

Union internationale des télécommunications

**UIT-R**

Secteur des Radiocommunications de l'UIT

**Recommandation UIT-R BT.2021-1**  
(02-2015)

**Méthodes d'évaluation subjective des  
systèmes de TV3D stéréoscopique**

**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
<b>BO</b>	Diffusion par satellite
<b>BR</b>	Enregistrement pour la production, l'archivage et la diffusion; films pour la télévision
<b>BS</b>	Service de radiodiffusion sonore
<b>BT</b>	<b>Service de radiodiffusion télévisuelle</b>
<b>F</b>	Service fixe
<b>M</b>	Services mobile, de radiorepérage et d'amateur y compris les services par satellite associés
<b>P</b>	Propagation des ondes radioélectriques
<b>RA</b>	Radio astronomie
<b>RS</b>	Systèmes de télédétection
<b>S</b>	Service fixe par satellite
<b>SA</b>	Applications spatiales et météorologie
<b>SF</b>	Partage des fréquences et coordination entre les systèmes du service fixe par satellite et du service fixe
<b>SM</b>	Gestion du spectre
<b>SNG</b>	Reportage d'actualités par satellite
<b>TF</b>	Emissions de fréquences étalon et de signaux horaires
<b>V</b>	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, 2016

© UIT 2016

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.2021-1

**Méthodes d'évaluation subjective des systèmes de TV3D stéréoscopique**

(2012-2015)

**Domaine d'application**

La présente Recommandation définit des méthodes d'évaluation des systèmes de TV3D stéréoscopique et porte en particulier sur les méthodes générales d'évaluation, les échelles d'évaluation et les conditions d'observation.

L'Assemblée des radiocommunications de l'UIT,

*considérant*

- a) qu'une grande quantité d'informations a été recueillie sur les méthodes utilisées dans divers laboratoires pour l'évaluation des caractéristiques essentielles de qualité de fonctionnement des systèmes de TV3D;
- b) que l'examen de ces méthodes montre qu'il existe un large accord entre les différents laboratoires sur un certain nombre d'aspects de ces évaluations;
- c) qu'il est important d'adopter des méthodes normalisées pour l'échange d'informations entre les divers laboratoires;
- d) que l'introduction de services de TV3D pourrait nécessiter le développement de nouveaux formats d'image et de nouvelles techniques de traitement et de transmission des images, dont la performance devra être évaluée par des méthodes subjectives,

*recommande*

**1** que les méthodes générales d'évaluation, les échelles d'évaluation et les conditions d'observation pour l'évaluation de la qualité des images de TV3D stéréoscopique décrites dans l'Annexe 1 ci-après soient utilisées pour les expériences en laboratoire ainsi que pour les évaluations en exploitation chaque fois que possible.

**Annexe 1****1 Dimensions (de perception) d'évaluation**

La TV3D stéréoscopique exploite les caractéristiques du système visuel binoculaire humain en recréant les conditions permettant de percevoir la profondeur relative des objets dans la scène visuelle. Actuellement, l'imagerie stéréoscopique nécessite principalement de prendre au moins deux vues de la même scène depuis deux caméras alignées horizontalement. Les images des objets de la scène auront des positions relatives différentes dans la vue de gauche et dans celle de droite. Cette différence de positions relatives dans les deux vues, généralement appelée disparité d'image (ou parallaxe), est habituellement exprimée par un nombre de pixels, une distance physique (par exemple en mm) ou une mesure relative (par exemple en pourcentage de la largeur d'écran). Il convient de distinguer la disparité d'image de la disparité angulaire (rétinienne). De fait, les mêmes informations de disparité d'image produisent des disparités angulaires (rétiniennes) différentes pour des distances d'observation différentes. L'amplitude et le sens de la profondeur perçue dépendent de l'amplitude et du sens des disparités rétiniennes provoquées par l'image stéréoscopique.

Les facteurs généralement utilisés pour l'évaluation des images de télévision monoscopiques, tels que la résolution, la restitution des couleurs, la restitution du mouvement, la qualité générale, la netteté, etc. pourraient aussi être utilisés pour les systèmes de télévision stéréoscopique. A cela s'ajouteraient de nombreux facteurs propres aux systèmes de télévision stéréoscopique, par exemple la résolution en profondeur, qui est la résolution spatiale dans le sens de la profondeur, le mouvement en profondeur, facteur qui indique si le mouvement dans le sens de la profondeur est restitué sans à-coups, et les distorsions spatiales, dont deux exemples sont bien connus: l'*effet spectacle de marionnettes* – lorsque des objets sont perçus comme étant anormalement grands ou petits – et l'*effet carton* – lorsque des objets sont perçus stéréoscopiquement mais paraissent anormalement minces.

Il est possible d'identifier trois dimensions de perception principales qui, ensemble, ont une incidence sur la qualité offerte par un système stéréoscopique: la *qualité de l'image*, la *qualité de la profondeur* et le *confort visuel*. D'après certains chercheurs, on pourrait aussi mesurer l'incidence psychologique des technologies d'imagerie stéréoscopique en utilisant des concepts plus généraux tels que le *naturel* et l'*impression de présence*.

### **Dimensions de perception principales**

La *qualité de l'image* désigne la qualité perçue de l'image fournie par le système. Principale composante de la qualité de fonctionnement d'un système vidéo, elle dépend essentiellement des paramètres techniques et des erreurs introduites par, entre autres, les processus de codage et/ou de transmission.

La *qualité de la profondeur* désigne la capacité du système à renforcer l'impression de profondeur. La présence d'indices monoculaires (perspective linéaire, flou, gradients, etc.) permet de percevoir une certaine profondeur même dans des images 2D standards. Toutefois, les images 3D stéréoscopiques contiennent aussi des informations de disparité qui offrent des informations supplémentaires de profondeur et, par là même, renforcent l'impression de profondeur par rapport aux images 2D.

Le *confort visuel* (l'*inconfort visuel*) désigne la sensation subjective de confort (d'inconfort) associée à l'observation d'images stéréoscopiques. Des images stéréoscopiques mal prises ou mal affichées peuvent créer beaucoup d'inconfort.

### **Autres dimensions de perception**

Le *naturel* désigne la perception de l'image stéréoscopique comme étant une représentation fidèle de la réalité (perception de réalisme). L'image stéréoscopique peut présenter différents types de distorsions qui la rendent moins naturelle. Par exemple, certains objets stéréoscopiques sont parfois perçus comme étant anormalement grands ou petits (effet spectacle de marionnettes), ou ils paraissent anormalement minces (effet carton).

L'*impression de présence* désigne l'impression subjective d'être à un endroit différent de l'endroit où l'on se trouve.

La présente Recommandation donne des informations sur les méthodes et les procédures d'évaluation des trois dimensions principales – qualité de l'image, qualité de la profondeur et confort visuel – décrites ci-dessus. Les méthodes d'évaluation du naturel et de l'impression de présence n'y sont pas traitées, mais il est prévu de les aborder dans une future version.

## **2 Méthodes subjectives**

La Recommandation UIT-R BT.500 décrit de nombreuses méthodes d'évaluation de la qualité d'image. Dans toutes les méthodes, un ensemble de séquences vidéo, traitées par les systèmes (par exemple un algorithme avec différents paramètres, une technologie de codage à différents débits binaires, différents scénarios de transmission, etc.) à l'étude, est présenté à un groupe d'observateurs

dans une série d'expériences d'évaluation. Dans chaque expérience, il est demandé aux observateurs d'évaluer une caractéristique particulière (par exemple la qualité de l'image) de la ou des séquences vidéo sur une échelle donnée. Les principales différences entre les diverses méthodes résident dans le mode de présentation, autrement dit la manière dont les séquences vidéo sont présentées aux observateurs, et dans l'échelle utilisée par les observateurs pour noter ces séquences.

Les images de test sont des images stéréo binoculaires choisies sur la base des éléments décrits au § 4. Les observateurs évaluent les trois éléments suivants:

- la qualité de l'image: l'effet produit sur la résolution d'images 3D stéréoscopiques par un système opérant entre les images de test et le moniteur utilisé pour afficher les images à évaluer;
- la qualité de la profondeur: l'effet produit sur la perception de la profondeur d'images 3D stéréoscopiques par un système opérant entre les images de test et le moniteur utilisé pour afficher les images à évaluer;
- le confort visuel: l'effet produit sur la facilité d'observation d'images 3D stéréoscopiques par un système opérant entre les images de test et le moniteur utilisé pour afficher les images à évaluer.

La présente Recommandation reprend six méthodes de la Recommandation UIT-R BT.500; ces méthodes ont donné de bons résultats au cours des vingt dernières années dans des travaux de recherche concernant la qualité de l'image, la qualité de la profondeur et le confort visuel en imagerie stéréoscopique. Il s'agit de:

- la méthode à un seul stimulus (SS);
- la méthode à double stimulus utilisant une échelle de dégradation (DSIS);
- la méthode à double stimulus utilisant une échelle de qualité continue (DSCQS);
- la méthode de comparaison de stimulus (SC);
- la méthode d'évaluation continue de la qualité avec stimulus unique (SSCQE);
- la méthode d'évaluation continue à double stimulus simultanés (SDSCE).

Les méthodes ont parfois été utilisées moyennant de légères modifications, par exemple des échelles différentes ont été utilisées pour le confort visuel. Le mode de présentation et les échelles associés à la méthode d'évaluation de la qualité de l'image, de la qualité de la profondeur et du confort visuel sont récapitulés respectivement dans les Tableaux 1, 2 et 3.

On trouvera ci-après une brève description de chaque méthode, puis une description des éléments communs à toutes les méthodes.

## 2.1 Méthode à un seul stimulus (SS)

La procédure consiste à mener une série d'expériences d'évaluation qui pourraient éventuellement être subdivisées en plusieurs séances séparées par des pauses. Dans chaque expérience, une seule séquence vidéo «*de test*», autrement dit une séquence qui a été traitée au moyen d'un système à l'étude, est présentée et notée de façon indépendante sur l'échelle proposée.

### 2.1.1 Structure d'expérience de la méthode SS

Dans chaque expérience, la présentation de la séquence vidéo «*de test*» à évaluer est précédée et suivie par la présentation d'une image gris moyen. L'image gris moyen précédente peut contenir une cible de fixation, par exemple le numéro de l'expérience, sans disparité et devrait avoir une durée  $\leq 3$  s. L'image gris moyen suivante peut contenir un rappel invitant, par exemple, à procéder à l'évaluation, et devrait avoir une durée suffisante pour permettre à l'observateur de donner une note (par

exemple  $\leq 10$  s). La séquence vidéo «*de test*» devrait généralement avoir une durée d'environ 10 s<sup>1</sup>. La structure d'une expérience type pour la méthode SS est représentée sur la Fig. 1.

### 2.1.2 Echelles d'évaluation de la méthode SS

Pour l'évaluation de la qualité de l'image, on peut utiliser deux échelles d'évaluation: une échelle discrète à cinq notes et l'échelle de qualité continue normalisée par l'UIT (voir le Tableau 1). Les notes de qualité sont: «excellent», «bon», «passable», «médiocre» et «mauvais».

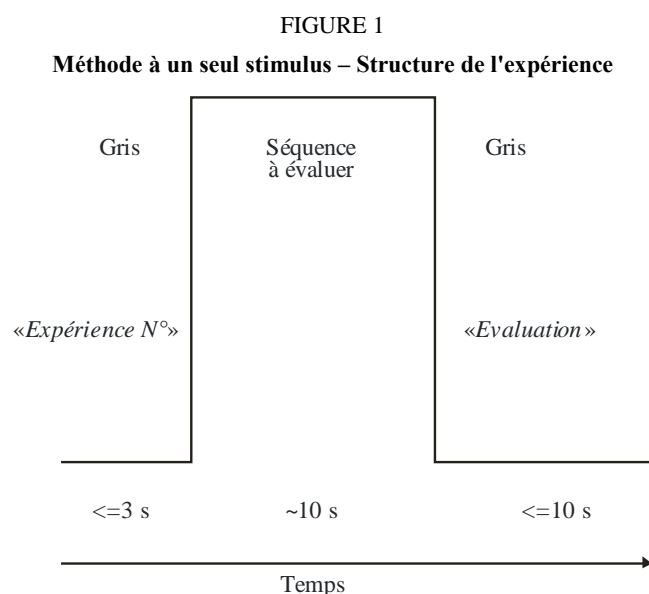
On peut utiliser les mêmes échelles pour l'évaluation de la qualité de la profondeur (voir le Tableau 2). Dans ce cas, il est demandé aux observateurs d'évaluer la qualité de la représentation de la profondeur et non la qualité de l'image proprement dite.

Pour l'évaluation du confort visuel, on peut utiliser deux échelles de notation: une échelle discrète à cinq notes et une échelle de confort continue (voir le Tableau 3). Les notes de confort sont: «très confortable», «confortable», «légèrement inconfortable», «inconfortable» et «extrêmement inconfortable».

### 2.1.3 Données d'opinion de la méthode SS

On appelle «*note d'opinion*» la note donnée à chaque séquence examinée et note moyenne d'opinion (MOS) la moyenne des notes d'opinion obtenues, en règle générale, pour chaque système étudié.

Les séquences vidéo «*de référence*», qui sont les versions des séquences de test qui n'ont subi aucun traitement (voir le § 8), peuvent être incluses dans l'ensemble des séquences. Cette inclusion permet de calculer la «*note différentielle d'opinion*», qui est la différence arithmétique entre la note donnée à la version «*de test*» et celle donnée à la version «*de référence*» de chaque séquence étudiée. On appelle note différentielle moyenne d'opinion (DMOS) la moyenne des notes différentielles d'opinion obtenues pour chaque système étudié.



BT.2021-01

<sup>1</sup> Certains chercheurs ont préconisé l'utilisation de séquences de durée plus longue en s'appuyant essentiellement sur l'hypothèse selon laquelle l'évaluation complète d'un contenu stéréoscopique nécessite plus de temps que celle d'un contenu monoscopique (2D) normal. A ce jour, les éléments empiriques militant en faveur ou contre cette hypothèse sont peu nombreux.

## 2.2 Méthode à double stimulus utilisant une échelle de dégradation (DSIS) (méthode UER)

La méthode à double stimulus (UER) est cyclique en ce sens que l'on présente d'abord une image de référence non dégradée puis la même image dégradée à l'observateur, qui est ensuite prié de donner son avis sur la seconde image, tout en gardant la première à l'esprit. On présente à l'observateur une série d'images ou de séquences d'images dans un ordre aléatoire au cours de séances durant au plus une demi-heure, les images ou les séquences d'images présentant des dégradations aléatoires couvrant toutes les combinaisons requises. L'image de référence fait partie des images ou des séquences d'images à évaluer. La note moyenne pour chaque condition d'essai et chaque image d'essai est calculée à l'issue de la série de séances.

Cette méthode fait appel à une échelle de dégradation, qui conduit généralement à des résultats plus stables pour des dégradations faibles que pour des dégradations importantes. Bien qu'elle ait parfois été utilisée avec des gammes de dégradations limitées, cette méthode est plus adaptée en association avec une gamme de dégradations complète. Le système d'essai aura la configuration indiquée à la Fig. 2.

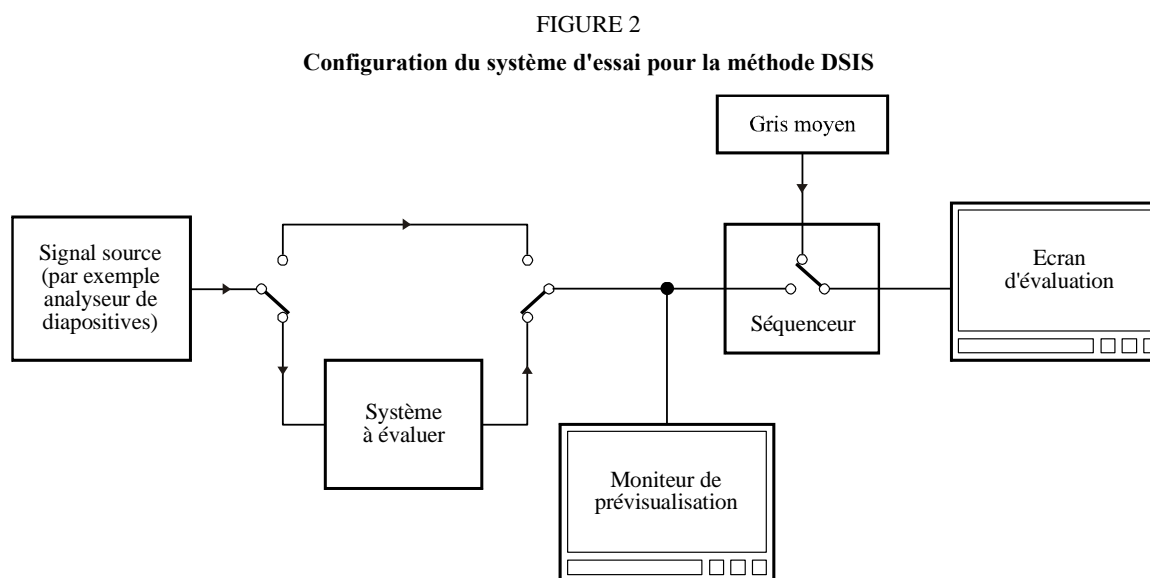
### 2.2.1 Présentation des séquences de test

Une séance d'évaluation se compose d'un certain nombre de présentations. Il existe deux variantes (I et II) de la structure des présentations.

Variante I: L'image ou la séquence de référence et l'image ou la séquence d'essai ne sont présentées qu'une seule fois (Fig. 3a)).

Variante II: L'image ou la séquence de référence et l'image ou la séquence d'essai sont présentées deux fois (Fig. 3b)).

La variante II, plus longue que la variante I, peut être utilisée si une discrimination de très petites dégradations est nécessaire ou si les essais comprennent des séquences animées.







de la même séquence vidéo sont présentées deux fois, soit quatre présentations au total. La configuration générale de l'évaluation devrait être telle que représentée sur la Fig. 4.

### 2.3.1 Présentation des séquences de test

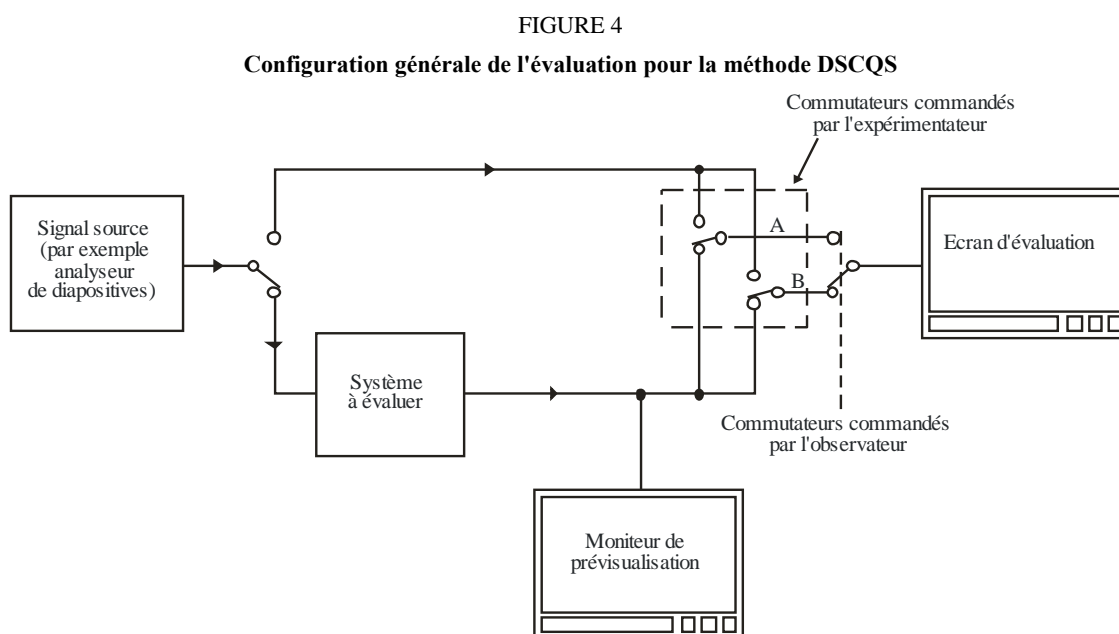
Une séance d'évaluation comprend un certain nombre de présentations. Dans le cas de la variante I, qui ne nécessite qu'un seul observateur, l'observateur peut, pour chaque présentation, passer du trajet A au trajet B et inversement jusqu'à ce qu'il ait une idée de la mesure de la qualité associée à chaque signal. Il peut répéter cette opération deux ou trois fois pendant des laps de temps ne dépassant pas 10 s. Dans la variante II, qui fait appel à plusieurs observateurs simultanément, avant d'enregistrer les résultats, chaque paire de conditions est présentée une ou plusieurs fois pendant un laps de temps égal, afin de permettre aux observateurs de se faire une idée de la mesure des qualités associées à ces conditions. Ensuite, chaque paire est à nouveau présentée une ou plusieurs fois en vue de l'enregistrement des résultats. Le nombre de répétitions dépend de la longueur des séquences de test. Pour des images fixes, une séquence de 3 à 4 s et cinq répétitions (avec évaluation pendant les deux dernières) peut convenir. Pour des images animées avec des artefacts variant dans le temps, une séquence de 10 s avec deux présentations (et évaluation pendant la seconde) peut être appropriée. La structure des présentations est illustrée à la Fig. 5.

### 2.3.2 Echelles d'évaluation de la méthode DSCQS

Dans la méthode DSCQS, il est demandé aux observateurs de noter les séquences vidéo A et B. Pour les évaluations de la qualité de l'image et de la qualité de la profondeur, on peut utiliser l'échelle de qualité continue normalisée par l'UIT (voir les Tableaux 1 et 2). Pour l'évaluation du confort visuel, il convient d'utiliser l'échelle de confort continue dont les notes sont «très confortable», «confortable», «légèrement inconfortable», «inconfortable» et «extrêmement inconfortable» (voir le Tableau 3).

### 2.3.3 Données d'opinion de la méthode DSCQS

Les notes de la version «*de test*» et de la version «*de référence*» de chaque séquence obtenues dans chaque expérience sont utilisées pour calculer les notes différentielles d'opinion, lesquelles sont ensuite utilisées pour calculer la note DMOS pour chaque système étudié.



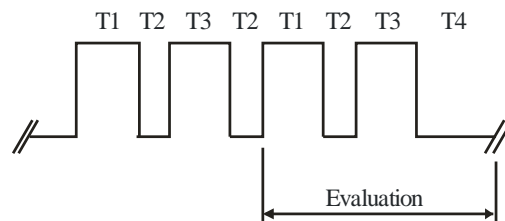
Cette méthode comporte deux variantes, I et II, exposées ci-après.

**Variante I:** L'observateur, en principe seul, est autorisé à passer de la condition A à la B et inversement jusqu'à ce qu'il se soit fait une opinion sur chacune d'elles. Les trajets A et B reçoivent l'image de référence directe ou l'image qui est passée par le système à évaluer, avec une alternance aléatoire d'une condition de test à l'autre. Cette alternance est notée par l'expérimentateur mais elle n'est pas annoncée à l'observateur.

**Variante II:** On présente consécutivement aux observateurs les images provenant des trajets A et B, afin qu'ils se fassent une opinion sur chacune d'elles. En ce qui concerne la fourniture des images sur les trajets A et B pour chaque présentation, on procède comme dans la variante I ci-dessus. La stabilité des résultats obtenus avec cette variante, qui utilise une plage de qualité limitée, est encore à l'étude.

FIGURE 5

Méthode à double stimulus utilisant une échelle de qualité continue – Structure de l'expérience



Phases de présentation:

T1 = 10 s	Séquence de test A
T2 = 3 s	Gris moyen
T3 = 10 s	Séquence de test B
T4 = 5-11 s	Gris moyen

BT.2021-05

## 2.4 Méthode de comparaison par paires (PC)

Dans la méthode de comparaison par paires, un ensemble de séquences «*de test*», autrement dit des séquences qui ont été traitées par des systèmes différents (par exemple avec des débits binaires différents, des algorithmes différents, etc.) sont comparées par paires (c'est-à-dire deux à la fois). Il est demandé aux observateurs de donner une appréciation concernant l'élément préféré de chaque paire dans le contexte du scénario d'évaluation. Le nombre d'appréciations requises est fonction du nombre de systèmes étudiés: les systèmes à évaluer (X, Y, Z, etc.) sont associés selon les  $n(n-1)$  combinaisons possibles (XY, ZY, YZ, etc.). De plus, toutes les paires de séquences doivent être présentées dans les deux ordres possibles (par exemple XY, YX).

### 2.4.1 Structure d'expérience de la méthode PC

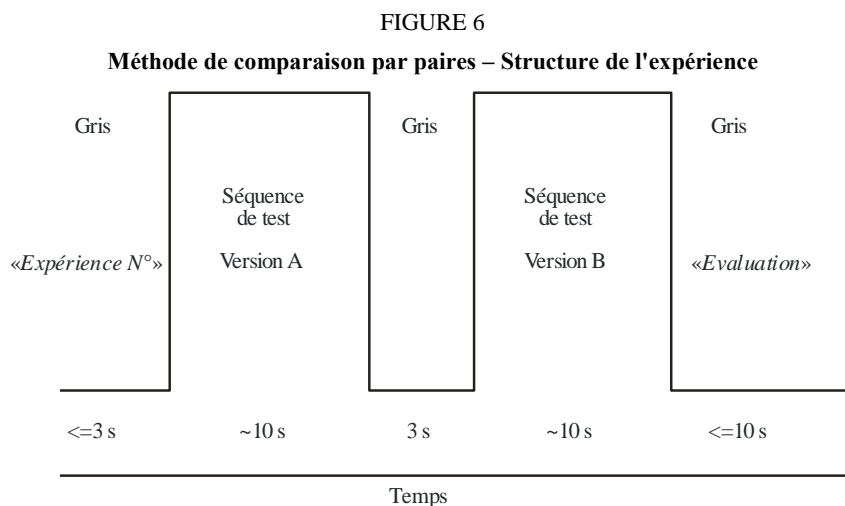
Une expérience commence par la présentation d'une image gris moyen qui peut contenir une cible de fixation, par exemple le numéro de l'expérience, sans disparité et qui devrait avoir une durée  $\leq 3$  s. Les séquences à comparer sont ensuite présentées. Chaque séquence à évaluer devrait généralement avoir une durée d'environ 10 s. Les séquences peuvent être présentées simultanément sur deux écrans (ou côte à côte sur le même écran) ou successivement (par exemple AB) sur le même écran. Dans le deuxième cas, les séquences sont séparées temporellement par la présentation d'une image gris moyen pendant 3 s. L'expérience se termine par la présentation d'une image gris moyen pouvant contenir un rappel invitant, par exemple, à procéder à l'évaluation et devrait avoir une durée suffisante pour permettre à l'observateur de donner une appréciation (par exemple  $\leq 10$  s). Un exemple d'expérience type de comparaison par paires est illustré sur la Fig. 6.

### 2.4.2 Echelles d'évaluation de la méthode PC

Il peut être demandé aux observateurs de donner une simple appréciation de préférence sur une échelle binaire (par exemple A est préféré) ou d'indiquer un niveau de préférence (par exemple A est nettement mieux que B). On peut utiliser les mêmes échelles pour la qualité de l'image, la qualité de la profondeur et le confort visuel (voir les Tableaux 1, 2 et 3).

### 2.4.3 Données d'opinion de la méthode PC

Dans la méthode PC, les appréciations sont données en termes de préférences.



BT.2021-06

## 2.5 Méthode d'évaluation continue de la qualité avec stimulus unique (SSCQE)

Même sur de courts extraits de séquences vidéo stéréoscopiques à codage numérique, les niveaux de la qualité de l'image, de la qualité de la profondeur et du confort visuel peuvent beaucoup fluctuer dans le temps, suivant le contenu de la scène et la durée des artefacts (par exemple courts ou longs) affectant ces trois dimensions de base. La méthode SSCQE a été conçue pour tenir compte de l'incidence de ces fluctuations. Dans cette méthode, la qualité de l'image, la qualité de la profondeur et le confort visuel des séquences vidéo stéréoscopiques sont évalués en continu (c'est-à-dire au fur et à mesure de leur évolution dans le temps). On estime en général que cette méthode est plus représentative de ce que le téléspectateur observe réellement à son domicile.

### 2.5.1 Forme générale du protocole d'évaluation

Les formats utilisés pour l'évaluation par les observateurs sont les suivants:

- *Segment de programme (PS)*: correspond à un type de programme (par exemple sport, actualités, théâtre) traité conformément à l'un des paramètres de qualité (QP) à évaluer (par exemple le débit binaire); chaque segment de programme devrait durer au moins 5 min.
- *Séance de test (TS)*: correspond à une série d'une ou de plusieurs combinaisons différentes PS/QP sans séparation et mises dans un ordre pseudo-aléatoire. Chaque séance de test contient au moins une fois tous les segments de programme et paramètres de qualité, mais pas nécessairement toutes les combinaisons PS/QP; chaque séance de test devrait durer entre 30 et 60 min.

- *Présentation de test (TP)*: correspond à l'intégralité d'un test. Elle peut être subdivisée en séances de test pour respecter les impératifs de durée maximale et pour pouvoir évaluer la qualité de toutes les paires PS/QP. Si le nombre de ces paires est limité, une présentation de test peut comporter plusieurs fois la même séance de test afin que le test dure assez longtemps.

Si l'on veut évaluer la qualité de service, on peut introduire une séquence audio. Dans ce cas, on apportera le même soin au choix de la séquence audio d'accompagnement et de la séquence vidéo avant de procéder au test.

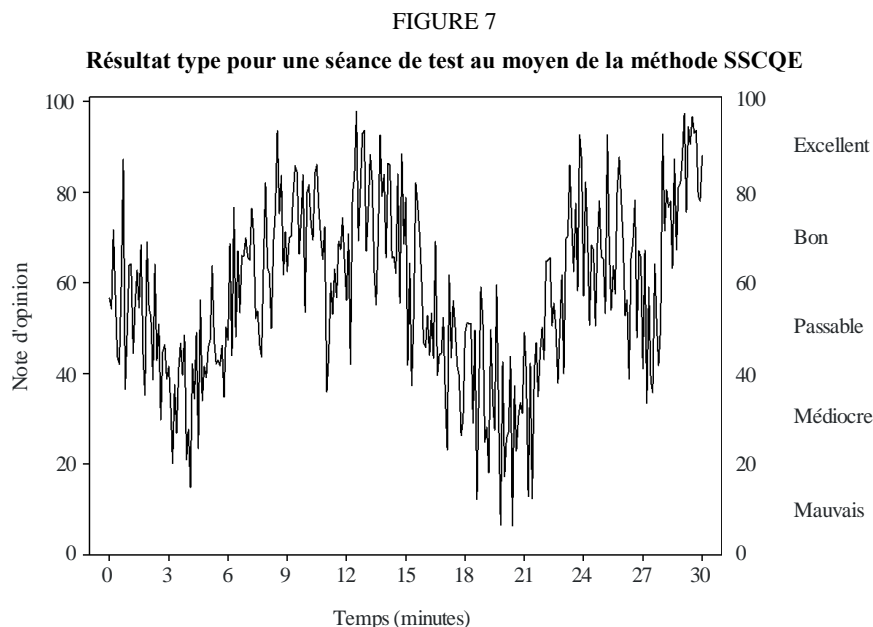
Le format le plus simple pour l'évaluation se composera d'un seul segment de programme et d'un seul paramètre de qualité.

### 2.5.2 Echelles d'évaluation de la méthode SSCQE

Pour l'évaluation de la qualité de l'image et de la qualité de la profondeur, il convient d'utiliser l'échelle de qualité continue normalisée par l'UIT (voir les Tableaux 1 et 2). Pour l'évaluation du confort visuel, il convient d'utiliser l'échelle de confort visuel présentée dans le Tableau 3. La Fig. 7 donne un exemple de résultat obtenu pour une *séance de test* sur une échelle de qualité.

### 2.5.3 Données d'opinion de la méthode SSCQE

Il convient de regrouper les données de toutes les séances de test afin de calculer la note moyenne de qualité en fonction du temps  $q(t)$ . Les résultats peuvent être exprimés en termes de moyenne de toutes les notes de qualité données par les observateurs par segment de programme, contenu vidéo ou séance de test.



BT.2021-07

## 2.6 Méthode d'évaluation continue à double stimulus simultanés (SDSCE)

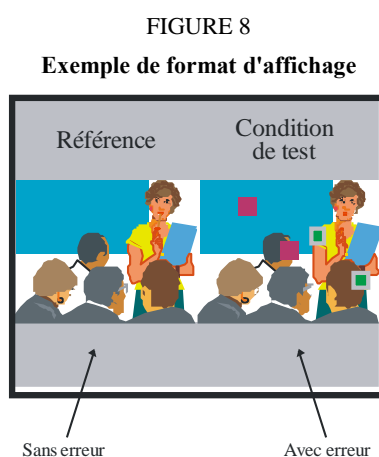
La méthode SSCQE permet de mesurer la qualité de vidéos stéréoscopiques sur des séquences plus longues, représentatives du contenu des vidéos stéréoscopiques et des statistiques d'erreur. Dans la méthode SSCQE, aucune référence n'est utilisée en vue de reproduire des conditions d'observation aussi proches que possible de situations réelles.

Il est nécessaire d'introduire des conditions de référence lorsque la fidélité doit être évaluée. La méthode SDSCE a été élaborée à partir de la méthode SSCQE, en lui apportant de légères

modifications en ce qui concerne la manière de présenter les images aux observateurs et l'échelle de notation. Cette méthode a été proposée au groupe MPEG pour évaluer la tolérance aux erreurs à un débit binaire très faible, mais elle peut tout à fait être employée dans tous les cas où la fidélité d'informations visuelles affectées par une dégradation variable dans le temps doit être évaluée.

### 2.6.1 Présentation des séquences de test

Un groupe d'observateurs a examiné deux séquences en même temps: la première était la séquence de référence, la seconde correspondait aux conditions de test. Si les deux séquences étaient présentées en format d'image standard (SIF, *standard image format*) ou en un format plus petit, elles pouvaient être affichées en parallèle sur le même écran, autrement, il fallait utiliser deux écrans placés côte à côte (voir la Fig. 8).



### 2.6.2 Echelles d'évaluation de la méthode SDSCE

On a demandé aux observateurs d'examiner les différences entre les deux séquences et de juger la fidélité des informations vidéo en déplaçant la glissière du mécanisme de notation. Lorsque la fidélité était parfaite, la glissière devait être placée au sommet de l'échelle de notation (soit 100), et lorsque la fidélité était nulle, elle devait être placée au bas de l'échelle (soit 0). Il était possible d'utiliser l'échelle de qualité continue normalisée par l'UIT (voir les Tableaux 1, 2 et 3).

Les observateurs savaient quelle était la séquence de référence et ils devaient faire connaître leur opinion tout en observant les séquences, pendant toute la durée de celles-ci.

### 2.6.3 Données d'opinion de la méthode SDSCE

Les définitions ci-après sont utilisées pour décrire le protocole de test:

- *Segment vidéo (SV)*: correspond à une séquence vidéo.
- *Condition de test (CT)*: peut être soit un processus vidéo spécifique, une condition de transmission ou les deux. Chaque SV doit être traité conformément à une CT au moins. En outre, les références doivent être ajoutées à la liste des conditions de test, afin de créer des paires référence/référence à évaluer.
- *Séance (S)*: série de paires différentes SV/CT non séparées et ordonnées de façon pseudo-aléatoire. Chaque séance contient au moins une fois tous les SV et les CT, mais pas nécessairement toutes les combinaisons SV/CT.
- *Présentation de test (PT)*: série de séances comprenant toutes les combinaisons SV/CT. Toutes les combinaisons SV/CT doivent être notées par le même nombre d'observateurs (mais pas nécessairement par les mêmes observateurs).

- *Période de notation*: chaque observateur est prié de noter de manière continue pendant une séance.
- *Segment de notation*: segment de 10 s de notation; tous les segments de notation sont obtenus par l'utilisation de groupes de 20 notes consécutives (l'équivalent de 10 s) sans aucun chevauchement.

Une fois que le test a été mené à bien, on établit un (ou plusieurs) fichier(s) de données regroupant toutes les notes des différentes séances (S) qui représentent l'ensemble du matériel de notation de la présentation de test (PT). On peut effectuer un premier contrôle de la validité des données en vérifiant que chaque paire SV/CT a été traitée et qu'un nombre équivalent de notes a été attribué à chacune des paires.

Les données collectées durant l'exécution des tests effectués conformément à ce protocole peuvent être traitées de trois manières différentes:

- analyse statistique de chaque SV distinct;
- analyse statistique de chaque CT distincte;
- analyse statistique globale de toutes les paires SV/CT.

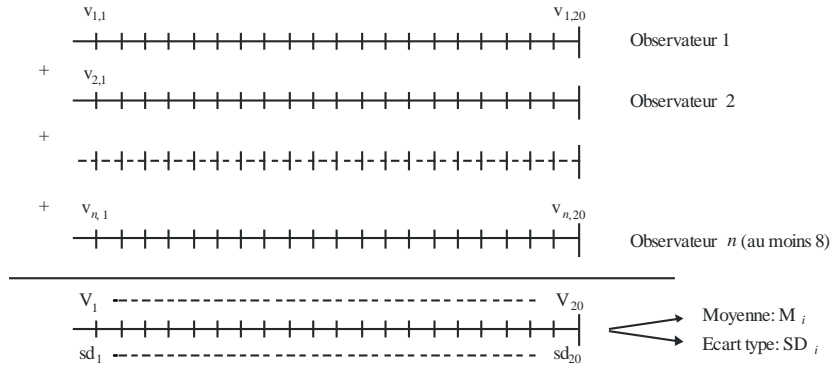
Pour chaque cas, il faut procéder à une analyse composée de quatre étapes:

- Les moyennes et les écarts types sont calculés pour chaque notation par cumul des observations.
- La moyenne et l'écart type sont calculés pour chaque segment de notation, comme illustré à la Fig. 9. Les résultats de cette étape peuvent être représentés par un diagramme temporel (Fig. 10).
- On analyse la distribution statistique des moyennes calculées à l'étape précédente (c'est-à-dire celles correspondant à chaque segment de notation) et leur fréquence d'occurrence. Afin d'éviter un effet de rémanence dû à la combinaison SV × CT précédente, on ne tient pas compte des 10 premières secondes de notation pour chaque échantillon SV × CT.
- On calcule les caractéristiques globales de gêne par cumul des fréquences d'occurrence. Il faut tenir compte dans ce calcul des intervalles de confiance, comme indiqué à la Fig. 11. Des caractéristiques globales de gêne correspondent à cette fonction de distribution statistique cumulative en indiquant la relation entre les moyennes pour chaque segment de notation et leur fréquence cumulative d'occurrence.

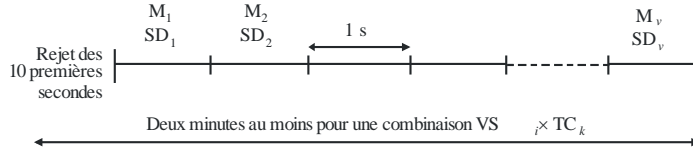
FIGURE 9

Traitement des données

a) Calcul de note moyenne,  $V_i$ , et de l'écart type, SD, par instant de vote pour les observateurs pour chaque séquence de notation de chaque combinaison VS  $\times$  TC



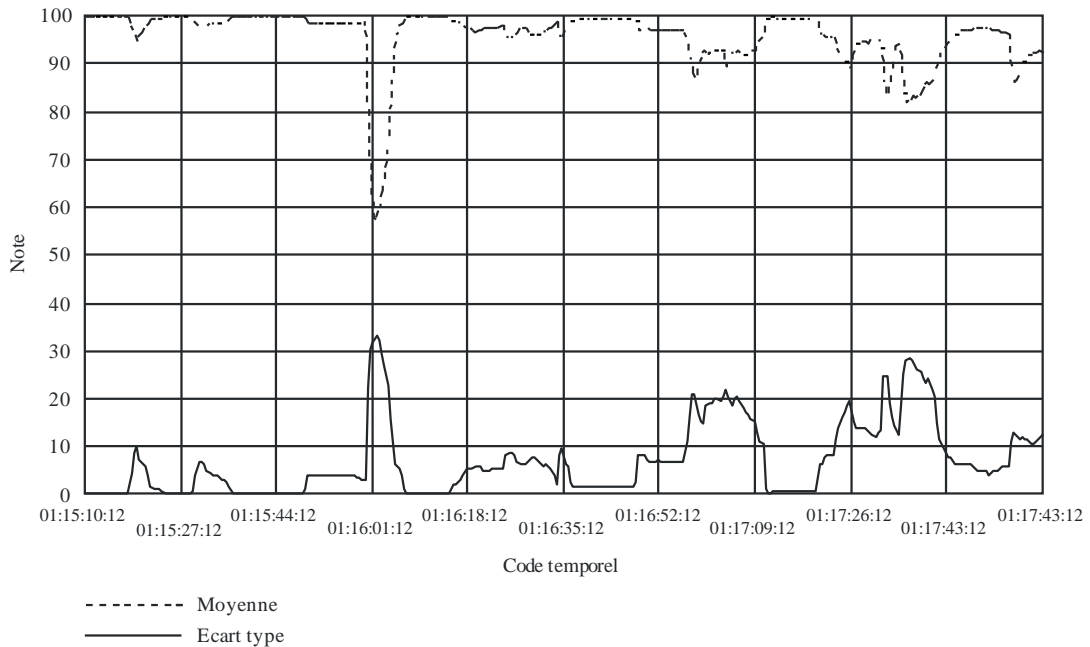
b) Calcul de la moyenne,  $M_i$ , et de l'écart, SD, par séquence de notation d'une seconde pour chaque combinaison VS  $\times$  TC



BT.2021-09.

FIGURE 10

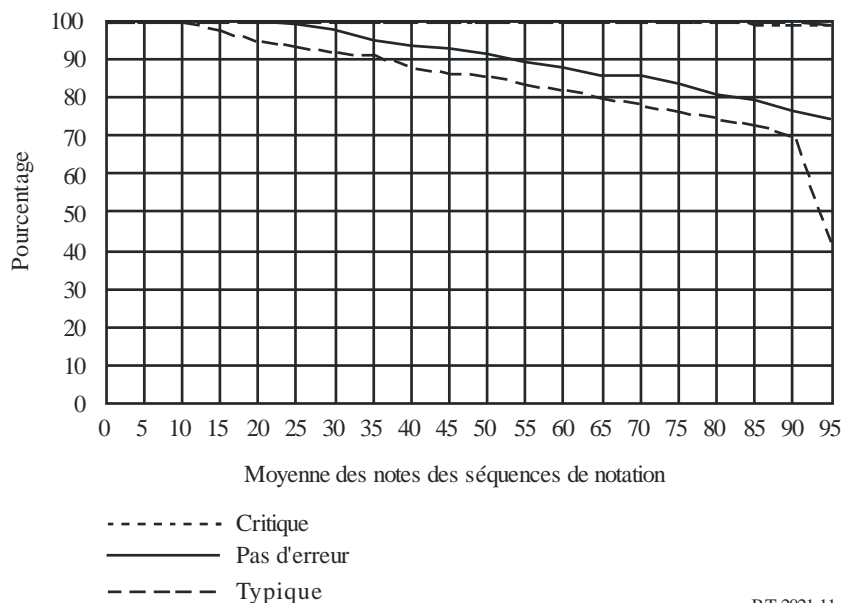
Diagramme temporel brut



BT.2021-10

FIGURE 11

Caractéristiques globales de gêne calculées à partir des distributions statistiques et comprenant l'intervalle de confiance



BT.2021-11

TABLEAU 1

## Méthode d'évaluation subjective de la qualité de l'image

Mode de présentation	Durée de la séquence	Echelle binaire	Echelle discrète	Echelle continue
Méthode à un seul stimulus (SS), telle que décrite au § 6.1 de l'Annexe 1 de la Recommandation UIT-R BT.500.	~10 s		5 Excellent 4 Bon 3 Passable 2 Médiocre 1 Mauvais	
Méthode à double stimulus utilisant une échelle de dégradation (DSIS), telle que décrite au § 4 de l'Annexe 1 de la Recommandation UIT-R BT.500.			5 Imperceptible 4 Perceptible mais non gênant 3 Légèrement gênant 2 Gênant 1 Très gênant	
Méthode à double stimulus utilisant une échelle de qualité continue (DSCQS), telle que décrite au § 5 de l'Annexe 1 de la Recommandation UIT-R BT.500.	~10 s			



TABLEAU 1 (*fin*)

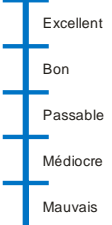

Mode de présentation	Durée de la séquence	Echelle binaire	Echelle discrète	Echelle continue
Méthode de comparaison de stimulus (SC), telle que décrite au § 6.2 de l'Annexe 1 de la Recommandation UIT-R BT.500.	~10 s	A en fonction de B	-3 Beaucoup moins bon -2 Moins bon -1 Un peu moins bon 0 Equivalent 1 Légèrement meilleur 2 Meilleur 3 Bien meilleur	
Méthode d'évaluation continue de la qualité avec stimulus unique (SSCQE), telle que décrite au § 6.3 de l'Annexe 1 de la Recommandation UIT-R BT.500.	~3-5 min			
Méthode d'évaluation continue à double stimulus simultanés (SDSCE), telle que décrite au § 6.4 de l'Annexe 1 de la Recommandation UIT-R BT.500.				

TABLEAU 2

## Méthode d'évaluation subjective de la qualité de la profondeur

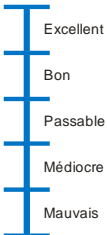
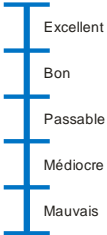
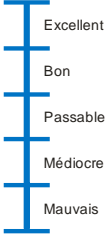
Mode de présentation	Durée de la séquence	Echelle binaire	Echelle discrète	Echelle continue
Méthode à un seul stimulus (SS), telle que décrite au § 6.1 de l'Annexe 1 de la Recommandation UIT-R BT.500.	~10 s		5 Excellent 4 Bon 3 Passable 2 Médiocre 1 Mauvais	
Méthode à double stimulus utilisant une échelle de dégradation (DSIS), telle que décrite au § 4 de l'Annexe 1 de la Recommandation UIT-R BT.500.			5 Imperceptible 4 Perceptible mais non gênant 3 Légèrement gênant 2 Gênant 1 Très gênant	
Méthode à double stimulus utilisant une échelle de qualité continue (DSCQS), telle que décrite au § 5 de l'Annexe 1 de la Recommandation UIT-R BT.500.	~10 s			
Méthode de comparaison de stimulus (SC), telle que décrite au § 6.2 de l'Annexe 1 de la Recommandation UIT-R BT.500.	~10 s	A en fonction de B	-3 Beaucoup moins bon -2 Moins bon -1 Un peu moins bon 0 Equivalents 1 Légèrement meilleur 2 Meilleur 3 Bien meilleur	
Méthode d'évaluation continue de la qualité avec stimulus unique (SSCQE), telle que décrite au § 6.3 de l'Annexe 1 de la Recommandation UIT-R BT.500.	~3-5 min			

TABLEAU 2 (*fin*)


Mode de présentation	Durée de la séquence	Echelle binaire	Echelle discrète	Echelle continue
Méthode d'évaluation continue à double stimulus simultanés (SDSCE), telle que décrite au § 6.4 de l'Annexe 1 de la Recommandation UIT-R BT.500.				

TABLEAU 3

## Méthode d'évaluation subjective du confort visuel

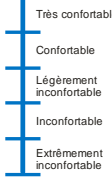
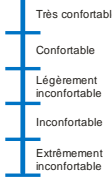
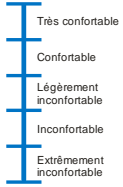

Mode de présentation	Durée de la séquence	Echelle binaire	Echelle discrète	Echelle continue
Méthode à un seul stimulus (SS), telle que décrite au § 6.1 de l'Annexe 1 de la Recommandation UIT-R BT.500.	~10 s		5 Très confortable 4 Confortable 3 Légèrement inconfortable 2 Inconfortable 1 Extrêmement inconfortable	
Méthode à double stimulus utilisant une échelle de dégradation (DSIS), telle que décrite au § 4 de l'Annexe 1 de la Recommandation UIT-R BT.500.			5 Imperceptible 4 Perceptible mais non gênant 3 Légèrement gênant 2 Gênant 1 Très gênant	
Méthode à double stimulus utilisant une échelle de qualité continue (DSCQS), telle que décrite au § 5 de l'Annexe 1 de la Recommandation UIT-R BT.500.	~10 s			
Méthode de comparaison de stimulus (SC), telle que décrite au § 6.2 de l'Annexe 1 de la Recommandation UIT-R BT.500.	~10 s	A en fonction de B	-3 Beaucoup moins bon -2 Moins bon -1 Un peu moins bon 0 Equivalent 1 Légèrement meilleur 2 Meilleur 3 Bien meilleur	

TABLEAU 3 (*fin*)

Mode de présentation	Durée de la séquence	Echelle binaire	Echelle discrète	Echelle continue
Méthode d'évaluation continue de la qualité avec stimulus unique (SSCQE), telle que décrite au § 6.3 de l'Annexe 1 de la Recommandation UIT-R BT.500.	~3-5 min			
Méthode d'évaluation continue à double stimulus simultanés (SDSCE) telle que décrite au § 6.4 de l'Annexe 1 de la Recommandation UIT-R BT.500.				

### 3 Conditions générales d'observation

Les conditions d'observation (luminance et contraste de l'écran, éclairage du fond, distance d'observation, etc.) devraient être cohérentes avec celles utilisées pour la 2D qui sont décrites dans la Recommandation UIT-R BT.2022 (Doc. 6/20) – Conditions générales d'observation pour l'évaluation subjective de la qualité des images de TVDN et de TVHD sur des écrans plats. Deux arguments militent en faveur de cette cohérence. Le premier est que, dans la pratique, les utilisateurs regarderont la TV3D sur les mêmes écrans et dans les mêmes conditions d'observation que pour la 2D. Le deuxième est que les progrès des technologies vidéo de TV3D devront souvent être mesurés par rapport aux progrès des technologies vidéo de la TVHD standard.

La Recommandation UIT-R BT.2022 (Doc. 6/20) spécifie deux critères possibles pour le choix de la distance d'observation. Il convient de choisir la distance d'observation nominale. Pour un système numérique, la distance d'observation nominale est la distance à laquelle deux pixels adjacents forment un angle de 1 minute d'arc depuis l'œil de l'observateur.

Exprimée en fonction de la hauteur de l'image, la distance d'observation nominale est de 4,8H pour une résolution d'image de 1 280 x 720 (Rec. UIT-R BT.1543 et UIT-R BT.1847) et de 3,1H (images statiques) pour une résolution d'image de TVHD de la famille 1 920 x 1 080 (Rec. UIT-R BT.709).

A titre d'illustration, le Tableau 4 donne la distance d'observation nominale en mètres pour un échantillon représentatif de tailles de diagonale de téléviseur.

TABLEAU 4

**Distance d'observation nominale en mètres pour diverses tailles  
de diagonale de téléviseur**

Taille de diagonale (pouces)	Système d'image 1 920 × 1 080	Système d'image 1 280 × 720
	Distance d'observation nominale (mètres)	Distance d'observation nominale (mètres)
32	1,24	1,88
42	1,62	2,47
52	2,01	3,06
62	2,39	3,64
72	2,78	4,23
82	3,17	4,82
92	3,55	5,41
102	3,94	5,99

Il convient de noter que, étant donné que deux pixels adjacents forment un angle de 1 minute d'arc depuis l'œil de l'observateur, la plus petite disparité angulaire (rétinienne) qui peut être représentée par le système (c'est-à-dire la résolution en profondeur du système) à la distance d'observation nominale est égale à 1 minute d'arc (ou, de façon équivalente, à 60 secondes d'arc). D'après des travaux de recherche, près de 97% de la population peut distinguer des disparités horizontales égales ou inférieures à 140 secondes d'arc, et au moins 80% peut détecter des disparités horizontales de 30 secondes d'arc. Par conséquent, la plupart des observateurs devraient pouvoir détecter la plus petite disparité représentable dans les systèmes vidéo 3D actuels à la distance d'observation nominale.

#### 4 Séquences de test

Le choix des séquences de test devrait dépendre de la question expérimentale traitée dans l'étude. D'une manière générale, le contenu des séquences de test (sport, théâtre, film, etc.) et leurs caractéristiques spatio-temporelles devraient être représentatifs des programmes diffusés par le service étudié.

En outre, le contenu des séquences de test stéréoscopiques choisies devrait aussi être relativement confortable à regarder. Le confort visuel des images stéréoscopiques dépend essentiellement des disparités d'image (parallaxe) et des conditions d'observation. Il convient donc de veiller à ce que les disparités ne dépassent pas les limites indiquées dans le paragraphe qui suit, sauf si l'étude vise expressément à mesurer le confort visuel. De plus, il convient, chaque fois que possible, de mesurer et d'indiquer la moyenne, l'écart type et la plage (min/max) de la distribution de disparité des séquences de test.

La parallaxe, les incohérences entre les images de gauche et de droite, et la distribution et la modification de parallaxe sont des éléments dont il convient de tenir compte lors du choix d'images 3D stéréoscopiques faciles à observer en vue d'une évaluation. La relation entre une image 3D stéréoscopique facile à observer et la parallaxe, les incohérences entre les images de gauche et de droite, et la distribution et la modification de parallaxe est décrite dans les paragraphes qui suivent.

#### 4.1 Limites de confort visuel

Une disparité/parallaxe excessive entraîne un inconfort visuel, peut-être en raison de l'accentuation du conflit entre accommodation et vergence. Il a donc été proposé, pour réduire ce conflit, de faire en sorte que les disparités dans l'image stéréoscopique soient suffisamment faibles pour que les profondeurs perçues des objets se situent dans les limites d'une «zone de confort». Pour définir ces limites, plusieurs approches ont été proposées. L'une des approches consiste à utiliser une mesure de la parallaxe relative à l'écran, exprimée en pourcentage de la largeur de l'écran. Des valeurs de 1% pour les disparités croisées/négatives et de 2% pour les disparités non croisées/positives (pour une valeur totale d'environ 3%) ont été proposées. Selon une autre approche, on utilise la profondeur de champ de l'œil pour délimiter la zone de confort. Pour les conditions d'observation types de programmes de télévision, les chercheurs ont pris comme hypothèse une profondeur de champ comprise entre  $\pm 0,2D$  (dioptries) et  $\pm 0,3D$  (dioptries). Pour une résolution d'image de TVHD de  $1\,920 \times 1\,080$  (Rec. UIT-R BT.709) et une distance d'observation nominale de  $3,1H$ , ces valeurs correspondent approximativement à une parallaxe relative à l'écran de  $\pm 2\%$  et de  $\pm 3\%$ . Enfin, une troisième approche consiste à spécifier les limites de confort en fonction de la disparité rétinienne et à fixer ces limites à  $\pm 1^\circ$  d'angle visuel à la fois pour les disparités positives et pour les disparités négatives.

Il est à noter que ces différentes approches ont tendance à converger vers les mêmes limites de confort. Rappelons que, à la distance d'observation nominale, deux pixels adjacents forment un angle de 1 minute d'arc depuis l'œil de l'observateur. Donc 60 pixels correspondent à  $1^\circ$  d'angle visuel. On peut ainsi spécifier facilement les limites de confort en termes de disparité rétinienne (pour un observateur moyen). Par exemple, pour une résolution d'image de TVHD de  $1\,920 \times 1\,080$  (Rec. UIT-R BT.709), 1% ( $\sim 19,2$  pixels) correspond approximativement à 20 minutes d'arc, 2% à  $\sim 40$  minutes d'arc et 3% à  $\sim 60$  minutes d'arc (ou, de façon équivalente, à  $1^\circ$ ).

Il convient de noter que même si, à la distance d'observation nominale, deux pixels adjacents forment toujours un angle de 1 minute d'arc, la distance physique (par exemple en mm) entre ces pixels est plus grande pour des écrans plus grands (le nombre de pixels reste identique, mais les dimensions physiques de l'écran augmentent). Par conséquent, pour les limites plus élevées (par exemple  $\pm 3\%$ ) et pour les écrans plus grands, la distance physique entre les points correspondants (autrement dit la parallaxe des deux vues en mm) pourrait dépasser l'écartement des yeux de l'observateur moyen ( $\sim 63$ - $65$  mm), ce qui pourrait donner lieu à un inconfort plus important.

#### 4.2 Discordances entre les images de gauche et de droite

Dans les systèmes 3D stéréo, une image 3D binoculaire est formée par la présentation de l'image de gauche et de celle de droite à l'œil correspondant. Toute discordance entre ces deux images peut entraîner un stress psychophysique et, dans certains cas, l'observation en 3D peut échouer. Par exemple, lors de la prise de vues et de l'affichage de programmes de TV3D stéréoscopique, il peut y avoir des distorsions géométriques, comme une incohérence de taille, un décalage vertical ou une erreur de rotation, entre l'image de gauche et celle de droite. Il est préférable que les images de test ne comportent pas de telles distorsions géométriques. Voir le § 3.2.1 de l'Annexe 4 du Rapport UIT-R BT.2160-2 pour plus d'informations.

Les discordances entre l'image de gauche et celle de droite dont il convient de tenir compte lors du choix d'images 3D stéréoscopiques faciles à observer en vue d'une évaluation sont les suivantes:

- discordance géométrique (taille, déplacement vertical, rotation);
- discordance de brillance (niveau de blanc et de noir);
- diaphonie.

### **4.3 Plage, distribution et modification de parallaxe**

Dans le cas des images stéréoscopiques, les distributions de parallaxe sont corrélées avec le confort visuel.

La distribution de parallaxe d'images stéréoscopiques est discontinue pendant les trames de changement de scène. Les cas de parallaxe extrême ou de modification subite de parallaxe entraînent un inconfort visuel; il est donc important de prendre en considération avec soin la parallaxe des images de test. Voir le § 3.2.2 de l'Annexe 4 du Rapport UIT-R BT.2160-2 pour plus d'informations.

D'une manière générale, étant donné que les études utilisant des séquences de test stéréoscopiques peuvent donner lieu à un certain inconfort visuel, il est recommandé d'utiliser, chaque fois que possible, des séquences de test dont la disparité ne dépasse pas les limites de confort, même si des dépassements occasionnels peuvent être autorisés.

## **5 Appareils expérimentaux**

Les appareils expérimentaux (serveur vidéo, écran, etc.) devraient être en mesure d'afficher des séquences de test HD pleine résolution, par exemple en utilisant un format de mise en trame HDMI, ce qui permettrait d'élargir l'éventail des études qui peuvent être réalisées.

A ce jour, aucun écran de référence pour l'évaluation de la TV3D n'a été normalisé. On s'attend donc à ce que la plupart des chercheurs utilisent les écrans actuels de TV3D grand public. Étant donné que les caractéristiques de ces écrans peuvent varier d'un fabricant à l'autre, les chercheurs sont vivement encouragés à indiquer les paramètres de l'écran utilisé dans l'étude.

## **6 Observateurs**

### **6.1 Taille de l'échantillon**

Il est généralement recommandé de faire appel à au moins 30 observateurs. Il est toutefois reconnu que le nombre effectif dépendra des objectifs spécifiques de l'étude, tout en sachant que les considérations relatives à la taille de l'échantillon pour les études de la 3D ne sont pas différentes de celles utilisées pour les études de la 2D.

### **6.2 Sélection des observateurs en fonction de leur vue**

Il convient de sélectionner les observateurs en fonction de leur acuité visuelle, de leur perception des couleurs et de leur vision stéréoscopique au moyen de tests cliniques actuels de la vue, par exemple l'équivalent des mires de Snellen pour l'acuité visuelle, les planches d'Ishihara ou un équivalent pour les couleurs et le test de Randot ou un équivalent pour la vision stéréoscopique. Il est à noter que les tests de vision stéréoscopique tels que les tests de Randot, Stereo Fly ou Frisby mesurent généralement des disparités rétinienne allant d'environ 20 à 400 secondes d'arc. Les chercheurs sont encouragés à indiquer les statistiques pertinentes concernant l'acuité stéréoscopique des observateurs participant à l'étude. Si une analyse plus détaillée de l'acuité stéréoscopique des participants est nécessaire, les chercheurs peuvent recourir aux images présentées dans l'Appendice 1.

## **7 Instructions à l'intention des observateurs**

Les instructions devraient être adaptées aux dimensions (par exemple qualité de la profondeur, confort, etc.) à étudier. En particulier, les directives déontologiques pour les études de la 3D sont plus strictes que celles qui sont généralement utilisées pour l'évaluation de la qualité d'images 2D car les participants peuvent ressentir un inconfort visuel. D'une manière générale, pour les études portant sur

la 3D, il faut veiller à bien informer les participants des motivations de chaque étude ainsi que des éventuelles conséquences négatives de l'exposition aux stimuli utilisés dans l'étude.

## 8 Durée d'une séance

Si les séquences à observer sont censées être confortables, la durée de la séance de test peut être aussi longue que celle utilisée pour les études de la 2D (à savoir ~20-40 minutes entrecoupées de pauses). Si on sait que les séquences contiennent une parallaxe excessive et qu'elles sont donc susceptibles d'être inconfortables, la durée devrait être limitée.

## 9 Utilisation des séquences vidéo de référence

Si la séquence de référence est disponible, les chercheurs souhaiteront peut-être l'inclure dans l'ensemble des séquences de test. La référence est généralement une version de la séquence de test qui n'a subi aucun traitement (autrement dit la séquence source d'origine). Pour les études stéréoscopiques, la principale référence est la séquence stéréoscopique non traitée d'origine. Le plan d'expérience pourrait toutefois aussi inclure la version monoscopique de la référence (autrement dit une seule vue de la séquence source d'origine); par exemple, pour l'étude du confort visuel, il pourrait être utile de faire une comparaison avec la référence monoscopique. La version monoscopique de la référence devrait être présentée en mode 3D (par exemple la vue de gauche est présentée à la fois à l'œil gauche et à l'œil droit et on utilise les mêmes paramètres matériels 3D que pour la vraie séquence stéréoscopique). L'inclusion de la référence dans le plan d'expérience comporte deux avantages importants: elle permet de mesurer la transparence (autrement dit la fidélité) offerte par l'algorithme ou la technologie à l'étude<sup>2</sup> et elle sert de repère de qualité élevée afin d'aider à stabiliser les notes<sup>3</sup>.

## 10 Variabilité des réponses

En général, les notes fournies par les observateurs dans les expériences d'évaluation subjective sont assez variables. Les différences entre les observateurs peuvent simplement être liées aux caractéristiques de la population de référence et, pour les atténuer, on peut augmenter la taille de l'échantillon.

Toutefois, une partie de la variabilité peut être liée à l'évolution des réponses fournies par les différents observateurs au cours de l'expérience. Cette évolution traduit une évolution des critères d'évaluation qui pourrait être due au fait d'acquérir davantage de pratique de la tâche, à un apprentissage des caractéristiques des artefacts, etc. Pour réduire autant que possible les effets négatifs de cette variabilité, les chercheurs devraient prévoir des procédures de formation appropriées (tâche, niveau de dégradation, etc.), recourir à de multiples randomisations (à savoir présenter les séquences de test dans différents ordres aléatoires aux différents observateurs), et répéter les présentations (ce qui permettrait aussi de mesurer l'éventuelle évolution des réponses fournies).

---

<sup>2</sup> La transparence (fidélité) décrit le fait qu'un codec ou un système n'entraîne aucune dégradation par rapport à un système de transmission idéal. Il va de soi qu'on peut mesurer la transparence en comparant les notes attribuées à la séquence de référence et celles attribuées à la séquence traitée par l'algorithme ou la technologie à l'étude.

<sup>3</sup> Il est reconnu que la stabilité des notes dans l'espace (c'est-à-dire dans des laboratoires différents) et dans le temps (c'est-à-dire dans le même laboratoire mais à des moments différents) pourrait aussi être améliorée avec l'utilisation de repères de faible qualité. Cela étant, l'UIT a l'intention de produire/définir dès maintenant des repères normalisés de faible qualité pour l'évaluation des technologies d'imagerie stéréoscopique.



## 11 Critères de rejet d'observateurs

Les critères de rejet d'observateurs pour les méthodes présentées au § 2 sont décrits dans la Recommandation UIT-R BT.500.

## 12 Analyse statistique

Les analyses statistiques pour l'étude des systèmes d'imagerie 3D sont les mêmes que pour les systèmes d'imagerie 2D.

# Appendice 1

## Images pour l'examen de la vue

### 1 Examen de la vue

Le Tableau 5 donne la liste des mires utilisées pour l'examen de la vue. Les 12 tests ont été choisis selon la hiérarchie du système visuel humain, depuis le plus bas niveau jusqu'au niveau le plus élevé. On trouvera ci-après la description de huit tests de la vue (VT, *vision test*) principaux, puis de quatre autres réservés à l'usage clinique. Les observateurs doivent avoir une stéréopsie normale, c'est-à-dire qu'ils doivent satisfaire au test VT-04 pour la stéréopsie fine et au test VT-07 pour la stéréopsie dynamique. Les six autres tests sont destinés à une caractérisation plus détaillée. Les mires doivent être regardées à une distance égale au triple de la hauteur de l'écran.

TABLEAU 5

### Images stéréoscopiques pour l'examen de la vue

N°	Propriété testée	Objet du test	Contenu
1	Perception simultanée	Aptitude à percevoir simultanément, et dans la position correcte, des images en présentation dichoptique	L'image d'une cage est présentée à un œil et l'image d'un lion à l'autre œil
2	Fusion binoculaire	Aptitude à percevoir deux images dichoptiques dans l'œil gauche et l'œil droit sous la forme d'une image unique	L'image destinée à l'un des yeux comporte deux points, celle destinée à l'autre œil en comporte trois, avec un point en commun
3	Stéréopsie grossière	Aptitude à percevoir des images en présentation dichoptique avec une parallaxe, sous la forme d'une image unique donnant une impression de relief (profondeur) grossière	Les images destinées aux deux yeux sont une paire stéréo d'images représentant une libellule dont les ailes sont déployées
4	Stéréopsie fine	Aptitude à percevoir des images en présentation dichoptique avec une parallaxe, sous la forme d'une image unique donnant une impression de relief (profondeur) fine	On présente neuf losanges contenant chacun quatre cercles dont un seul a une petite parallaxe

TABLEAU 5 (fin)

N°	Propriété testée	Objet du test	Contenu
5	Limite de fusion avec croisement	Aptitude à percevoir comme une image unique des images en présentation dichoptique avec des disparités croisées	On présente une paire stéréo de barres dont la parallaxe croisée varie de 10'/s
6	Limite de fusion sans croisement	Aptitude à percevoir comme une image unique des images en présentation dichoptique avec des disparités non croisées	On présente une paire stéréo de barres dont la parallaxe non croisée varie de 11'/s
7	Stéréopsie dynamique	Aptitude à percevoir la profondeur dans des images formées par des stéréogrammes en points aléatoires en mouvement	Stéréogramme dynamique en points aléatoires
8	Acuité binoculaire	Acuité binoculaire, y compris la dissymétrie éventuelle de l'acuité monoculaire qui pourrait empêcher une bonne stéréopsie	Caractères E d'orientation et de dimensions variées
9	Strabisme horizontal	Déviations horizontales de l'œil que le patient ne parvient pas à maîtriser	Segment vertical et segment horizontal
10	Strabisme vertical	Déviations verticales de l'œil que le patient ne parvient pas à maîtriser	Segment vertical et segment horizontal
11	Aniséiconie	Perception d'images différentes en termes de taille et de forme par les deux yeux fixant un même objet	L'image de gauche est constituée des caractères «[o]» et celle de droite des caractères «o]», les caractères «o]» ayant la même position
12	Cyclophorie	Rotation d'un œil, ou de l'autre, autour de l'axe antéropostérieur lorsque la fusion est empêchée	L'image de gauche représente un cadran d'horloge et celle de droite, les aiguilles d'une horloge marquant 6 h

NOTE 1 – Ces images de test sont au format 1125/60/I (voir la Recommandation UIT-R BT.709).

NOTE 2 – On peut se procurer ces images de test auprès de l'Institute of Image Information and Television Engineers (ITE), 3-5-8 Shibakoen, Minato-ku, Tokyo 105-0011, Japon, téléphone: 81-3-3432-4675, courriel: [ite@ite.or.jp](mailto:ite@ite.or.jp).

On trouvera ci-après, côte à côte, les images miniatures de droite et de gauche destinées à la fusion libre avec croisement, aux fins d'explication.

### 1) VT-01: Perception simultanée (test du lion)

Test de l'aptitude à percevoir simultanément, et dans la position correcte, des images en présentation dichoptique. A un œil, on présente l'image d'une cage et à l'autre œil, l'image d'un lion dont la position se déplace de 12'/s. La taille de chaque image est fixée à 10°, ce qui permet aux observateurs de capter les images dans leur zone paramaculaire. Les observateurs dotés d'une vue normale peuvent voir le lion dans la cage à un certain instant de la période de présentation.

FIGURE 12  
Mire pour le test VT-01



BT.2021-12

## 2) VT-02: Fusion binoculaire (test aux 4 points de Worth)

Test de l'aptitude à percevoir deux images dichoptiques dans l'œil gauche et l'œil droit comme une seule image. L'image destinée à l'un des yeux comporte deux points, celle destinée à l'autre œil en comporte trois, avec un point en commun. Les observateurs dotés d'une vue normale voient quatre points.

FIGURE 13  
Mire pour le VT-02



BT.2021-13

## 3) VT-03: Stéréopsie grossière (test de la libellule)

Test de l'aptitude à percevoir des images en présentation dichoptique avec une parallaxe, sous la forme d'une image unique donnant une impression de relief (profondeur) grossière. Les images destinées aux deux yeux sont une paire stéréo d'images représentant une libellule dont les ailes sont déployées. Les observateurs dotés d'une vue normale voient les ailes en avant de l'écran.

FIGURE 14  
Mire pour le test VT-03

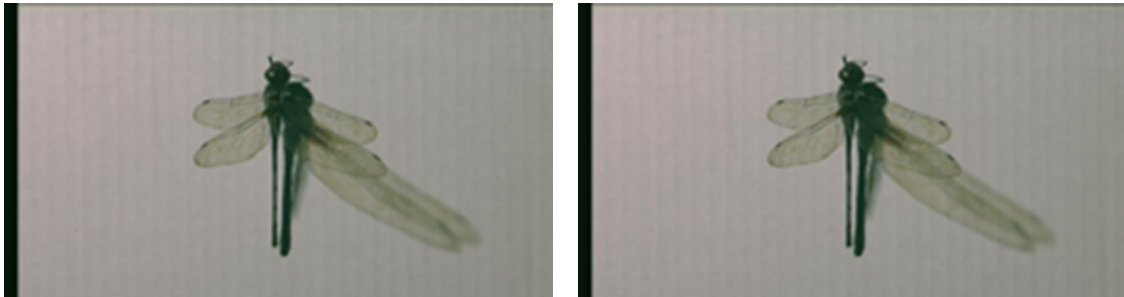


Image de droite

Image de gauche

BT.2021-14

#### 4) VT-04: Stéréopsie fine (test des cercles)

Test de l'aptitude à percevoir des images en présentation dichoptique avec une parallaxe, sous la forme d'une image unique donnant une impression de relief (profondeur) fine. On présente neuf losanges contenant chacun quatre cercles dont un seul a une petite parallaxe. Les observateurs dotés d'une vue normale voient le cercle avec une petite parallaxe en avant de l'écran. Le Tableau 6 donne le numéro des tests, les réponses correctes et l'angle de stéréopsie à la distance  $3 H$ .

TABLEAU 6

#### Réponses correctes et parallaxe

Test N°	Réponses correctes	Angle de stéréopsie à la distance $3 H$ (")
1	En bas	480
2	A gauche	420
3	En bas	360
4	En haut	300
5	En haut	240
6	A gauche	180
7	A droite	120
8	A gauche	60
9	–	0

FIGURE 15

## Mire pour le test VT-04

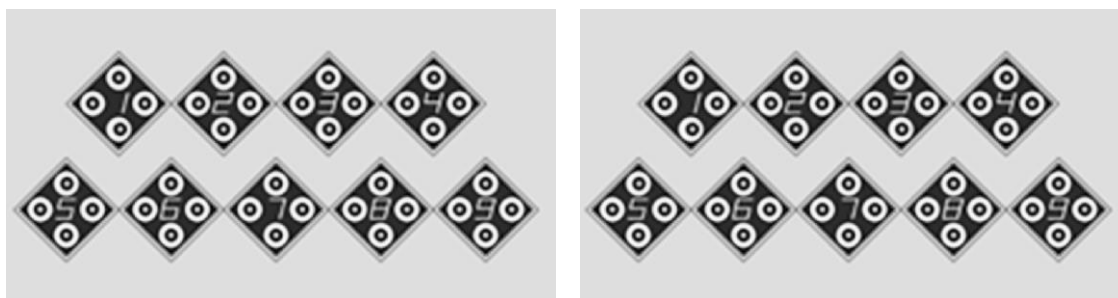


Image de droite

Image de gauche

BT.2021-15

### 5) VT-05: Limite de fusion avec croisement (test des barres)

Test de l'aptitude à percevoir comme une image unique des images en présentation dichoptique avec des disparités croisées. On présente une paire stéréo de barres dont la parallaxe varie de 10'/s. Il est possible de mesurer les limites de la fusion pour les séries croissante et décroissante. Il est demandé aux observateurs de signaler leur rupture de fusion dès qu'ils voient deux images dans la série croissante et le rétablissement de la fusion dès qu'ils voient les images dichoptiques comme une image unique dans la série décroissante.

FIGURE 16

## Mire pour le test VT-05



Image de droite

Image de gauche

BT.2021-16

### 6) VT-06: Limite de fusion sans croisement (test des barres)

Test de l'aptitude à percevoir comme une image unique des images en présentation dichoptique avec des disparités non croisées. Les images présentées sont les mêmes que pour le test précédent avec croisement, mais les images de droite et de gauche sont interverties.

FIGURE 17  
Mire pour le test VT-06



Image de droite

Image de gauche

BT.2021-17

**7) VT-07: Stéréopsie dynamique (test avec stéréogramme dynamique en points aléatoires)**

Test de l'aptitude à percevoir la profondeur dans des images formées par des stéréogrammes en points aléatoires en mouvement. Les observateurs dotés d'une vue normale perçoivent une forme rectangulaire et un mouvement sinusoïdal en profondeur dans le stéréogramme.

FIGURE 18  
Mire pour le test VT-07

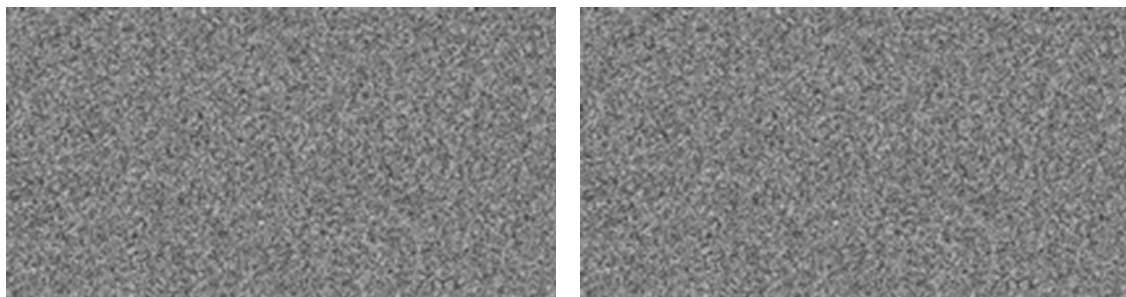


Image de droite

Image de gauche

BT.2021-18

**8) VT-08: Acuité binoculaire (test d'acuité)**

Test de l'acuité binoculaire avec fusion binoculaire, y compris une dissymétrie éventuelle de l'acuité monoculaire qui pourrait empêcher une bonne stéréopsie. Les images contiennent quatre colonnes et cinq lignes composées de caractères E d'orientation et de dimensions variées. Les deux colonnes centrales peuvent être vues avec les deux yeux, les deux colonnes de gauche peuvent être vues seulement avec l'œil gauche, et les deux colonnes de droite seulement avec l'œil droit. Les observateurs dotés d'une vue normale peuvent distinguer correctement l'orientation des caractères E. Les dimensions des caractères correspondent à des acuités d'environ 1,0, 0,5, 0,33, 0,25 et 0,125 à la distance 3 H.

FIGURE 19  
Mire pour le test VT-08

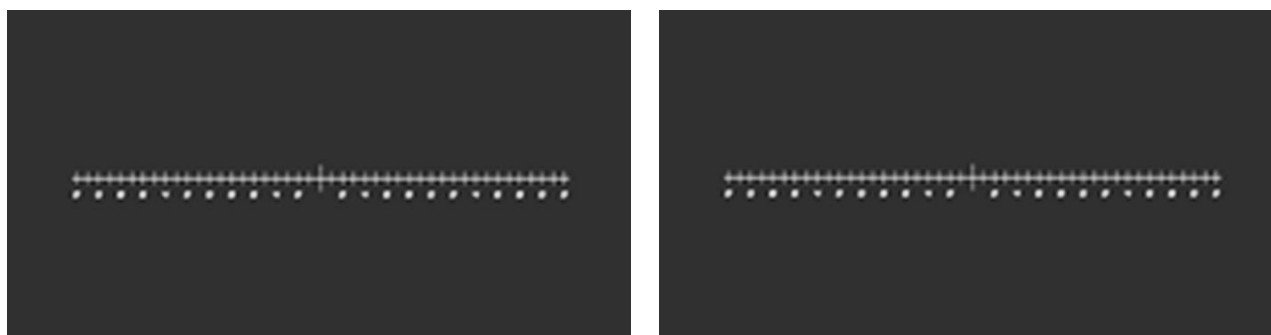


BT.2021-19

**9 & 10) VT-09: Strabisme horizontal (test de Maddox horizontal) et VT-10: Strabisme vertical (test de Maddox vertical)**

Ces mires permettent de mesurer la déviation horizontale et verticale de l'œil. La position des axes visuels l'un par rapport à l'autre est différente de celle requise par les conditions physiologiques. Les images sont constituées d'un segment vertical et d'un segment horizontal. Les observateurs dotés d'une vue normale voient le point d'intersection des segments aux alentours du milieu des segments. Le nombre de graduations donne la dioptrie prismatique pour un écartement des yeux de 65 mm à la distance 3,02 H.

FIGURE 20  
Mire pour le test VT-09



BT.2021-20

FIGURE 21  
Mire pour le test VT-10



BT.2021-21

**11) VT-II: Aniséiconie (test des caractères «[ ]»)**

Perception d'images différentes en termes de taille et de forme par les deux yeux fixant un même objet. L'image de gauche est constituée des caractères «[o» et celle de droite des caractères «o]», les caractères «o» ayant la même position. Les observateurs dotés d'une vue normale voient les caractères «[«et]» comme ayant la même taille et la même hauteur.

FIGURE 22

Mire pour le test VT-11



BT.2021-22

**12) VT-12: Cyclophorie (test de l'horloge)**

Rotation d'un œil autour de l'axe antéropostérieur uniquement lorsqu'il est couvert et que la fusion est empêchée. L'image de gauche représente un cadran d'horloge et celle de droite, les aiguilles d'une horloge marquant six heures. Les observateurs dotés d'une vue normale voient une horloge indiquant exactement six heures.

FIGURE 23

Mire pour le test VT-12



BT.2021-23