

Union internationale des télécommunications

UIT-T

SECTEUR DE LA NORMALISATION
DES TÉLÉCOMMUNICATIONS
DE L'UIT

P.800.2

(07/2016)

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

Méthodes d'évaluation objective et subjective de la qualité
vocale et de vidéo

**Interprétation des notes moyennes d'opinion et
informations à communiquer**

Recommandation UIT-T P.800.2

UIT-T



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

Vocabulaire et effets des paramètres de transmission sur l'opinion des usagers	P.10–P.19
Caractéristiques des terminaux vocaux	P.30–P.39
Systèmes de référence	P.40–P.49
Appareils de mesures objectives	P.50–P.59
Mesures électroacoustiques objectives	P.60–P.69
Mesures de la sonie vocale	P.70–P.79
Méthodes d'évaluation objective et subjective de la qualité vocale	P.80–P.89
Caractéristiques des terminaux vocaux	P.300–P.399
Appareils de mesures objectives	P.500–P.599
Méthodes d'évaluation objective et subjective de la qualité vocale et de vidéo	P.800–P.899
Qualité audiovisuelle dans les services multimédias	P.900–P.999
Aspects relatifs à la qualité de transmission et à la qualité de service aux points de terminaison des réseaux à protocole Internet	P.1000–P.1099
Communications impliquant des véhicules	P.1100–P.1199
Modèles et outils d'évaluation de la qualité de flux médias transmis en continu	P.1200–P.1299
Évaluation des téléconférences	P.1300–P.1399
Lignes directrices pour l'analyse et l'évaluation statistiques des mesures de qualité et le compte rendu de ces mesures	P.1400–P.1499
Méthodes d'évaluation objective et subjective de la qualité des services autres que les services vocaux ou de vidéo	P.1500–P.1599

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

Recommandation UIT-T P.800.2

Interprétation des notes moyennes d'opinion et informations à communiquer

Résumé

La Recommandation UIT-T P.800.2 présente certains types de note moyenne d'opinion (MOS) parmi les plus courants et décrit les informations minimales qui doivent accompagner les valeurs MOS pour qu'elles puissent être interprétées correctement.

Historique

Édition	Recommandation	Approbation	Commission d'études	ID unique*
1.0	UIT-T P.800.2	14-05-2013	12	11.1002/1000/11934
2.0	UIT-T P.800.2	29-07-2016	12	11.1002/1000/12973

Mots clés

Évaluation par catégories absolues, ACR, note moyenne d'opinion, MOS, modèle objectif, informations à communiquer, expérience subjective.

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

AVANT-PROPOS

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

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

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

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

NOTE

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

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

DROITS DE PROPRIÉTÉ INTELLECTUELLE

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

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

© UIT 2019

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

Table des matières

	Page
1	Domaine d'application 1
2	Références..... 1
3	Définitions 1
3.1	Termes définis ailleurs 1
3.2	Termes définis dans la présente Recommandation 1
4	Abréviations et acronymes 1
5	Conventions 2
6	Introduction..... 2
7	Valeurs MOS subjectives 2
8	Interprétation des valeurs MOS 4
9	Considérations relatives aux vidéos..... 5
10	Analyse statistique des notes MOS..... 6
11	Valeurs MOS objectives 6
12	Informations à communiquer avec les valeurs MOS subjectives 6
13	Informations à communiquer avec les valeurs MOS objectives 8
14	Notation 8
	Bibliographie..... 9

Recommandation UIT-T P.800.2

Interprétation des notes moyennes d'opinion et informations à communiquer

1 Domaine d'application

La présente Recommandation présente certains types de note moyenne d'opinion (MOS) parmi les plus courants et décrit les informations minimales qui doivent accompagner les valeurs MOS pour qu'elles puissent être interprétées correctement.

Il convient de noter que le présent texte n'a pas pour objectif de fournir des lignes directrices définitives pour les évaluations subjectives ou objectives. La bibliographie située à la fin de la présente Recommandation est constituée de documents contenant des renseignements plus détaillés.

2 Références

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

[UIT-T P.800.1] Recommandation UIT-T P.800.1 (2006), *Terminologie des notes moyennes d'opinion*.

3 Définitions

3.1 Termes définis ailleurs

Aucun.

3.2 Termes définis dans la présente Recommandation

La présente Recommandation définit les termes suivants:

3.2.1 condition: un des ensembles de cas d'utilisation évalués lors d'une expérience subjective; on parle souvent de circuit fictif de référence (HRC) dans le cadre d'expériences portant sur des vidéos.

3.2.2 sous-condition: sous-ensemble d'une condition défini par une caractéristique particulière du cas d'utilisation, par exemple le matériel vocal d'un locuteur donné.

3.2.3 sujet: participant d'une expérience subjective.

3.2.4 vote: réponse d'un sujet à une question sur une échelle d'évaluation pour un échantillon d'essai individuel ou une interaction.

4 Abréviations et acronymes

La présente Recommandation utilise les abréviations et les acronymes suivants:

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

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

DMOS note moyenne d'opinion de dégradation (*degradation mean opinion score*)

HRC circuit fictif de référence (*hypothetical reference circuit*)

MOS	note moyenne d'opinion (<i>mean opinion score</i>)
MUSHRA	test multi stimuli avec référence et repère cachés (<i>multi-stimulus test with hidden reference and anchor</i>)
QCIF	quart de format intermédiaire commun (<i>quarter common intermediate format</i>)
SSCQE	évaluation continue de la qualité avec stimulus unique (<i>single stimulus continuous quality evaluation</i>)
VGA	réseau vidéo graphique (<i>video graphics array</i>)

5 Conventions

Aucune.

6 Introduction

Les qualités audio et vidéo sont, par nature, des grandeurs subjectives. Par conséquent, la référence en matière de qualités audio et vidéo est l'opinion de l'utilisateur. Toutefois, l'opinion d'une personne quant à ce qui est "bon" peut être sensiblement différente de celle d'une autre personne – elles n'ont raison ni l'une ni l'autre, pas plus qu'elles n'ont tort.

Avant qu'une nouvelle technique de transmission audio ou vidéo ne soit déployée, il est conseillé d'évaluer la qualité de la transmission au moyen d'une ou de plusieurs expériences subjectives. Une expérience subjective vise à recueillir les opinions de plusieurs personnes (les "sujets") quant à la qualité de fonctionnement d'un système, pour un certain nombre de cas d'utilisation bien définis (les "conditions")¹. La note moyenne d'opinion (MOS) d'une condition donnée est simplement la moyenne des opinions (les "votes") recueillies pour ce cas d'utilisation.

Les algorithmes de mesure objective de la qualité ont pour objectif de prévoir la valeur MOS qui serait obtenue pour un signal d'entrée donné au cours d'une expérience subjective. Par conséquent, lorsque l'on interprète une valeur MOS obtenue par une procédure objective, il est essentiel de comprendre le modèle de base de l'expérience dont on prévoit les résultats.

Il existe plusieurs types différents de valeurs MOS et de nombreuses méthodes d'essai différentes permettant de les obtenir. L'objet de la présente Recommandation est de donner au lecteur une vision d'ensemble des principaux points à prendre en compte lors de l'interprétation de valeurs MOS et de lui indiquer les renseignements minimaux qui doivent accompagner les valeurs MOS lorsqu'elles sont communiquées.

7 Valeurs MOS subjectives

Types de notes MOS

On pense en général, à tort, que les valeurs MOS ne concernent que les services vocaux. Pourtant, la procédure consistant à demander à des sujets d'évaluer la qualité d'un contenu peut être appliquée aux services vidéo et aux services audio généraux aussi simplement qu'aux services vocaux. Il est aussi possible de demander aux sujets d'attribuer une note à la qualité audiovisuelle globale d'un service. L'UIT a élaboré plusieurs normes décrivant différents aspects relatifs à l'évaluation subjective des applications vidéo et des applications audio générales, outre celle des applications vocales. Ces textes sont indiqués dans la bibliographie.

¹ Lors d'expériences portant sur des vidéos, les conditions sont souvent appelées "circuits fictifs de référence" (HRC).

Les expériences subjectives peuvent être réparties en deux grandes catégories: les expériences passives et les expériences interactives. Au cours d'une expérience subjective passive, des échantillons d'essai préenregistrés présentant les conditions étudiées sont soumis aux sujets. On demande à ces derniers d'écouter et/ou de regarder de manière passive le matériel expérimental et de donner leur opinion au moyen de l'échelle d'évaluation mise à leur disposition. Au cours d'une expérience interactive, deux sujets ou plus prennent part de manière active à une conversation au moyen d'un équipement conçu pour simuler les cas d'utilisation étudiés. Il est souvent demandé aux sujets de réaliser certaines tâches afin de stimuler la conversation et les interactions. La plupart des expériences ont tendance à être de nature passive. Toutefois, certains aspects de l'expérience utilisateur, tels que les conséquences du retard de réception et de l'écho, ne peuvent être pris en compte qu'au moyen de scénarios impliquant une conversation.

Méthode d'essai et échelle d'évaluation

Lors d'une expérience subjective, il est demandé aux sujets d'exprimer leurs opinions au moyen d'une "échelle d'évaluation". Cette échelle sert à traduire l'évaluation de la qualité émise par un sujet en une valeur numérique dont on peut calculer la moyenne sur l'ensemble des sujets et en fonction d'autres facteurs expérimentaux.

Plusieurs échelles d'évaluation sont couramment utilisées. Leurs avantages respectifs n'entrent pas dans le cadre de la présente Recommandation. La plus couramment utilisée est l'échelle d'évaluation par catégories absolues (ACR) à cinq niveaux:

Excellent	5
Bon	4
Satisfaisant	3
Médiocre	2
Mauvais	1

L'échelle ACR est une échelle discrète, ce qui signifie que la réponse du sujet est limitée à l'une des cinq valeurs ci-dessus. Toutefois, le processus de calcul de la moyenne utilisé pour combiner les résultats des différents sujets produit des valeurs MOS qui ne sont pas nécessairement des nombres entiers. Certaines échelles d'évaluation sont constituées de plus de cinq niveaux discrets, alors que d'autres permettent aux sujets de donner des réponses situées entre les niveaux définis.

La partie "absolue" de l'évaluation ACR reflète le fait qu'il est demandé aux sujets d'attribuer une note pour chaque échantillon pris individuellement. Certaines échelles d'évaluation, telles que l'échelle d'évaluation par catégories de dégradation (DCR), consistent à demander aux sujets leur opinion concernant la différence entre un échantillon traité de façon à présenter la condition étudiée et une version non traitée du même échantillon. La valeur MOS découlant d'une telle expérience est souvent appelée note MOS de dégradation ou note DMOS.

Dans la plupart des modèles expérimentaux, il est demandé aux sujets de noter la qualité de courts échantillons audio ou vidéo. La durée de ces échantillons est généralement de l'ordre de 6 à 10 secondes, car cela laisse suffisamment de temps aux sujets pour se faire une opinion, sans toutefois introduire de distorsion vers la fin de l'échantillon. Il est difficile de refléter la totalité d'une condition avec un seul échantillon d'une telle durée. Par conséquent, il est généralement demandé aux sujets de noter plusieurs échantillons représentant un même cas d'utilisation. Par exemple, dans le cas d'une expérience portant sur la voix, chaque condition de réseau étudiée peut être reflétée dans des échantillons vocaux provenant de trois locuteurs de sexe masculin et de trois locutrices de sexe féminin. Des valeurs MOS peuvent être obtenues pour la totalité de la condition, en calculant la moyenne sur l'ensemble des sujets et l'ensemble des locuteurs, ou pour une sous-condition, par exemple pour un locuteur particulier ou pour les locuteurs d'un genre donné.

Les méthodes d'essai telles que l'évaluation continue de la qualité avec stimulus unique (SSCQE) utilisent des échantillons d'une durée nettement plus longue et consistent à demander au sujet de mettre à jour son opinion concernant la qualité de façon continue au cours de la présentation de l'échantillon. Cette méthode permet d'obtenir, pour chaque sujet, une séquence de notes relatives à la qualité, plutôt qu'une valeur unique reflétant son opinion.

Certaines méthodes d'essai consistent à demander au sujet de répondre à plusieurs questions. Non seulement cela permet d'obtenir davantage de renseignements concernant les conditions étudiées, mais il peut aussi s'agir d'une partie nécessaire du modèle expérimental. Par exemple, la méthode d'essai de la Recommandation UIT-T P.835 consiste à demander aux sujets de donner leur opinion sur la qualité de la voix d'une part et, d'autre part, sur le bruit, puis de donner une note sur la qualité de manière globale. Il a été constaté que cette procédure produisait des résultats plus stables dans le cas des dispositifs de suppression du bruit que la méthode d'essai ACR à une seule question.

Il convient de noter que certaines questions peuvent ne pas porter directement sur la qualité, mais concerner un aspect différent des communications. Par exemple, la Recommandation [b-UIT-T P.800] définit une échelle d'effort d'écoute pour les expériences portant sur la voix. De la même manière, dans certaines expériences impliquant une conversation, les questions posées aux sujets concernent leur expérience pendant leur temps de parole plutôt que pendant leur temps d'écoute.

8 Interprétation des valeurs MOS

Dans ce qui suit, on s'intéresse en premier lieu aux valeurs MOS relatives à la voix. Cependant, beaucoup d'éléments mentionnés dans les sous-sections s'appliquent de la même manière pour les valeurs MOS relatives aux formats vidéo, audio et audiovisuels. Les principales différences concernant le cas des formats vidéo sont décrites dans le paragraphe suivant.

On considère aussi souvent, à tort, qu'un codec vocal particulier se voit attribuer une note MOS donnée. Cela vient notamment de l'utilisation courante de modèles d'évaluation objective de la qualité, qui produisent des résultats très reproductibles. Ces modèles sont conçus pour prévoir ou estimer les résultats d'expériences subjectives. Néanmoins, pour tout codec donné à un débit binaire donné, la valeur MOS obtenue lors d'une expérience subjective peut varier de manière significative d'une expérience à l'autre. Il y a un certain nombre de raisons à cela.

Premièrement, les valeurs MOS exactes obtenues pour une condition particulière lors d'une expérience subjective peuvent être influencées par un grand nombre de facteurs, dont voici une liste non exhaustive:

- les instructions données aux sujets et la formulation utilisée pour l'échelle d'opinion;
- l'équipement utilisé pour présenter le matériel (téléphone, casque, haut-parleurs);
- présentation monaurale, diotique binaurale ou stéréophonique;
- volume de la présentation;
- environnement acoustique;
- préparation des sujets;
- profil du sujet, par exemple son âge et son niveau d'exposition à la technologie;
- différences relatives à l'interprétation et à l'utilisation des échelles d'évaluation en fonction des cultures;
- matériel vocal (contenu phonétique et caractéristiques du locuteur);
- langue (présence/absence, prévalence et importance de certains sons et de certaines transitions en particulier).

Deuxièmement, la valeur MOS exacte obtenue pour une condition donnée lors d'une expérience subjective dépend de la qualité des autres conditions de l'expérience. Par exemple, une condition

correspondant au codec vocal de la Recommandation UIT-T G.729 peut obtenir une note supérieure à 3,9 lors d'une expérience ACR où la plupart des autres conditions présentées sont de qualité inférieure; mais, au contraire, elle peut obtenir une note nettement inférieure à 3,9 si la plupart des autres conditions s'avèrent de qualité supérieure.

Troisièmement, si une expérience est réalisée avec des codecs fonctionnant avec des largeurs de bande audio différentes, la présence de conditions avec une largeur de bande plus importante aura pour effet de réduire la note MOS obtenue pour les conditions qui présentent une largeur de bande audio inférieure. La plus grande largeur de bande audio d'une expérience portant sur la voix est souvent appelée "contexte" de l'expérience. Par exemple, une condition correspondant au codec vocal de la Recommandation UIT-T G.711 obtiendra généralement une note supérieure à 4,0 dans une expérience ACR à bande étroite (300-3 700 Hz); cependant, elle obtiendra vraisemblablement une note comprise entre 3,5 et 3,7 lors d'une expérience ACR à bande étendue (50-7 000 Hz), en raison de la présence d'échantillons à bande étendue de meilleure qualité.

Ces deux derniers points mettent en évidence le fait que les sujets des expériences ont tendance à adapter leur utilisation de l'échelle d'évaluation en fonction du contenu de l'expérience. En effet, les expériences bien conçues débutent par une phase d'exercice au cours de laquelle des exemples de plusieurs conditions, y compris la meilleure et la pire, sont présentés aux sujets.

L'une des conséquences les plus importantes découlant des considérations décrites ci-dessus est qu'il n'est pas approprié de comparer directement des valeurs MOS obtenues lors d'expériences différentes, à moins que ces expériences aient été conçues spécifiquement en vue d'être comparées et, quand bien même, des analyses statistiques doivent être effectuées sur les données pour garantir la validité de la comparaison.

9 Considérations relatives aux vidéos

Une grande partie des considérations décrites plus haut au sujet des expériences subjectives relatives à la voix s'appliquent aussi dans le cas d'expériences portant sur des vidéos. Les conditions expérimentales, souvent appelées circuits fictifs de référence (HRC), définissent généralement différentes combinaisons de codec vidéo, débit binaire, fréquence d'image et conditions de transmission. Les facteurs ayant une incidence sur les valeurs MOS exactes obtenues pour une condition particulière comprennent, sans s'y limiter:

- l'équipement utilisé pour présenter le matériel (technologie d'affichage, fréquence de rafraîchissement, contraste, etc.);
- l'environnement d'observation (couleur, température et niveau de luminosité);
- distance d'observation (généralement exprimée comme le rapport entre la distance d'observation et la hauteur d'affichage); et
- le contenu vidéo.

Le dernier point revêt une importance particulière dans le cas des expériences portant sur des vidéos. Le choix du matériel expérimental est un facteur bien plus important pour les expériences portant sur des vidéos que pour celles portant sur la voix. Cela provient du fait que le contenu d'une séquence vidéo peut avoir une incidence considérable sur l'efficacité avec laquelle elle peut être codée. Par exemple, les informations contenues dans une séquence mettant en scène des images sportives avec des déplacements rapides sont bien plus importantes que dans le cas d'une séquence extraite d'une conférence où l'on ne voit que la tête et les épaules de l'orateur.

Dans le cas des expériences portant sur des vidéos, le contexte primaire est déterminé par la résolution de l'image vidéo. En général, une expérience subjective ne porte pas sur plusieurs résolutions différentes et, par conséquent, les valeurs MOS de vidéos concernent une résolution en particulier, par exemple 480 p ou 1080 p. Dans les cas où l'on considère plusieurs résolutions dans la même expérience, le contexte sera défini comme la résolution comportant le plus grand nombre de lignes.

Dans ce cas de figure, il est important d'indiquer si les images de résolution inférieure sont affichées dans leur format d'origine ou si elles sont redimensionnées vers le format de l'expérience ayant la plus haute résolution.

10 Analyse statistique des notes MOS

L'analyse statistique de valeurs MOS subjectives sort du cadre de la présente Recommandation. Cependant, les valeurs MOS doivent être accompagnées de suffisamment de renseignements pour permettre la réalisation d'une analyse statistique de base, par exemple le calcul d'un intervalle de confiance pour chaque condition. Pour toute condition ou sous-condition donnée, ces renseignements comprennent le nombre de votes ainsi que leur moyenne et leur écart type.

11 Valeurs MOS objectives

Un modèle de qualité objectif a pour objet de prévoir la valeur MOS qu'obtiendrait un signal audio ou vidéo lors d'une expérience subjective. Comme indiqué plus haut, la valeur MOS exacte découlant d'une expérience donnée pour un codec particulier ou pour une chaîne de transmission particulière dépend de nombreux facteurs relatifs à la façon dont l'expérience a été conçue et réalisée. Il s'agit donc pour les personnes élaborant un modèle objectif de prévoir les résultats d'une expérience idéale. Cette expérience est généralement réalisée selon une méthode d'essai particulière, le plus souvent une évaluation ACR, et comprend un échantillon équilibré des distorsions rencontrées dans le domaine d'application étudié.

Par exemple, la conversion définie dans la Recommandation [b-UIT-T P.862.1] utilise la sortie brute provenant du modèle objectif de la Recommandation UIT-T P.862 et la convertit de sorte qu'elle s'inscrive dans un intervalle déterminé en calculant la moyenne des résultats d'un grand nombre d'expériences subjectives réalisées selon la méthode ACR, telle qu'elle est décrite dans la Recommandation [b-UIT-T P.800]. Une conversion similaire est appliquée sur les résultats du modèle de la Recommandation UIT-T P.863.

L'un des avantages du modèle objectif est que les résultats sont reproductibles et, par conséquent, que les mesures réalisées à des moments et dans des endroits différents peuvent être directement comparées. Il convient toutefois de rester attentif, car des facteurs tels que le choix du matériel expérimental ainsi que tout pré- ou post-traitement peuvent tout de même altérer les résultats.

Pour des raisons qui devraient maintenant sembler évidentes, dans les mêmes conditions, des modèles objectifs différents peuvent produire des prévisions différentes pour les valeurs MOS. Par exemple, les modèles des Recommandations UIT-T P.862.1 et UIT-T P.863 ne produisent pas exactement les mêmes prévisions pour les valeurs MOS correspondant à la parole codée selon la Recommandation UIT-T G.729, quand bien même ce codec entre dans le champ d'application des deux modèles. Cela est en partie dû au fait que les deux modèles ont été entraînés et optimisés sur la base d'expériences subjectives différentes. Par conséquent, lorsque l'on compare des prévisions MOS objectives avec des seuils, par exemple en vue d'assurer la surveillance d'accords sur les niveaux de service ou pour le déclenchement d'une alarme, les seuils doivent être choisis dans le contexte du modèle ayant fourni la prévision.

12 Informations à communiquer avec les valeurs MOS subjectives

On trouvera dans le Tableau 1 la description des informations qui doivent accompagner les valeurs MOS subjectives ainsi que les informations complémentaires qu'il est recommandé d'indiquer.

Dans le cas où une expérience a été réalisée conformément à une Recommandation de l'UIT, les informations concernant la méthode utilisée peuvent généralement se résumer à une simple référence à la norme pertinente et à la méthode particulière employée; les écarts par rapport aux procédures normalisées doivent toutefois être indiqués.

Il est important de toujours fournir des informations sur les échantillons d'essai utilisés pour les expériences passives. Dans le cas d'échantillons vidéo, il peut être utile de fournir des informations plus détaillées, par exemple en indiquant si des séquences particulières comportent des mouvements panoramiques ou des changements de scène.

Tableau 1 – Informations minimales à communiquer avec les valeurs MOS subjectives

Information	Type d'expérience	Fourniture
<p>Méthode</p> <ul style="list-style-type: none"> Passive ou interactive Fondée sur des échantillons ou évaluation continue Évaluation absolue ou relative des échantillons Instructions et question(s) soumises aux sujets Désignation des niveaux de l'échelle d'évaluation Échelle d'évaluation discrète ou continue Durée des échantillons <p>Ou</p> <ul style="list-style-type: none"> Recommandation de l'UIT et méthode utilisée 	Tous	Obligatoire
<p>Protocole expérimental</p> <ul style="list-style-type: none"> Objectif de l'expérience Date et lieu de l'expérience Informations sur le déroulement de l'expérience Modèle expérimental, par exemple constitution de bloc Nombre de sessions et durée Nombre de sujets Profil des sujets, distributions en fonction de l'âge et du sexe Type de sujets, par exemple non avertis ou experts Information concernant l'équipement utilisé Environnement de présentation, c'est-à-dire le bruit de fond, le niveau de luminosité, etc. 	Tous	Recommandée
<p>Informations relatives aux conditions HRC</p> <ul style="list-style-type: none"> Nombre de conditions Liste des conditions Moyenne des votes pour chaque condition (note MOS) Écart type des votes pour chaque condition Nombre de votes pour chaque condition 	Tous	Obligatoire
<p>Informations sur les sous-conditions</p> <ul style="list-style-type: none"> Liste des facteurs des sous-conditions Valeurs MOS des sous-conditions Nombre de votes et variance pour chaque sous-condition 	Tous	Facultative
<p>Présentation audio</p> <ul style="list-style-type: none"> Largeur(s) de bande audio Canaux audio, par exemple monophoniques, stéréophoniques, etc. Volume de la présentation audio Méthode de présentation audio, par exemple haut-parleurs ou casque (monaurale, diotique binaurale, etc.) 	Vocale, audio, audiovisuelle	Obligatoire

Tableau 1 – Informations minimales à communiquer avec les valeurs MOS subjectives

Information	Type d'expérience	Fourniture
Présentation vidéo Résolution(s) des images vidéo (Notes 1 et 2) Distance d'observation en fonction de la hauteur d'affichage, par exemple 3H Type et taille du dispositif, c'est-à-dire télévision, tablette, téléphone, etc. Application, c'est-à-dire visiophonie, vidéo à la demande, télévision linéaire, etc.	Vidéo, audiovisuelle	Obligatoire
Langue	Passive audio, audiovisuelle	Obligatoire
Nombre et sexe des locuteurs	Passive audio, audiovisuelle	Obligatoire
Type de matériel vidéo, par exemple sport, tête et torse	Passive vidéo, audiovisuelle	Obligatoire
Type d'audio, par exemple musique classique, musique populaire, bande-son de film	Passive audio, audiovisuelle	Obligatoire
NOTE 1 – L'utilisation d'images entrelacées doit être signalée. NOTE 2 – Si l'expérience contient différentes résolutions d'image, il doit être indiqué si les images de résolution inférieure sont affichées dans leur format d'origine ou si elles sont redimensionnées vers une résolution plus haute.		

13 Informations à communiquer avec les valeurs MOS objectives

Dans le cas d'une valeur MOS obtenue au moyen d'un modèle objectif de l'UIT-T, il suffit généralement d'indiquer le modèle utilisé et tout paramètre modifié par rapport aux conditions par défaut. Lorsqu'un modèle non normalisé est utilisé, les informations indiquées dans la section "Méthode" du Tableau 1 doivent être fournies afin de décrire le modèle expérimental dont on prévoit les résultats. Il est aussi recommandé de fournir les informations concernant le type de matériel expérimental utilisé dans les expériences visant à tester et/ou entraîner le modèle objectif.

14 Notation

La Recommandation [UIT-T P.800.1] fournit un système de notation générique de haut niveau pouvant être utilisé pour aider à identifier la provenance d'une valeur MOS. Le système de notation de la Recommandation UIT-T P.800.1 est utile pour donner une vue d'ensemble du contexte dont découle une valeur MOS, mais il ne remplace pas la description détaillée du contexte, devant être fournie conformément à la présente Recommandation, dans tous les cas, lorsque c'est possible.

Bibliographie

L'UIT a normalisé un certain nombre de méthodes d'évaluation subjective pour diverses applications. Celles qui sont le plus couramment utilisées sont indiquées ci-dessous.

[b-UIT-T G.729] Recommandation UIT-T G.729 (2012), *Codage de la parole à 8 kbit/s par prédiction linéaire avec excitation par séquences codées à structure algébrique conjuguée.*

La série UIT-T P.800 contient de nombreuses Recommandations portant sur les évaluations subjectives et objectives de la qualité de la voix, en particulier les Recommandations suivantes:

[b-UIT-T P.800] Recommandation UIT-T P.800 (1996), *Méthodes d'évaluation subjective de la qualité de transmission.*

[b-UIT-T P.805] Recommandation UIT-T P.805 (2007), *Évaluation subjective de la qualité conversationnelle.*

[b-UIT-T P.835] Recommandation UIT-T P.835 (2003), *Méthodologie d'évaluation subjective des systèmes de communication vocale avec algorithmes de suppression du bruit.*

[b-UIT-T P.862.1] Recommandation UIT-T P.862.1 (2001), *Fonction de conversion des notes brutes P.862 en notes moyennes d'opinion de qualité de liaison objective (MOS-LQO).*

[b-UIT-T P.863] Recommandation UIT-T P.863 (2014), *Évaluation objective de la qualité d'écoute perçue.*

La série UIT-T P.900 comprend des Recommandations portant sur l'évaluation de contenus multimédias.

[b-UIT-T P.910] Recommandation UIT-T P.910 (2008), *Méthodes subjectives d'évaluation de la qualité vidéographique pour les applications multimédias.*

[b-UIT-T P.911] Recommandation UIT-T P.911 (1998), *Méthodes d'évaluation subjective de la qualité audiovisuelle pour applications multimédias.*

[b-UIT-T P.912] Recommandation UIT-T P.912 (2016), *Méthodes d'évaluation subjective de la qualité vidéo pour les tâches de reconnaissance.*

L'UIT-R a aussi publié des Recommandations portant sur l'évaluation subjective de la qualité audio et vidéo.

[b-UIT-R BS.1116-1] Recommandation UIT-R BS.1116-1 (1997), *Méthodes d'évaluation subjective des dégradations faibles dans les systèmes audio y compris les systèmes sonores multivoies.*

[b-UIT-R BS.1534-1] Recommandation UIT-R BS.1534-1 (2003), *Méthode d'évaluation subjective du niveau de qualité intermédiaire des systèmes de codage.*

[b-UIT-R BT.500-13] Recommandation UIT-R BT.500-13 (2012), *Méthodologie d'évaluation subjective de la qualité des images de télévision.*

[b-UIT-R BT.710-4] Recommandation UIT-R BT.710-4 (1998), *Méthodes d'évaluation subjective de la qualité d'image en télévision à haute définition.*

NOTE – Les méthodes normalisées décrites par l'UIT-R ne portent pas toutes sur le calcul de notes moyennes d'opinion. Les Recommandations UIT-R publiées fournissent une documentation et des références complètes pour toutes les méthodes correspondantes. Pour une meilleure description et une clarification des méthodes citées ci-dessus, le lecteur pourra se reporter aux Recommandations pertinentes publiées par l'UIT-R.

Le manuel suivant décrit de manière approfondie les méthodes d'évaluation subjective et les bonnes pratiques associées.

[b-Manuel UIT-T] *Procédures pratiques d'évaluation subjective* (2011).

SÉRIES DES RECOMMANDATIONS UIT-T

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