

Union internationale des télécommunications

UIT-R

Secteur des Radiocommunications de l'UIT

Recommandation UIT-R BT.2095-1
(06/2017)

**Evaluation subjective de la qualité vidéo
au moyen du protocole d'observation
par des spécialistes**

Série BT
Service de radiodiffusion télévisuelle



Union
internationale des
télécommunications

Avant-propos

Le rôle du Secteur des radiocommunications est d'assurer l'utilisation rationnelle, équitable, efficace et économique du spectre radioélectrique par tous les services de radiocommunication, y compris les services par satellite, et de procéder à des études pour toutes les gammes de fréquences, à partir desquelles les Recommandations seront élaborées et adoptées.

Les fonctions réglementaires et politiques du Secteur des radiocommunications sont remplies par les Conférences mondiales et régionales des radiocommunications et par les Assemblées des radiocommunications assistées par les Commissions d'études.

Politique en matière de droits de propriété intellectuelle (IPR)

La politique de l'UIT-R en matière de droits de propriété intellectuelle est décrite dans la «Politique commune de l'UIT-T, l'UIT-R, l'ISO et la CEI en matière de brevets», dont il est question dans l'Annexe 1 de la Résolution UIT-R 1. Les formulaires que les titulaires de brevets doivent utiliser pour soumettre les déclarations de brevet et d'octroi de licence sont accessibles à l'adresse <http://www.itu.int/ITU-R/go/patents/fr>, où l'on trouvera également les Lignes directrices pour la mise en oeuvre de la politique commune en matière de brevets de l'UIT-T, l'UIT-R, l'ISO et la CEI et la base de données en matière de brevets de l'UIT-R.

Séries des Recommandations UIT-R

(Egalement disponible en ligne: <http://www.itu.int/publ/R-REC/fr>)

Séries	Titre
BO	Diffusion par satellite
BR	Enregistrement pour la production, l'archivage et la diffusion; films pour la télévision
BS	Service de radiodiffusion sonore
BT	Service de radiodiffusion télévisuelle
F	Service fixe
M	Services mobile, de radiorepérage et d'amateur y compris les services par satellite associés
P	Propagation des ondes radioélectriques
RA	Radio astronomie
RS	Systèmes de télédétection
S	Service fixe par satellite
SA	Applications spatiales et météorologie
SF	Partage des fréquences et coordination entre les systèmes du service fixe par satellite et du service fixe
SM	Gestion du spectre
SNG	Reportage d'actualités par satellite
TF	Emissions de fréquences étalon et de signaux horaires
V	Vocabulaire et sujets associés

Note: Cette Recommandation UIT-R a été approuvée en anglais aux termes de la procédure détaillée dans la Résolution UIT-R 1.

Publication électronique
Genève, 2018

© UIT 2018

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.

RECOMMANDATION UIT-R BT.2095-1

Evaluation subjective de la qualité vidéo au moyen du protocole d'observation par des spécialistes

(2016-2017)

Domaine d'application

La présente Recommandation décrit la méthode d'évaluation subjective de la qualité vidéo d'images animées au moyen du protocole d'observation par des spécialistes, avec la participation d'un nombre réduit d'observateurs, tous choisis parmi les spécialistes du domaine correspondant du traitement vidéo.

Mots clés

Télévision, qualité vidéo, évaluation subjective, observation par des spécialistes

L'Assemblée des radiocommunications de l'UIT,

considérant

- a) que l'efficacité et la qualité visuelle des technologies de codage de sources pour les applications de télévision numérique ne cessent de s'améliorer;
- b) que cette évolution permanente des technologies de codage vidéo se traduit par une demande en hausse constante concernant les méthodes d'évaluation des performances techniques et visuelles;
- c) que l'efficacité de compression et les performances visuelles des nouvelles technologies de codage de sources vidéo appellent des méthodes nouvelles et plus efficaces d'évaluation et de notation de la qualité visuelle;
- d) que les méthodes d'évaluation définies dans les Recommandations UIT-R en vigueur sont très exigeantes en termes de temps et de ressources humaines et ne tiennent souvent pas compte de l'évolution technique des écrans et de l'installation chez l'utilisateur final;
- e) que, récemment, de nouvelles approches dans le domaine des protocoles d'observation par des spécialistes se sont avérées plus efficaces et plus efficaces, en termes de temps et de coût global, que les méthodes faisant appel à des observateurs non spécialistes;
- f) que, si l'on ne peut pas considérer que les résultats obtenus avec un protocole d'observation par des spécialistes remplacent ceux obtenus avec un protocole d'évaluation subjective formelle, on peut toutefois considérer qu'ils donnent une première indication précieuse de la qualité de fonctionnement des systèmes testés;
- g) que l'évolution technologique de plus en plus soutenue dans le domaine des écrans plats modifie considérablement les conditions d'observation normalement appliquées par les spécialistes;
- h) que l'ISO/CEI a déjà utilisé avec succès de nouveaux protocoles fondés sur l'observation par des spécialistes pour évaluer de nouvelles technologies de codage de sources vidéo,

recommande

1 en ce qui concerne l'évaluation des nouvelles technologies de codage vidéo numérique, d'envisager d'utiliser le protocole d'observation par des spécialistes décrit dans l'Annexe 1;

2 de mettre en œuvre le protocole d'observation par des spécialistes en utilisant des écrans plats professionnels et le montage de laboratoire décrit dans l'Annexe 1.

Note 1 – L'Annexe 2 (informative) présente les résultats d'expériences d'évaluation subjective réalisées conformément au protocole d'observation par des spécialistes, tel que mentionné au point h) du *considérant*.

Annexe 1

Protocole d'observation par des spécialistes pour l'évaluation subjective de la qualité des séquences vidéo

1 Montage de laboratoire

1.1 Choix de l'écran et montage

L'écran utilisé devrait être un écran plat ayant les caractéristiques types d'applications professionnelles (par exemple, studio de radiodiffusion ou car-régie); sa dimension en diagonale pourra être comprise entre 22 pouces (minimum) et 40 pouces (recommandée), mais elle pourra aller jusqu'à 50 pouces ou plus pour l'évaluation de systèmes d'images avec une résolution de TVHD ou plus élevée.

Il est possible d'utiliser une partie réduite de la zone active d'affichage de l'écran; dans ce cas, la zone située autour de la partie active de l'écran devrait être mise à «gris moyen». Dans cette configuration, il ne devrait pas être accepté que la résolution de l'écran soit réglée sur une résolution autre que celle d'origine.

L'écran devrait permettre un montage et un étalonnage de la luminance et des couleurs appropriés au moyen d'un luxmètre professionnel. L'étalonnage de l'écran devrait être conforme aux paramètres définis dans la Recommandation applicable pour le test effectué.

1.2 Distance d'observation

La distance d'observation devrait être choisie en fonction de la résolution de l'écran et de la hauteur de la partie active de l'écran, conformément à la distance d'observation nominale définie dans la Recommandation UIT-R BT.2022 ou être plus courte, selon les exigences concernant les mauvaises conditions d'observation.

1.3 Conditions d'observation

Une expérience menée selon le protocole d'observation par des spécialistes (EVP) ne devrait pas nécessairement se dérouler dans un laboratoire de test, mais il est important que l'emplacement choisi soit protégé des perturbations sonores et/ou visuelles (on peut par exemple utiliser un bureau ou une salle de réunion au calme).

Il convient d'éliminer tout reflet sur l'écran provenant d'une source de lumière directe ou indirecte; l'éclairage ambiant devrait être faible, au niveau minimum suffisant pour remplir les feuilles de notation (le cas échéant).

Le nombre de spécialistes assis devant l'écran pourra varier en fonction de la taille de l'écran, l'objectif étant de garantir que le rendu d'image et l'exposition aux stimuli soient identiques pour tous les observateurs.

2 Observateurs

Les observateurs participant à une expérience EVP devraient être des spécialistes du domaine à l'étude.

Les observateurs ne devraient pas nécessairement être sélectionnés pour leur acuité visuelle ou leur perception des couleurs, mais devraient être choisis parmi des personnes qualifiées.

Il devrait y avoir au minimum neuf observateurs différents.

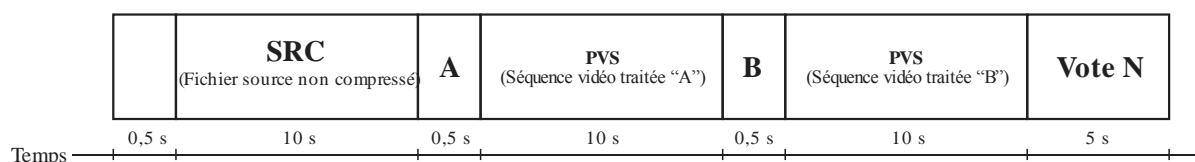
Pour atteindre le nombre minimum d'observateurs, il sera possible d'effectuer la même expérience à plusieurs reprises au même endroit ou dans plusieurs endroits. Les notes obtenues dans les différents lieux utilisés pour une session d'observation par des spécialistes pourront être traitées ensemble sur le plan statistique.

3 Cellule test de base

La séquence qui sera présentée aux spécialistes devrait être organisée de manière à créer une cellule test de base (BTC) pour chaque ensemble de conditions de codage à évaluer (voir la Fig. 1).

Les extraits de la séquence source de référence (SRC) et des séquences vidéo traitées (PVS) à examiner dans une cellule BTC devraient toujours être tirés de la même séquence vidéo, afin que les spécialistes soient en mesure de repérer toute amélioration de la qualité visuelle offerte par les algorithmes de compression testés.

FIGURE 1
Organisation d'une cellule test de base pour le protocole d'observation
par des spécialistes



BT.2095-01

La cellule BTC devrait être organisée comme suit:

- écran mis à gris moyen (valeur moyenne sur l'échelle de luminance) pendant 0,5 seconde;
- présentation de 10 secondes de l'extrait vidéo de référence non compressé;
- affichage pendant 0,5 seconde du message «A» (première vidéo à évaluer) sur fond gris moyen;
- présentation de 10 secondes d'une version altérée de l'extrait vidéo;
- affichage pendant 0,5 seconde du message «B» (deuxième vidéo à évaluer) sur fond gris moyen;
- présentation de 10 secondes d'une version altérée de l'extrait vidéo;
- affichage pendant 5 secondes d'un message demandant aux observateurs de donner leur avis.

Le message «Vote» devrait être suivi d'un numéro facilitant le report sur la feuille de notation.

3.1 Feuille de notation et échelle de notation

Comme le montre la Fig. 1, la présentation des extraits vidéo devraient être organisée de telle sorte que la séquence de référence non altérée (séquence SRC) soit diffusée en premier et suivie des deux séquences vidéo altérées (séquence PVS). L'ordre de présentation des séquences PVS devrait être modifié de manière aléatoire pour chaque cellule BTC et les observateurs ne devraient pas connaître l'ordre de présentation.

FIGURE 2

**Exemple de feuille de notation pour une session d'observation
par des spécialistes avec 24 cellules BTC**

Session 1

Vote 1	Vote 2	Vote 3	Vote 4	Vote 5
A B <input type="checkbox"/> <input type="checkbox"/>	A B <input type="checkbox"/> <input type="checkbox"/>	A B <input type="checkbox"/> <input type="checkbox"/>	A B <input type="checkbox"/> <input type="checkbox"/>	A B <input type="checkbox"/> <input type="checkbox"/>
Vote 6	Vote 7	Vote 8	Vote 9	Vote 10
A B <input type="checkbox"/> <input type="checkbox"/>	A B <input type="checkbox"/> <input type="checkbox"/>	A B <input type="checkbox"/> <input type="checkbox"/>	A B <input type="checkbox"/> <input type="checkbox"/>	A B <input type="checkbox"/> <input type="checkbox"/>
Vote 11	Vote 12	Vote 13	Vote 14	Vote 15
A B <input type="checkbox"/> <input type="checkbox"/>	A B <input type="checkbox"/> <input type="checkbox"/>	A B <input type="checkbox"/> <input type="checkbox"/>	A B <input type="checkbox"/> <input type="checkbox"/>	A B <input type="checkbox"/> <input type="checkbox"/>
Vote 16	Vote 17	Vote 18	Vote 19	Vote 20
A B <input type="checkbox"/> <input type="checkbox"/>	A B <input type="checkbox"/> <input type="checkbox"/>	A B <input type="checkbox"/> <input type="checkbox"/>	A B <input type="checkbox"/> <input type="checkbox"/>	A B <input type="checkbox"/> <input type="checkbox"/>
Vote 21	Vote 22	Vote 23	Vote 24	
A B <input type="checkbox"/> <input type="checkbox"/>	A B <input type="checkbox"/> <input type="checkbox"/>	A B <input type="checkbox"/> <input type="checkbox"/>	A B <input type="checkbox"/> <input type="checkbox"/>	

Siège	Sujet
1 2 3 <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>

BT.2095-02

Une échelle numérique comprenant 11 niveaux allant de 10 (dégradation imperceptible) à 0 (dégradation très gênante) est utilisée.

Le Tableau 1 donne des indications sur la signification des 11 niveaux de l'échelle numérique.

TABLEAU 1

Signification des 11 niveaux de l'échelle numérique

Note	Dégradation	
10	Imperceptible	
9	Légèrement perceptible	A un endroit
8		Partout
7	Perceptible	A un endroit
6		Partout
5	Clairement perceptible	A un endroit
4		Partout
3	Gênante	A un endroit
2		Partout
1	Extrêmement gênante	A un endroit
0		Partout

Il est demandé aux observateurs de remplir un questionnaire comprenant deux cases («A» et «B») pour chaque cellule BTC, en indiquant dans chacune de ces deux cases une note choisie selon l'échelle numérique à 11 niveaux.

La Figure 2 donne un exemple de feuille de notation pour une session comprenant 24 cellules BTC.

Pour chaque cellule BTC, les observateurs remplissent la case identifiée par la lettre **A** (pour noter l'extrait vidéo apparaissant en premier) et la case identifiée par la lettre **B** (pour noter l'extrait vidéo apparaissant en deuxième).

La présentation de l'extrait vidéo original non altéré permet aux spécialistes d'évaluer plus facilement les éventuelles dégradations.

La signification des 11 niveaux de l'échelle numérique devrait être expliquée en détail lors de «sessions de formation», comme indiqué ci-après.

3.2 Conception du test et création d'une session

L'ordre de présentation des cellules BTC devrait être fixé de manière aléatoire par le concepteur du test, de telle sorte que le même extrait vidéo ou le même extrait altéré ne soit pas présenté deux fois de suite.

Toutes les sessions d'observation devraient commencer par une «phase de stabilisation» comprenant la «meilleure» cellule BTC, la pire cellule BTC et deux cellules BTC de «qualité moyenne» figurant dans chaque session de test. Les observateurs auront ainsi immédiatement un aperçu des différentes qualités, dès le début de la session.

Si la session d'observation dure plus de 20 minutes, le concepteur du test devrait la scinder en deux (ou plus) sessions d'observation séparées, ne dépassant pas 20 minutes chacune. Dans ce cas, il devrait y avoir une «phase de stabilisation» avant chaque session.

3.3 Formation

Même si cette procédure s'adresse à des spécialistes, il est préférable d'organiser une courte session d'observation de formation (5 à 6 cellules BTC) avant chaque expérience.

Les séquences vidéo utilisées pour la session de formation pourront être les mêmes que celles utilisées pour les véritables sessions, mais l'ordre de présentation devrait être différent.

Il convient de former les observateurs à l'utilisation de l'échelle à 11 niveaux en leur demandant de regarder attentivement les extraits vidéo diffusés immédiatement après les messages «A» et «B» sur l'écran et de déterminer s'ils peuvent voir une différence par rapport à l'extrait vidéo diffusé en premier (la séquence SRC).

4 Collecte et traitement des données

Les notes devraient être collectées à la fin de chaque session et saisies dans un tableur pour calculer les valeurs MOYENNES.

Il est souhaitable d'effectuer une sélection a posteriori des observateurs moyennant une corrélation linéaire de Pearson.

Cette fonction de corrélation devrait être appliquée compte tenu de toutes les notes de chaque sujet par rapport aux notes moyennes d'opinion (MOS); un seuil pourra être fixé afin de définir pour chaque observateur s'il est «acceptable» ou «refusé» (la Recommandation UIT-T P.913 propose d'utiliser une valeur seuil de «refus» égale à 0,75).

5 Conditions d'utilisation des résultats obtenus avec le protocole d'observation par des spécialistes

Le protocole d'observation par des spécialistes (EVP) pourra être utilisé lorsque le temps et les ressources disponibles ne permettent pas de mener à bien une expérience d'évaluation subjective formelle.

Une évaluation EVP prend moins de temps qu'une évaluation subjective formelle et peut être menée dans un environnement «informel», pour autant que cet environnement soit protégé de toute perturbation visuelle et sonore extérieure.

Les seules conditions à respecter impérativement concernent l'éclairage ambiant et les conditions d'observation (écran, angle et distance d'observation) comme indiqué ci-dessus.

6 Limites de l'utilisation des résultats des tests EVP

Bien qu'il soit établi que le protocole EVP peut fournir des résultats acceptables avec seulement neuf observateurs, on ne peut pas considérer que les notes moyennes d'opinion obtenues avec une expérience EVP remplacent les résultats que l'on peut obtenir avec une expérience d'évaluation subjective formelle.

Les notes moyennes d'opinion obtenues avec le protocole EVP peuvent être utilisées pour avoir une première indication du niveau de dégradation.

Les notes moyennes d'opinion obtenues avec le protocole EVP peuvent être utilisées pour faire un premier classement des mécanismes de traitement vidéo testés.

Lorsqu'on le juge pratique ou nécessaire, une expérience EVP peut se dérouler à plusieurs endroits en parallèle, pour autant que les conditions et la distance d'observation ainsi que la conception du test soient identiques.

Si le nombre d'observateurs spécialistes participant à la même expérience EVP, qu'elle se déroule dans un seul ou dans plusieurs endroits, est égal ou supérieur à 15, les données subjectives brutes peuvent être traitées pour obtenir une note moyenne d'opinion, l'écart type et l'intervalle de confiance, qui pourront aider à classer de manière plus précise les cas testés. Dans ce dernier cas, une analyse statistique déductive plus précise pourra être réalisée, par exemple un test T de Student.

Annexe 2 (informative)

Application du protocole d'observation par des spécialistes et incidences liées à la présence d'un grand nombre d'observateurs spécialistes

La présente Annexe informative fournit des informations sur les résultats de deux sessions différentes d'évaluation subjective de séquences vidéo HD et UHD codées, organisées conformément au protocole EVP lors de la 117ème réunion du Groupe d'expert pour les images animées (MPEG), en application des dispositions de la Recommandation UIT-R BT.2095, afin de classer de manière rapide et fiable deux méthodes différentes de codage de la source.

Etant donné qu'un grand nombre de spécialistes ont participé à la 117ème réunion du Groupe MPEG, le nombre d'observateurs participant aux deux sessions EVP était largement supérieur au nombre de

neuf observateurs préconisé dans la Recommandation UIT-R BT.2095; 30 spécialistes ont participé à la session de test EVP HD et 32 spécialistes ont participé à la session de test EVP UHD.

Grâce à la large participation d'observateurs spécialistes, on a pu analyser les notes moyennes d'opinion, afin de vérifier le niveau de fiabilité lié à l'utilisation de la Recommandation UIT-R BT.2095 lorsqu'il s'agit de classer des séquences vidéo codées.

Dans le cadre de cette évaluation, on considère quatre groupes d'observateurs (composés respectivement de 9, 12, 15 et 18 observateurs) et on effectue une comparaison entre les valeurs des notes moyennes d'opinion obtenues à partir du groupe de neuf spécialistes et celles obtenues à partir des groupes de 12, 15 et 18 observateurs.

L'objectif consistait à comparer le classement obtenu avec neuf spécialistes (protocole EVP) et les classements obtenus avec 12, 15 et 18 spécialistes (expérience d'évaluation subjective formelle).

Il ressort de la Figure 3 (expérience portant sur un contenu UHD) et de la Figure 4 (expérience portant sur un contenu HD) que les résultats des classements obtenus dans les quatre cas considérés sont très similaires.

Si l'on prend comme référence les résultats obtenus avec le groupe de 18 observateurs, on peut tracer les graphiques des Figures 3 et 4 en classant les séquences évaluées selon les valeurs des notes moyennes d'opinion obtenues avec le groupe de 18 observateurs (courbe continue rouge).

Les autres courbes des graphiques représentent les résultats obtenus avec le groupe de neuf observateurs (courbe en pointillés rouge), de 12 observateurs (courbe en traits discontinus bleue) et de 15 observateurs (courbe continue verte).

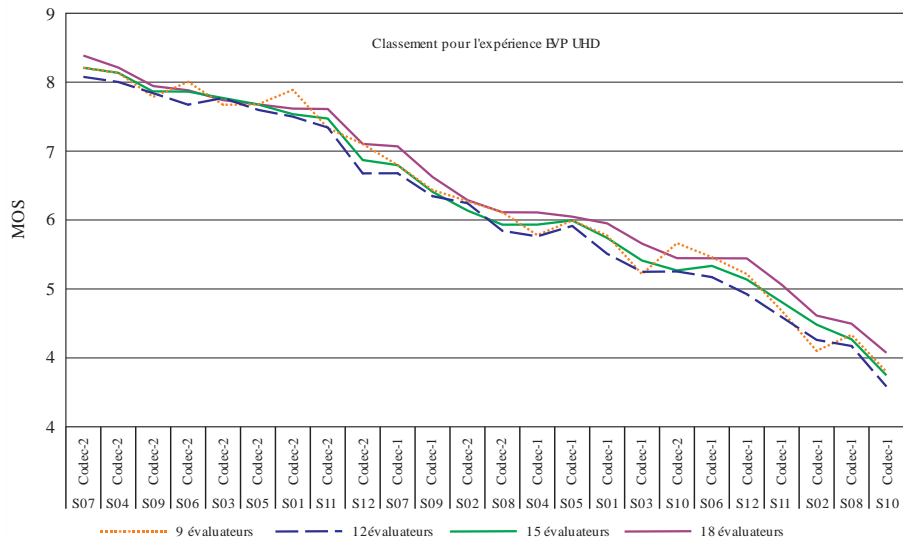
En observant les résultats représentés dans les Figures 3 et 4, il convient de noter que:

- les courbes représentant les résultats obtenus avec les groupes de 15 et de 18 observateurs présentent une pente homogène, concernant aussi bien les notes moyennes d'opinion pour les séquences de bonne qualité que celles pour les séquences de mauvaise qualité;
- les courbes représentant les résultats obtenus avec les groupes de neuf et de 12 observateurs présentent certaines «inversions» dans le classement par rapport à la courbe correspondant au groupe de 18 observateurs, bien que ces variations aient une portée limitée.

En conclusion, les expériences EVP décrites ci-avant témoignent de la très bonne qualité des résultats obtenus au moyen du protocole EVP, ce qui vient confirmer les explications données dans la Recommandation UIT-R BT.2095, à savoir que, même si l'on ne peut pas considérer qu'il remplace en tout point une expérience d'évaluation subjective formelle, le protocole EVP pourrait être considéré comme une procédure d'évaluation stable, qui fournit des résultats très proches de ceux obtenus lorsqu'un nombre beaucoup plus important d'observateurs est disponible et qu'une évaluation subjective formelle est réalisée.

FIGURE 3

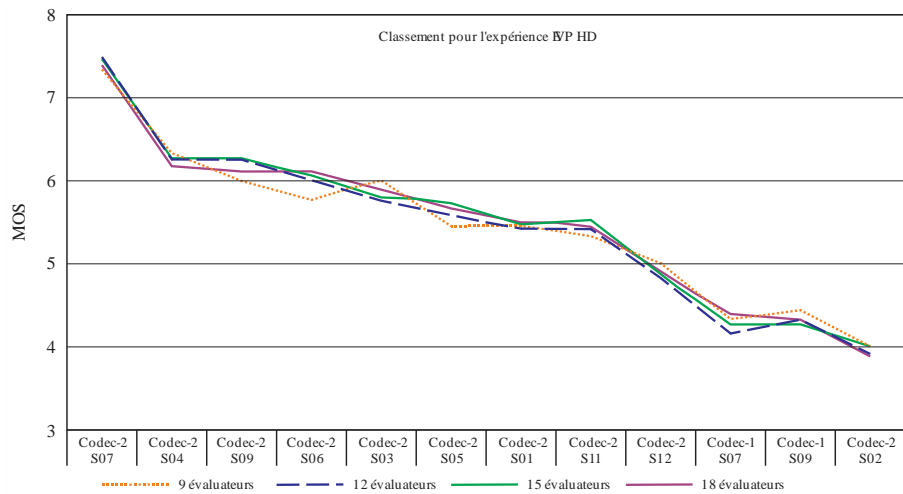
Classement pour l'expérience UHD en fonction du nombre d'observateurs



BT.2095-03

FIGURE 4

Classement pour l'expérience HD en fonction du nombre d'observateurs



BT.2095-04