivage de données, la télé-médecine etc., aux débits spécifiés dans les Tableaux B.2 et B.4 pour les classes TV 3, MM 4, MM 5 et MM 6. Ces méthodes peuvent être utilisées à diverses fins telles que la sélection d'algorithmes, la catégorisation des performances audiovisuelles d'un système, l'évaluation du niveau de qualité lors d'une communication audiovisuelle, etc. Lorsque des aspects interactifs doivent être évalués, il y a lieu d'utiliser les méthodes de test conversationnel décrites dans la Recommandation P.920.

2 Références normatives

Les Recommandations UIT-T 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.601-5 (1995), *Paramètres de codage en studio de la télévision numérique pour des formats standards d'image 4:3 (normalisé) et 16:9 (écran panoramique).*
- [4] Recommandation UIT-R BT.500-8 (1998), *Méthodologie d'évaluation subjective de la qualité des images de télévision.*
- [5] Publication CEI/TR3 60268-13 (1998), *Equipements pour systèmes électroacoustiques – Partie 13: essais d'écoute des haut-parleurs.*
- [6] Recommandation UIT-T P.800 (1996), *Méthodes d'évaluation subjective de la qualité de transmission.*
- [7] Recommandation UIT-R BT.814-1 (1993), *Spécifications et méthodes de réglage de la brillance et du contraste des dispositifs de visualisation.*
- [8] Recommandation UIT-R BT.1128-2 (1997), *Evaluation subjective des systèmes de télévision classiques.*
- [9] Recommandation CCITT J.61 (1990), *Qualité de transmission des circuits de télévision destinés à être utilisés dans les communications internationales.*

- [10] Recommandation UIT-T P.810 (1996), *Appareil de référence à bruit modulé (MNRU)*.
- [11] Recommandation UIT-T P.910 (1996), *Méthodes subjectives d'évaluation de la qualité vidéographique pour les applications multimédias*.
- [12] Recommandation CCITT G.722 (1988), *Codage audiofréquence à 7 kHz à un débit inférieur ou égal à 64 kbit/s*.
- [13] Recommandation CCITT G.711 (1988), *Modulation par impulsions et codage (MIC) des fréquences vocales*.
- [14] Recommandation CCITT G.728 (1992), *Codage de la parole à 16 kbit/s en utilisant la prédiction linéaire à faible délai avec excitation par code*.
- [15] Recommandation UIT-T G.114 (1996), *Temps de transmission dans un sens*.
- [16] Recommandation UIT-R BS.775-1 (1993), *Système de son stéréophonique multicanal avec ou sans image associée*.
- [17] Publication CEI 60651 (1979), *Sonomètres*.
- [18] Recommandation UIT-T P.931 (1998), *Mesure du temps de transmission, de la synchronisation et du débit de trames dans les communications multimédias*.

3 Termes et définitions

La présente Recommandation définit les termes suivants:

3.1 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, 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".

3.2 gamma: la luminance de l'écran n'est pas en relation linéaire avec la tension du signal d'entrée, mais est proportionnelle à la puissance gamma de cette tension. Afin de compenser cette non-linéarité, on applique généralement dans la caméra un facteur de correction qui est une fonction inverse de gamma. Gamma possède également une influence sur le rendu des couleurs.

3.3 tests d'optimisation: tests 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 systèmes étudiés.

3.4 tests 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.

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

3.6 fiabilité d'un test 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 interindividuelle ("entre sujets différents") représente la constance d'un certain nombre d'évaluations de sujets différents dans la même condition d'essai.

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

3.8 information de perception spatiale (SI, *spatial perceptual information*): mesure 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. L'information de perception spatiale (SI) est fondée sur le filtre de Sobel. Chaque trame vidéo (plan de luminance) au temps n (F_n) passe par le filtre de Sobel [Sobel (F_n)]. L'écart type de la répartition des pixels ($\text{std}_{\text{space}}$) est ensuite calculé pour chaque trame issue du filtre de Sobel. Cette opération est répétée pour chaque trame de la séquence vidéo et donne une série temporelle d'informations spatiales relatives à la scène. La valeur maximale de cette série temporelle (max_{time}) est choisie pour représenter le contenu en informations spatiales de la scène. Ce processus peut être représenté sous forme d'équation par la relation suivante:

$$SI = \text{max}_{\text{time}} \{ \text{std}_{\text{space}} [\text{Sobel}(F_n)] \}$$

3.9 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. Elle n'est pas censée mesurer l'entropie ni être associée aux informations définies en théorie de la communication.

La mesure des informations temporelles (TI) est calculée comme la valeur maximale dans le temps (max_{time}) de l'écart type de la répartition dans l'espace ($\text{std}_{\text{space}}$) des points $M_n(i,j)$ pour tous les i et tous les j .

$$TI = \text{max}_{\text{time}} \{ \text{std}_{\text{space}} [M_n(i,j)] \}$$

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

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

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

3.10 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, au moyen d'un critère mathématique, le degré de conformité du signal traité avec le signal d'entrée ou avec le signal idéal. S'il n'y a aucune différence (c'est-à-dire distance = 0), 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 transparents à la perception.

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

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 (<i>confidence interval</i>)
CIF	format intermédiaire commun (format d'image défini dans la Recommandation H.261 pour la vidéophonie: 352 lignes × 288 pixels) (<i>common intermediate format</i>)
CRT	tube à rayons cathodiques (<i>cathode ray tube</i>)
DCR	évaluation par catégories de dégradation (<i>degradation category rating</i>)
%GOB	pourcentage d'opinions favorables ("bon" ou "excellent") (<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 d'opinions défavorables ("médiocre" ou "mauvais") (<i>percent of poor or worse</i>)
QCIF	quart de format CIF (format d'image défini dans la Recommandation H.261 pour la vidéophonie: 176 lignes × 144 pixels) (<i>quart CIF</i>)
SI	information spatiale (<i>spatial information</i>)
SIF	format intermédiaire normalisé [format d'image défini dans l'ISO/CEI 11172 (MPEG-1): 352 lignes × 288 pixels × 25 trames/s et 352 lignes × 240 pixels × 30 trames/s] (<i>standard intermediate format</i>)
S/N	rapport signal sur bruit (<i>signal-to-noise ratio</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. Il faut choisir des scènes audiovisuelles d'essai claires et appropriées de façon que leurs parties aussi bien vidéo qu'audio soient compatibles avec les services destinés à être fournis par le canal de service de transmission numérique. Il y a lieu que l'ensemble des scènes d'essai couvre toute l'étendue des informations spatiales et temporelles et comporte tout type de signal audio pouvant intéresser les utilisateurs des dispositifs mis à l'essai.

La durée des séquences de source devrait être d'environ 10 s mais sans être inférieure à 8 s. Il devrait s'agir de la durée réelle de la séquence, c'est-à-dire que les séquences ne peuvent pas être obtenues par répétition d'une séquence plus courte. La fin de la scène ne devrait pas tronquer une phrase vocale ou musicale. Une période de silence initiale et finale, non supérieure à 500 ms, peut rendre la séquence plus naturelle à l'écoute.

La qualité des séquences de source devrait être aussi élevée que possible: il y a lieu d'enregistrer le signal vidéo dans le format 4:2:2 décrit dans la Recommandation UIT-R BT.601 et le signal audio à une fréquence d'échantillonnage de 48 kHz avec au moins 16 bits par échantillon. Il convient que les signaux audio et vidéo soient synchronisés lorsque leur contenu l'exige.

L'Annexe A énumère les catégories audio et vidéo qui peuvent être utilisées pour caractériser une séquence audiovisuelle.

Lorsque l'expérimentateur souhaite comparer les résultats de différents laboratoires, il est nécessaire d'utiliser un ensemble commun de séquences d'essai pour éliminer toute autre source de variation.

Il y a lieu de définir les caractéristiques de l'environnement et des systèmes d'enregistrement conformément aux 5.1/P.910 et 5.2/P.910 et aux B.1.1/P.800 et B.1.3/P.800, respectivement pour la partie vidéo et pour la partie audio. Des exemples de scènes d'essai appropriées sont donnés dans l'Annexe C.

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

6 Méthodes d'essai et modèles expérimentaux

Le mesurage de la qualité perçue des séquences audiovisuelles 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 la nature de celui-ci (en l'occurrence la qualité audiovisuelle globale).

Un certain nombre de méthodes expérimentales ont été validées pour différents objets. Trois méthodes sont recommandées ci-dessous pour la gamme de débits et les applications indiquées dans le domaine d'application de la présente Recommandation.

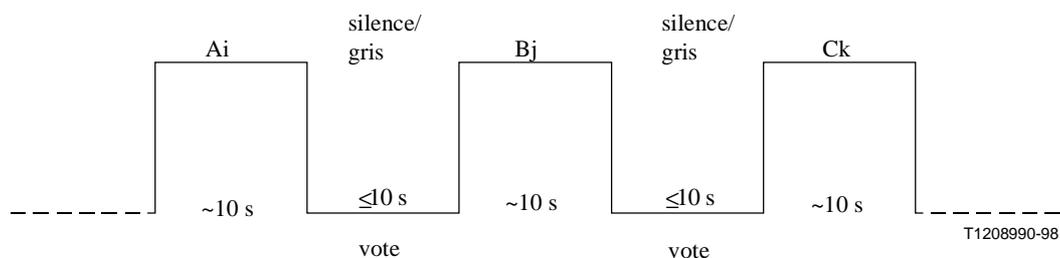
Le choix final d'une de ces méthodes pour une application particulière dépend de plusieurs facteurs, comme le contexte, le but recherché et le point du processus de développement auquel l'essai doit être exécuté.

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. (Ce protocole est du type décrit dans d'autres Recommandations de l'UIT-T et de l'UIT-R [11], [6] et [4]. En [4], ce type de protocole est appelé "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. Elle ne donne pas de référence explicite bien que les sujets utilisent toujours une référence implicite.

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.



Ai séquence A dans la condition d'essai i
 Bj séquence B dans la condition d'essai j
 Ck séquence C dans la condition d'essai k

Figure 1/P.911 – 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 (voir le Tableau 1):

Tableau 1/P.911 – Echelle de qualité à cinq niveaux

5	Excellent
4	Bon
3	Satisfaisant
2	Médiocre
1	Mauvais

Si une plus grande puissance de discrimination est requise (comme dans le cas de l'évaluation du codage à bas débit), il y a lieu de faire appel à l'échelle à neuf niveaux suivante (voir le Tableau 2):

Tableau 2/P.911 – Echelle de qualité à neuf niveaux

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

L'Annexe B/P.910 donne des exemples d'échelles numériques ou analogiques 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, *degradation category rating*)

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. (Ce protocole est du type décrit dans d'autres Recommandations de l'UIT-T et de l'UIT-R [11], [6] et [4]. En [4], ce type de protocole est appelé "méthode à double stimulus utilisant une échelle de dégradation".)

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.

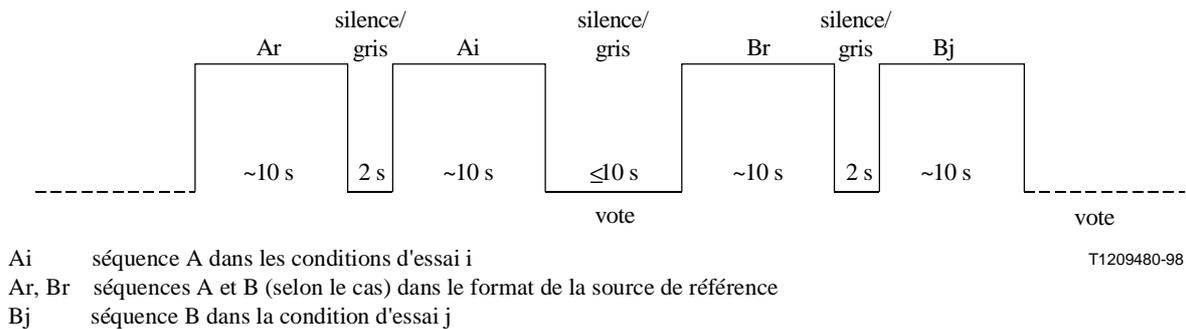


Figure 2/P.911 – Présentation du stimulus dans la méthode DCR

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 (voir le Tableau 3):

Tableau 3/P.911 – Echelle de dégradation à cinq niveaux

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

Dans la méthode DCR, 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. La séquence de source peut être incluse et sera traitée comme un système à l'essai additionnel.

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 devrait être d'environ 10 s.

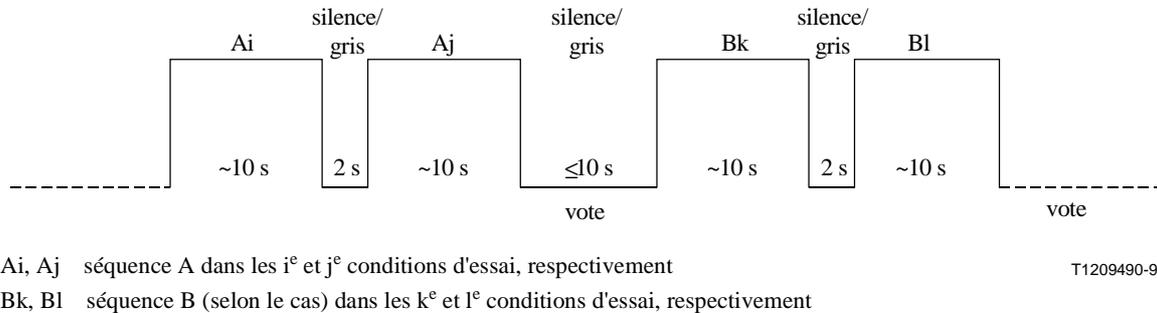


Figure 3/P.911 – Présentation du stimulus dans la méthode PC

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 [6].

6.4 Évaluation continue de la qualité par stimulus unique (SSCQE)

La méthode d'évaluation continue de la qualité par stimulus unique (SSCQE, *single stimulus continuous quality evaluation*) a été incluse dans la référence [4] afin de tenir compte des variations temporelles de la qualité dans les transmissions numériques.

Cette méthode porte sur des séquences de longue durée (de 3 à 30 min). Lorsque ces séquences sont affichées, l'évaluation de la qualité subjective est continue. Cette évaluation continue est effectuée au moyen d'éléments coulissants que les sujets doivent déplacer pendant qu'ils regardent et écoutent les programmes ou scénarios. Il importe de noter qu'aucune référence n'est donnée pour servir de base à l'évaluation subjective.

Les résultats sont présentés par tracé de courbes indiquant le pourcentage de temps pendant lequel la note subjective a été plus élevée qu'une note donnée sur une échelle de 0 à 100. Dans cette échelle, la valeur 100 représente une qualité parfaite du ou des services considérés.

Les avantages de cette méthode sont qu'elle est bien adaptée à tenir compte des variations temporelles de la qualité et à effectuer des évaluations globales de la qualité. Elle permet d'effectuer une comparaison entre services lorsque ceux-ci sont paramétrables. Elle peut donc être un outil utile pour la définition d'un service. De plus, cette méthode peut être utilisée aussi bien en laboratoire qu'au domicile de l'utilisateur. Un inconvénient de cette méthode est qu'elle n'a aucune référence, ce qui la rend peu adaptée aux essais qui nécessitent un degré élevé de discrimination, comme la comparaison entre codeurs.

L'applicabilité de la méthode d'évaluation continue de la qualité par stimulus unique (SSCQE) à la gamme de débits et au type des applications visés par la présente Recommandation fera l'objet d'un complément d'étude.

6.5 Comparaison des méthodes

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, PC et SSCQE). 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. 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.

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 les Recommandations P.930 [1] et P.810 [10] 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. (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) 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 "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 répliques permettent de calculer la fiabilité intra-individuelle et, si nécessaire, d'ignorer des résultats non fiables issus de certains sujets. Une estimation de l'écart type des lois aussi bien intra-individuelle qu'interindividuelle 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 et d'écoute

Le Tableau 4 énumère les conditions de visualisation et d'écoute normalement utilisées en évaluation de la qualité audiovisuelle. Il y a lieu de spécifier les valeurs particulières de paramétrisation. Pour la comparaison des résultats d'essai, toutes les conditions de visualisation et d'écoute doivent être identiques et constantes entre les laboratoires pour le même type d'essais.

Il convient qu'aussi bien les dimensions que le type du moniteur conviennent à l'application étudiée. Lorsque des séquences sont présentées au moyen d'un système informatique, les caractéristiques des transducteurs vidéo et audio doivent être spécifiées, par exemple le pas de masque du moniteur, le type de carte d'affichage vidéo utilisée, les caractéristiques des écouteurs ou des casques ou des haut-parleurs, etc.

Dans le cas d'une présentation par haut-parleurs, il convient en particulier d'indiquer le nombre et la position des haut-parleurs par rapport à l'image.

Pour les séquences d'essais, les paramètres d'exploitation tels que le niveau du signal doivent être corrélés à ceux du signal d'alignement utilisé pour vérifier les conditions de visualisation et d'écoute. Il convient de signaler tout ajustement opérationnel effectué pour que les séquences source ou les séquences traitées soient conformes à cette prescription.

Concernant le format d'affichage, il est préférable d'utiliser tout l'écran pour présenter les séquences. Lorsque cependant, pour une raison ou une autre, les séquences doivent être affichées dans une fenêtre de l'écran, la couleur du fond d'écran devrait être un gris à 50% correspondant aux coordonnées $Y = U = V = 128$ (U et V étant des valeurs non signées).

Il convient de mesurer la synchronisation entre le signal audio et le signal vidéo conformément à la Recommandation P.931 [18] et de la signaler.

Tableau 4/P.911 – Conditions normales de visualisation et d'écoute utilisées en évaluation de la qualité audiovisuelle

Paramètre	Réglage
Dimensions de la pièce (Note 7)	Spécifier L × W × H
Distance de visualisation (Note 5)	de 1 à 8 H
Luminance de l'écran (valeur de crête)	de 100 à 200 cd/m ²
Rapport de luminance d'écran inactif à luminance de crête	≤ 0,05
Rapport de luminance de l'écran au niveau de crête du blanc (lors de l'affichage d'un niveau de noir total dans une salle complètement obscure)	≤ 0,1
Rapport entre luminance de l'arrière-fond du moniteur d'image à la valeur de crête de la luminance d'image (Note 1)	≤ 0,2
Chromaticité de l'arrière-fond (Note 6)	D ₆₅
Eclairement lumineux d'ambiance de la salle (Note 1)	≤ 20 lx
Niveau de bruit de fond (Note 2)	≤ 30 dBA
Niveau d'écoute (Note 3)	~ 80 dBA
Durée de réverbération (Note 4)	< 500 ms, ∇f > 150 Hz
<p>NOTE 1 – Cette valeur correspond à un réglage autorisant une détectabilité maximale des distorsions. Pour certaines applications, des valeurs plus élevées sont autorisées ou sont déterminées par l'application.</p> <p>NOTE 2 – Si les niveaux de bruit de fond utilisés dans l'application sont nettement plus élevés, il y a lieu d'utiliser le bruit à spectre de Hoth (voir la Bibliographie [9]) pour les environnements comme les bureaux. Le bruit de Hoth doit être produit dans une salle ayant de bas niveaux de bruit de fond (≤ 30 dBA). Il doit être mesuré dans le domaine acoustique. Si, pour certains types d'application particuliers, un autre type de bruit de fond est utilisé, il conviendra d'en spécifier la puissance spectrique et le niveau en dBA.</p> <p>NOTE 3 – Cette valeur correspond à un réglage autorisant une détectabilité maximale des distorsions. Pour certaines applications, des valeurs moins élevées ou plus élevées sont autorisées. Le réglage de niveau est mesuré par intégration rapide de la valeur maximale sur toute la séquence audio, selon la publication de la CEI [17]. Lorsque des haut-parleurs sont utilisés, le niveau sonore peut être ajusté en fonction de la distance de visualisation (voir la Bibliographie [11]).</p> <p>NOTE 4 – Cette valeur n'est applicable qu'à la présentation par haut-parleurs. Des durées de réverbération plus longues conduisent généralement à une diminution de la détectabilité des distorsions.</p> <p>NOTE 5 – Pour une hauteur d'écran donnée, il est probable que la distance de visualisation préférée par les sujets augmentera en même temps que la dégradation de la qualité visuelle. A ce propos, il y a lieu de prédéfinir la distance de visualisation préférée pour les essais de qualification. La distance de visualisation dépend généralement des applications.</p> <p>NOTE 6 – Pour les moniteurs de PC, la chromaticité de l'arrière-plan peut être adaptée à celle du moniteur.</p> <p>NOTE 7 – Les dimensions de la salle ne sont importantes que pour la présentation par haut-parleurs.</p>	

7.2 Système de traitement et de reproduction

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

- a) 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, d'écouter et de donner leurs impressions;
- b) 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'enregistrer le signal en format numérique afin de minimiser les dégradations pouvant être produites par le processus d'enregistrement. Les conditions d'essai peuvent être enregistrées sur le dispositif de stockage de composantes selon UIT-R BT.601. 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.

7.3 Sujets

Le nombre de sujets pouvant participer à un essai de visualisation (ainsi qu'à des essais ergonomiques portant sur des terminaux ou sur des services) est compris entre 6 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 très 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.

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.

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.

7.4 Instructions données aux sujets 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 5 pour la méthode ACR.

Tableau 5/P.911 – 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*, au 2.6.2C et dans [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.

ANNEXE A

Détails relatifs à la caractérisation des séquences d'essai

La caractérisation des séquences audiovisuelles sera fondée sur les caractéristiques de contenu aussi bien audio que vidéo. Les Tableaux A.1 et A.2 décrivent respectivement les catégories de la partie vidéo et de la partie audio. Ces tableaux ne visent pas l'exhaustivité.

Conformément à cette classification, une scène vidéophonique appartiendra à la catégorie A-I.

Tableau A.1/P.911 – Catégories de contenu vidéo

Catégorie	Description
A	Une personne, portrait (tête et épaules) principalement, détail et mouvement limités
B	Une personne avec des données graphiques et/ou plus de détails
C	Plusieurs personnes
D	Données graphique avec pointage
E	Objet et caméra à contenu cinétique élevé, au-delà de la limite habituellement constatée en télévidéoconférence

Tableau A.2/P.911 – Catégories de contenu audio

Catégorie	Description
I	Paroles/orateur unique
II	Paroles/orateurs multiples
III	Paroles + musique d’ambiance
IV	Musique/instrument 1
V	Musique/instruments multiples

ANNEXE B

Classes vidéo et audio et leurs attributs

B.1 Classes vidéo et leurs attributs

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/échantillon selon UIT-R BT.601.

Voir les Tableaux B.1 et B.2.

Tableau B.1/P.911 – Définition des classes vidéo

TV 0	Classe vidéo sans pertes selon la Recommandation UIT-R BT.601, 8 bit/échantillon, 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 ≥ 25 images/s pour les systèmes à 625 lignes et ≥ 30 images/s pour les systèmes à 525 lignes.
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 B.2/P.911 – 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	Rec. UIT-R BT.601	Max FR	(Note 2)	270
TV 1	Rec. UIT-R BT.601	Max FR	(Note 2)	18 à 50
TV 2	Rec. UIT-R BT.601	Max FR	(Note 2)	10 à 25
TV 3	Rec. UIT-R BT.601	Max FR occasionnel Répétition de trames	(Note 2)	1,5 à 8
MM 4a	Rec. UIT-R BT.601	~30 ou ~25 fps	Délai \approx 150 ms Variation \approx 50 ms	~1,5
MM 4b	CIF	~30 ou ~ 25 fps	Délai \approx 150 ms Variation \approx 50 ms	~0,7
MM 5a	CIF	10-30 fps	Délai \approx 1000 ms Variation \approx 500 ms	~0,2
MM 5b	\leq CIF	1-15 fps	Délai \approx 1000 ms Variation \approx 500 ms	~0,05
MM 6a	CIF-16CIF	Limite \rightarrow 0 fps	Pas de restrictions (Note 3)	<0,05, Limite \rightarrow 0 fps
MM 6b	CIF-16CIF	Limite \rightarrow 0 fps	Pas de restrictions (Note 3)	<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 [15]. Des variations de délai sont autorisées dans l'étendue indiquée mais ne devraient pas conduire à des effets de désynchronisation trop perceptibles.

NOTE 3 – Cette classe ne diffère de la précédente que par son débit audio.

B.2 Classes audio et leurs attributs

Dans la présente Recommandation, la qualité audio la plus élevée qui est prise en considération est l'audio codé MIC linéaire à 20 bits/échantillon de son ambiophonique dans la présentation 5,1 [16]. Pour faire référence aux classes audio, il convient d'indiquer la disposition des voies (par exemple monophonie, stéréophonie, 3 voies, 3/1 voies, etc.).

Voir les Tableaux B.3 et B.4.

Tableau B.3/P.911 – Définitions des classes audio

TV 0	MIC linéaire de qualité studio à 20 bits/échantillon, 48 kHz sans perte.
TV 1	Classe audio 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	Distribution primaire. Classe audio utilisée pour des modifications simples, des corrections peu nombreuses, des mixages 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 audio 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	Légers défauts audibles par rapport à TV 3 en modes parole et audio. Vidéoconférence de qualité moyenne. Largeur de bande audio normalement complète (de 20 à 20 000 Hz) mais des largeurs de bande ^(Note) réduites à la bande téléphonique élargie (de 50 à 7000 Hz) sont acceptables.
MM 5	Légers défauts audibles par rapport à une bande normale de référence (bande téléphonique de 300 à 3400 Hz) en modes parole et musique. Apparition possible de défauts perceptibles mais niveau de qualité utile pour des tâches précises, comme la vidéoconférence de qualité inférieure.
MM 6	Défauts audibles importants par rapport à une application de téléphonie en bande normale (300-3400 Hz). Les paroles restent toutefois intelligibles.
NOTE – La largeur de bande est définie comme étant la bande passante perceptible (c'est-à-dire celle du système mis à l'essai sans tenir compte des composantes inaudibles).	

Tableau B.4/P.911 – Attributs des classes audio

Classe audio	Fréquence d'échantillonnage (kHz)	Variation typique du délai dû à la latence ^(Note 1)	Débit nominal (kbit/s/voie) ^(Note 2)
TV 0	48,0	(Note 1)	960
TV 1	44,1 ou 48,0 ou 32,0	(Note 1)	250-500
TV 2	44,1 ou 48,0 ou 32,0	(Note 1)	120-300
TV 3	44,1 ou 48,0 ou 32,0	(Note 1)	50-150
MM 4a	44,1 ou 48,0 ou 32,0	Délai ≤ 150 ms Variation $\leq \pm 20$ ms	20-100
MM 4b	12-32	Délai ≤ 150 ms Variation $\leq \pm 20$ ms	10-50
MM 5a	12-20	Délai ≤ 400 ms ^(Note 3) Variation $\leq \pm 100$ ms	4-32

Tableau B.4/P.911 – Attributs des classes audio (fin)

Classe audio	Fréquence d'échantillonnage (kHz)	Variation typique du délai dû à la latence (Note 1)	Débit nominal (kbit/s/voie) (Note 2)
MM 5b	8-12	Délai ≈ 400 ms (Note 3) Variation $\approx \pm 200$ ms	2-16
MM 6a	8	Délai ≈ 400 ms (Note 4) Variation $\approx \pm 200$ ms	<8
MM 6b	8	Délai ≈ 4000 ms (Note 4) Variation $\approx \pm 2000$ ms	<8

NOTE 1 – Les systèmes de radiodiffusion possèdent tous une latence constante mais non nécessairement faible, ainsi qu'une variation de délai (temps de propagation) constante. Dans la plupart des applications de diffusion, 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 dans un seul sens soit de préférence inférieure à 150 ms. Des variations de délai sont autorisées dans l'étendue indiquée mais ne devraient pas conduire à des effets de désynchronisation trop perceptibles.

NOTE 2 – La voie d'amélioration à basse fréquence (voie 0,1 dans la présentation 5,1) ne nécessite qu'un débit légèrement plus élevé. Les corrélations mutuelles seront exploitées dans un signal audio à N voies, ce qui donnera un débit total nettement inférieur à N fois le débit de chaque voie.

NOTE 3 – Pour cette application, la synchronicité entre audio et vidéo peut ne pas donner la plus grande qualité conversationnelle de la liaison de communication.

NOTE 4 – Conférences audiographiques.

ANNEXE C

Analyse de la Relation entre qualité audio, qualité vidéo et qualité audiovisuelle

La présente Recommandation traite de la qualité audiovisuelle globale de séquences audiovisuelles perçues. Les relations entre qualité audio, qualité vidéo et qualité audiovisuelle sont à l'étude, mais certains résultats stables existent, dont le plus important est le mappage de la qualité audio perçue et de la qualité vidéo perçue (telles qu'obtenues à l'aide d'expériences subjectives audio seulement et vidéo seulement) avec la qualité audiovisuelle globale perçue. Quatre laboratoires différents ont trouvé des résultats de corrélation similaires, en dépit de conditions expérimentales très différentes (voir la Bibliographie [13]-[17]). On a fait une série d'expériences en procédant à une défocalisation pour les dégradations vidéo et le bruit multiplicatif pour les dégradations audio (voir la Bibliographie [13] et [14]). Une expérience analogue a été faite avec des séquences publicitaires télévisées en utilisant la qualité de diffusion de télévision comme qualité de référence audio et vidéo la plus élevée (voir la Bibliographie [15]). Des distorsions similaires ont été utilisées pour les dégradations vidéo (VIRIS 601, ANSI [8] – extensions V3, V6, V8) mais avec une limitation de bande pour les distorsions audio (allant de la largeur de bande complète, de 20 à 20 000 Hz, à la bande téléphonique normale de 300 à 3400 Hz). On a procédé à une évaluation de la qualité de téléconférence vidéo avec plusieurs types de distorsions de codage (voir la Bibliographie [16]). Trois codecs vidéo différents (deux privés et un H.261) ainsi que quatre codecs audio différents ont été utilisés. Une évaluation de la qualité de téléconférence vidéo a aussi été effectuée avec plusieurs types de distorsions de codage (voir la Bibliographie [17]), mais également sans codage (largeur de bande complète pour le son et signal PAL pour la vidéo). Les résultats de ces expériences, qui faisaient appel à l'échelle de qualité ACR à neuf ou à cinq degrés numériques comme indiqué au 6.1, montrent ce qui suit:

- 1) la vidéo est globalement mieux perçue que l'audio. Lorsque les variances de qualité audio et de qualité vidéo sont à peu près égales, la variance de qualité audiovisuelle globale est dominée par la variance de qualité vidéo. La corrélation entre qualité vidéo et qualité audiovisuelle globale est supérieure à la corrélation entre qualité audio et qualité audiovisuelle globale;
- 2) la qualité audiovisuelle globale perçue peut être prédite d'après la qualité audio perçue et d'après la qualité vidéo perçue sur la base d'expériences subjectives portant sur des séquences purement audio ou purement vidéo;
- 3) le mappage interlaboratoire le plus stable entre qualité audio pure ou qualité vidéo pure et qualité audiovisuelle globale a été constaté dans les termes suivants: $MOS_{av} = \alpha + \beta * MOS_a * MOS_v$. Ce mappage est fondé sur quatre ensembles d'expériences subjectives et la corrélation entre qualités audiovisuelles globales prédite et mesurée a varié de 0,93 à 0,99. La valeur de α variait entre 1,1 et 1,5 et la valeur de β entre 0,107 et 0,121 lorsque toutes les échelles étaient mappées à l'échelle de neuf degrés. Provisoirement, les valeurs recommandées sont fixées à 1,3 pour α et à 1,1 pour β .

La validité de ce modèle n'a été démontrée que pour des signaux audio et vidéo synchronisés et uniquement dans les limites des expériences où les dégradations de la qualité audio et vidéo avaient à peu près les mêmes effets perturbateurs. Il est possible de concevoir une expérience où la relation [point 3) ci-dessus] ne serait pas valable. Si les traitements portent sur une gamme de qualités audio avec des variations d'intelligibilité et sur une gamme restreinte de qualités vidéo, la qualité audio devrait être prépondérante par rapport à la qualité audio et vidéo.

APPENDICE I

Séquences audiovisuelles d'essai

La sélection de séquences d'essai appropriées est une phase essentielle dans la planification des évaluations subjectives. Lorsque les résultats d'essais effectués avec différents groupes d'observateurs ou dans différents laboratoires doivent être mis en corrélation, il importe de pouvoir disposer d'un ensemble commun de séquences d'essai.

Un premier ensemble de telles séquences est décrit dans le Tableau I.1. Dans ce tableau, les informations suivantes sont données pour chaque séquence:

- la catégorie (définie dans les Tableaux A.1 et A.2);
- une brève description de la scène;
- le format de la source (soit 625 ou 525 lignes en format UIT-R BT.601 ou BetacamSP);
- la fréquence d'échantillonnage du signal audio;
- les valeurs des informations spatiales et temporelles (définies respectivement en 5.3.1/P.910 et 5.3.2/P.910).

Toutes les séquences énumérées dans le Tableau I.1 sont dans le domaine public et peuvent être utilisées librement pour les évaluations et les démonstrations.

**Tableau I.1/P.911 – Séquences d'essai pour évaluation de la qualité audiovisuelle
dans les applications multimédias
(à déterminer)**

Séquence	Catégorie (A + V)	Description	Format de source	Langue	Fréquence d'échan- tillonnage	SI	TI
	A-I						
	B-III						
	D-I						
	D-II						
	etc.						

APPENDICE II

Instructions pour les essais audiovisuels

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 implémentation spécifique.

II.1 Méthode ACR

Bonjour et merci de votre participation.

Dans cette expérience, de brèves séquences audiovisuelles vous seront présentées. Chaque fois que vous verrez et auditionnez une séquence, vous devrez en juger la qualité en cochant l'un des cinq niveaux de l'échelle suivante:

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

Votre évaluation doit refléter votre opinion quant à la qualité globale, **audio et vidéo**, combinée.

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

II.2 Méthode DCR

Bonjour et merci de votre participation.

Dans cette expérience, de brèves séquences audiovisuelles vous seront présentées. Chaque séquence sera présentée deux fois en succession rapide: la première séquence est la référence, 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, quant à la différence entre la seconde et la première séquence:

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

Votre évaluation doit refléter votre opinion quant à la qualité globale, **audio et vidéo**, combinée.

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

II.3 Méthode PC

Bonjour et merci de votre participation.

Dans cette expérience, de brèves séquences audiovisuelles vous seront présentées. Chaque séquence sera présentée deux fois en succession rapide, chaque fois par un système 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
---	---

Votre évaluation doit refléter votre opinion quant à la qualité globale, **audio et vidéo**, combinée.

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

APPENDICE III

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] Gram-Field Catalogue Number 13-1240.
- [4] Pseudo Isochromatic Plates, engraved and printed by – *The Beck Engraving Co., Inc.*, Philadelphie et New York, Etats-Unis.
- [5] KIRK (R.E.): Experimental Design – Procedures for the Behavioural Sciences, 2nd Editions, *Brooks/Cole Publishing Co.*, Californie, 1982.
- [6] VIRTANEN (M.T.), GLEISS (N.) et GOLDSTEIN (M.): On the use of Evaluative Category Scales in Telecommunications, HFT 1995, *Human Factors in Telecommunication Conference*, Melbourne, 1995.
- [7] GUILFORD (P.): Psychometric methods, MCGRAW-HILL, New York, 1954.
- [8] ANSI T1A1.5 contribution 96-109: VIRIS for ITU-R 601 digital images, mai 1996.

- [9] HOTH (D.F.): Room noise spectra at subscribers' telephone locations, *J.A.S.A.*, Volume 12, p. 99-504, avril 1941.
- [10] BECH (S.): Calibration of Relative Level Differences of a Domestic Multichannel Sound Reproduction System, *Journal of Audio Engineering Society*, avril 1998.
- [11] UIT-T Delayed Contribution D.40 (1990), *Some speech quality aspects to be considered in multimedia services*, NTT.
- [12] *Handbook on Telephony*, UIT Genève, 1992.
- [13] UIT-T COM 12-20 (décembre 1993), *Experimental combined audio/video subjective test method*, Bellcore.
- [14] UIT-T COM 12-37 (septembre 1994), *Extension of combined audio/video quality model*, Bellcore.
- [15] UIT-T COM 12-19 (février 1998), *Relations Between Audio, Video and Audiovisual Quality*, KPN Research, Netherlands.
- [16] UIT-T COM 12-64 (novembre 1998), *Results of an audiovisual desktop video teleconferencing subjective experiment*, USA.
- [17] UIT-T COM 12-61 (novembre 1998), *Study of the influence of experimental context on the relationship between audio, video and audiovisual subjective quality*, France Telecom/CNET.

SERIES DES RECOMMANDATIONS UIT-T

Série A	Organisation du travail de l'UIT-T
Série B	Moyens d'expression: définitions, symboles, classification
Série C	Statistiques générales des télécommunications
Série D	Principes généraux de tarification
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	Transmission des signaux radiophoniques, télévisuels et autres signaux multimédias
Série K	Protection contre les perturbations
Série L	Construction, installation et protection des câbles et autres éléments des installations extérieures
Série M	RGT et maintenance des réseaux: systèmes de transmission, de télégraphie, de télécopie, circuits téléphoniques et circuits loués internationaux
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
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 pour données et communication entre systèmes ouverts
Série Y	Infrastructure mondiale de l'information
Série Z	Langages de programmation