

UIT-R

Secteur des Radiocommunications de l'UIT

Recommandation UIT-R BS.2132-0

(10/2019)

**Méthode d'évaluation subjective de la
qualité en termes de différences audibles
pour les systèmes sonores utilisant
plusieurs stimuli sans référence connue**

Série BS

Service de radiodiffusion sonore



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 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 œuvre 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

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

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

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

Publication électronique
Genève, 2020

© UIT 2020

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 BS.2132-0

Méthode d'évaluation subjective de la qualité en termes de différences audibles pour les systèmes sonores utilisant plusieurs stimuli sans référence connue

(2019)

Domaine d'application

La présente Recommandation décrit une méthode utilisant plusieurs stimuli sans référence connue pour l'évaluation subjective de la qualité en termes de différences audibles pour les systèmes audio. Cette méthode reflète de nombreux aspects de la méthode MUSHRA décrite dans la Recommandation UIT-R BS.1534, mais contrairement à la Recommandation UIT-R BS.1534, elle étend l'évaluation des systèmes pour intégrer des conditions sans références ni repères cachés connus.

Mots clés

Test d'écoute, qualité audio, systèmes sonores évolués, évaluation subjective, évaluation de la perception

L'Assemblée des radiocommunications de l'UIT,

considérant

- a) que de nombreuses méthodes d'évaluation subjective existent dans des Recommandations UIT-R et UIT-T pour l'évaluation subjective de la qualité pour les systèmes audio, vidéo et vocaux;
- b) que l'utilisation de méthodes normalisées de test subjectif est importante pour l'échange, la compatibilité et l'évaluation correcte des résultats des tests;
- c) que l'utilisation de méthodes normalisées de test est demandée pour l'évaluation des systèmes sonores évolués;
- d) que dans certaines applications, aucun signal de référence n'est disponible ni adapté, de sorte que l'évaluation subjective de la qualité des systèmes sonores ne peut pas être effectuée relativement à un signal connu et, au lieu de cela, doit être effectuée sans référence;
- e) que le processus de production de programmes nécessite d'utiliser des systèmes technologiques pour créer des signaux audio et pour exprimer une intention créative, et qu'alors, il peut exister des conditions dans lesquelles aucun signal de référence cible ou comportement de système n'est disponible;
- f) que l'introduction de systèmes sonores évolués, tels que décrits dans la Recommandation UIT-R BS.2051, offre de nouveaux outils pour l'expression créative en termes de production, et nécessite de nouvelles méthodes d'évaluation subjective, y compris des méthodes pour l'association des attributs de perception à la qualité audio globale perçue,

recommande

que les procédures de test et d'évaluation indiquées dans l'Annexe 1 de la présente Recommandation soient utilisées pour l'évaluation subjective des différences audibles pour les systèmes audio, lorsqu'aucun signal ou système de référence approprié n'est disponible.

Annexe 1

1 Introduction

Les essais d'écoute subjectifs sont un moyen fiable de mesurer la qualité de perception des systèmes audio. Il existe des méthodes éprouvées et bien décrites d'évaluation de la qualité audio dans un contexte de radiodiffusion lorsque les systèmes sont comparés à une référence connue non altérée, à des niveaux de qualité à la fois élevée et intermédiaire. La Recommandation UIT-R BS.1116 – Méthodes d'évaluation subjective des dégradations faibles dans les systèmes audio y compris les systèmes sonores multivoies – a pour but d'évaluer les systèmes audio de haute qualité avec des dégradations faibles par rapport à un signal de référence donné, et la Recommandation UIT-R BS.1534 – Méthode d'évaluation subjective du niveau de qualité intermédiaire des systèmes audio – vise à évaluer les systèmes audio à un niveau intermédiaire, adapté aux applications de radiodiffusion, mais clairement différents d'un signal de référence. Il convient aussi de faire remarquer que l'élaboration de ces deux méthodes a été en grande partie motivée par le besoin d'évaluer les effets de systèmes de codage audio à faible débit binaire.

Dans certaines applications, il n'existe aucun signal de référence disponible ou approprié; l'évaluation de la qualité subjective des systèmes ne peut donc être effectuée en termes de fidélité par rapport à une référence. La Recommandation UIT-R BS.1284 – Méthodes générales d'évaluation subjective de la qualité du son – ne contient que des méthodes consacrées à la gamme audio de haute qualité, ou qui ne fournissent pas de barème absolu pour la qualité audio.

La présente Recommandation décrit une méthode utilisant plusieurs stimuli pour l'évaluation subjective de la qualité en termes de différences audibles pour les systèmes audio dans des applications où aucune référence connue n'est disponible.

La méthode se base sur l'approche consistant à présenter plusieurs stimuli, employée dans la Recommandation UIT-R BS.1534, pour permettre la comparaison avec les systèmes sonores à l'étude. Il est demandé à l'estimateur d'évaluer chaque système à l'étude en termes de:

- 1) Qualité sonore subjective globale.
- 2) Notes attribuées aux attributs (ensembles prédéfinis d'attributs choisis).

Les évaluations de la qualité sonore subjective globale sont effectuées au moyen de l'échelle de qualité continue définie dans la Recommandation UIT-R BS.1534.

Dans la mesure où l'étape d'évaluation des attributs, décrite dans les Recommandations UIT-R BS.1116, UIT-R BS.1284, UIT-T P.835 et UIT-T P.806, est facultative, plusieurs attributs pertinents en matière de qualité sonore, tirés de préférence de lexiques existants et validés, sont sélectionnés en amont de chaque expérience. Les estimateurs notent ces attributs sur des échelles linéaires à 100 niveaux.

L'analyse statistique de ces deux types de données permet de déduire les points suivants:

- la qualité sonore subjective globale relative de chaque système sonore;
- éventuellement, les caractéristiques de perception d'attributs sélectionnés de chaque système sonore;
- éventuellement, le poids relatif des différentes caractéristiques de perception par rapport à la qualité perçue des systèmes à l'étude.

2 Terminologie

Qualité subjective globale – attribut rendant à lui seul tous les aspects de la qualité sonore évaluée, en l'occurrence la «qualité audio de base» telle que définie dans la Recommandation UIT-R BS.1284.

Le terme de «qualité subjective globale» est utilisé ici pour éviter toute confusion potentielle avec la «qualité audio de base» définie dans la Recommandation UIT-R BS.1116.

La principale différence entre la «qualité audio de base» et la «qualité subjective globale» est la différence de processus d'évaluation de la qualité, et non les différents attributs perçus résumés sous ces deux termes généraux liés à la qualité. En ce qui concerne la «qualité audio de base», l'évaluation est effectuée en comparant deux stimuli ou plus, l'un d'eux étant pris pour référence (par exemple, juger la différence, sur le plan quantitatif, entre la version compressée d'un élément audio et l'original non compressé). En revanche, la «qualité subjective globale» correspond au jugement quantitatif comparé seulement à une référence interne, en l'occurrence l'attente de l'auditeur, sans référence extérieure connue (par exemple différentes reproductions binaurales).

Variables contrôlées – Variables contrôlées au sein de l'expérience, permettant de concevoir l'expérience de façon structurée et contrôlée. Également appelées variables indépendantes, car leur valeur est indépendante de celle des autres variables de l'expérience.

Variables de réponse – Variables pour lesquelles les estimateurs fournissent une réponse, en notant le stimulus perçu sur une échelle donnée. Également appelées variables dépendantes, car leur valeur dépend d'autres variables de l'expérience, en l'occurrence les variables indépendantes ou contrôlées.

Condition – Ensemble de valeurs des variables contrôlées utilisé dans le cadre de l'évaluation.

Essai – Étape du processus d'évaluation pendant laquelle les systèmes à l'étude (ou un sous-ensemble de ces systèmes) sont présentés dans certaines conditions, et notés par l'estimateur sous forme de variables de réponse.

Reproduction – Condition d'essai reproduite dans laquelle la ou les mêmes variables de réponse sont notées par un estimateur individuel avec les mêmes valeurs de variables contrôlées.

Descriptif – Décivant d'une façon objective et sans porter de jugement.

Attribut – Caractéristique spécifique de la qualité perçue pouvant être évaluée au moyen d'une échelle de notation. La qualité subjective globale perçue peut découler de l'évaluation de multiples attributs.

Élément de programme – Morceau de matériel audio utilisé dans le cadre de l'évaluation, associé à d'autres variables contrôlées.

Stimulus – Présentation individuelle d'un élément de programme au moyen d'un système répondant à un ensemble de valeurs de variables contrôlées.

Lexique – Ensemble d'attributs de perception descriptifs associés à des noms, définitions et échelles de notation précis.

3 Pratiques générales

Si l'on veut rassembler des informations fiables dans un domaine d'intérêt scientifique, on dispose d'un grand nombre de stratégies de recherche différentes. Pour l'évaluation subjective de la qualité des systèmes audio, il faut recourir à des méthodes expérimentales strictes, afin de garantir la fiabilité des résultats et de leur interprétation. La collecte de données fiables à partir d'expériences subjectives exige la maîtrise des conditions expérimentales, de sorte que lorsqu'elle est présentée aux estimateurs de façon contrôlée, l'expérience produise des données de haute qualité. La conception et la planification des expériences doivent être soignées afin que les facteurs non maîtrisés, qui peuvent créer du bruit de fond non désiré dans l'expérience, soient minimisés. Par exemple, si toutes les conditions d'une expérience sont présentées à tous les estimateurs dans un ordre fixe et identique, cela peut entraîner un effet de biais lié à l'ordre de présentation qui ne pourra être supprimé des données ni de leur interprétation. Une meilleure pratique à cet égard consiste à veiller à ce que les conditions soient présentées à chaque estimateur dans un ordre aléatoire, ou bien au moyen d'un plan équilibré,

afin de minimiser tout effet de biais potentiel dû à l'ordre. La procédure de test recommandée est présentée en détail au § 5.

Afin de garantir la meilleure qualité des données dans de telles expériences, il est important de tenir compte de certains des détails suivants, qui font partie de la présente Recommandation.

Il convient d'avoir recours à des estimateurs expérimentés, car ils produisent généralement des données de haute qualité. Les estimateurs expérimentés sont sélectionnés et contrôlés selon la description fournie dans la Recommandation UIT-R BS.1534. Afin d'être en mesure d'évaluer efficacement la performance des systèmes à l'étude, il est aussi important de recourir à des éléments de programme critiques, et de choisir les attributs de perception qui différencient le mieux les systèmes à l'étude et contribuent à la qualité perçue de l'expérience.

En ce qui concerne la conception des expériences, l'expérimentateur doit planifier attentivement la durée de l'expérience. Le fait d'inclure un nombre et une gamme suffisants d'éléments de test critiques donnera un aperçu plus généralisable de la performance des systèmes sonores à l'étude. Il peut également être souhaitable de comparer de nombreux systèmes sonores. De tels objectifs sont courants, mais ils s'accompagnent également d'une pénalité en termes de temps et de coût, en plus du risque de surcharger les estimateurs. La Pièce jointe 1 de l'Annexe 1 (pour information) de la présente Recommandation contient des méthodes visant à faciliter la planification des ressources (y compris une estimation de la durée de l'essai).

Afin de pouvoir poursuivre ou reproduire fidèlement les expériences ultérieurement ou ailleurs, le rapport de test ne devrait pas seulement contenir les résultats, mais tous les détails liés à l'expérience. Les lignes directrices pour l'établissement de rapports sont décrites dans les Recommandations UIT-R BS.1116 et UIT-R BS.1534.

4 Paramètres expérimentaux

Cette partie propose des définitions des paramètres expérimentaux clés, pour permettre de concevoir des expériences contrôlées de façon structurée. Ces paramètres sont répartis en deux catégories principales, à savoir les variables expérimentales contrôlées et les variables de réponse.

4.1 Variables expérimentales contrôlées

Les variables expérimentales contrôlées (ou variables indépendantes) sont utilisées pour définir les paramètres contrôlés au sein de l'expérience, permettant ainsi de concevoir l'expérience de façon structurée et contrôlée, ce qui mènera à une analyse statistique approfondie. En règle générale, on définit des variables contrôlées pour des paramètres tels que les systèmes à l'étude, les séquences de test, les estimateurs et les reproductions. Pour chaque variable contrôlée, le nombre de niveaux doit être défini par l'expérimentateur. Par exemple, un essai comportant dix éléments de programme différents revient à avoir dix niveaux de variables d'éléments de programme. Le nombre de niveaux est ensuite utilisé dans la conception de l'expérience et dans l'analyse statistique qui en découle.

4.1.1 Systèmes à l'étude

Dans de telles expériences, l'expérimentateur cherche à étudier la qualité de perception de la technologie ou du système à l'étude.

D'après la Loi de Miller (Miller, G.A., 1956), le nombre de systèmes à l'étude devrait se situer entre cinq et neuf, afin de minimiser les erreurs dans la notation de l'estimateur. Dans le cas où l'on souhaite étudier plus de neuf systèmes sonores, des orientations supplémentaires figurent au § 5.1.

Lorsque cela est possible, l'expérimentateur devrait inclure un ou plusieurs systèmes dont la qualité est bien connue afin de pouvoir considérer les résultats relatifs aux systèmes à l'étude en contexte.

4.1.2 Séquences de test

Le choix des séquences de test devrait suivre les procédures indiquées dans les Recommandations UIT-R BS.1116 et UIT-R BS.1534. Bien qu'aucune séquence de programme ne soit universellement adaptée et susceptible d'être utilisée pour évaluer tous les systèmes dans toutes les conditions, les séquences de test décisives doivent faire l'objet d'une recherche explicite. La recherche de séquences adaptées prend en général beaucoup de temps. Cependant, à moins de trouver des séquences réellement décisives pour chaque système, les expériences échoueront à faire apparaître les différences entre les systèmes, et ne permettront pas de dégager de conclusions.

4.1.3 Estimateurs

Il est recommandé de faire appel à des estimateurs expérimentés afin de garantir la qualité des données recueillies. Ces estimateurs doivent être compétents en matière d'écoute critique du son. Des résultats plus fiables seront ainsi obtenus plus rapidement qu'avec des estimateurs non expérimentés. Il est également important de noter que la plupart des estimateurs non expérimentés acquièrent en général une plus grande sensibilité aux différents types d'artéfacts après les avoir fréquemment rencontrés. Un estimateur expérimenté est choisi pour son aptitude à procéder à un test d'écoute. Cette aptitude doit être qualifiée et quantifiée par rapport à ses compétences, pour ce qui est de la fiabilité et du discernement dont il fait preuve lors de la répétition des évaluations. Ceux-ci sont définis comme suit:

- 1) Le discernement: aptitude à percevoir des différences entre les éléments testés.
- 2) La fiabilité: fait pour les évaluations répétées du même élément testé d'être proches les unes des autres.

Seuls les estimateurs classés parmi les estimateurs expérimentés pour un test donné doivent participer à l'analyse finale des données (voir la Recommandation UIT-R BS.1116). Ces techniques, qui sont fondées sur la répétition par chaque estimateur d'au moins une évaluation, permettent de qualifier et de quantifier sa compétence dans le cadre d'une expérience. Ces méthodes doivent être appliquées soit à la présélection des estimateurs lors d'une expérience pilote, soit de préférence à la présélection et au cours du test lui-même (au moyen des réponses obtenues dans le cadre de l'évaluation principale). Une expérience pilote est souvent une évaluation de plus faible ampleur associée à une expérience principale. Elle comporte un ensemble représentatif d'échantillons de test à évaluer dans le cadre de l'expérience principale. Aux fins de l'évaluation de la compétence de l'auditeur, l'expérience pilote doit comprendre un sous-ensemble pertinent de stimuli de test, représentatif de la gamme entière des stimuli et des artéfacts à évaluer au cours de l'expérience principale.

Une vingtaine (ou un nombre supérieur, de préférence) d'estimateurs expérimentés doivent être employés nominalement.

4.1.4 Reproduction

Un moyen d'évaluer la qualité des données collectées et la performance des estimateurs consiste à demander à chaque estimateur de formuler plusieurs jugements successifs pour une condition. On suggère de reproduire au moins deux échantillons pour vérifier la performance de l'estimateur (voir partie précédente) sans augmenter excessivement l'ampleur d'une expérience.

4.1.5 Variables expérimentales contrôlées supplémentaires

Il peut arriver que l'expérience nécessite des variables expérimentales contrôlées supplémentaires. C'est assez normal et acceptable, et ces variables peuvent être ajoutées d'une façon structurée et similaire à celles exposées dans le paragraphe 4.1. L'expérimentateur devrait garder à l'esprit que le fait d'augmenter le nombre de variables contrôlées accroîtra la taille et la durée de l'expérience.

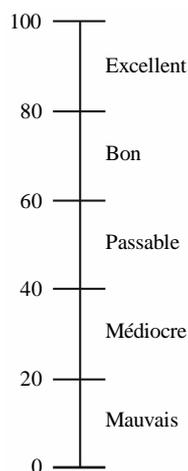
4.1.6 Variables de réponse

Pour chaque condition, il est demandé aux estimateurs de donner leur évaluation au moyen de variables de réponse. Deux types différents de variables de réponses doivent être utilisés; ils sont décrits ci-dessous, ainsi que leur dimensionnalité:

- Qualité subjective globale (par système).
- Notes des attributs (facultatif, par système).

4.1.7 Qualité sonore subjective globale

FIGURE 1
Échelle de qualité continue



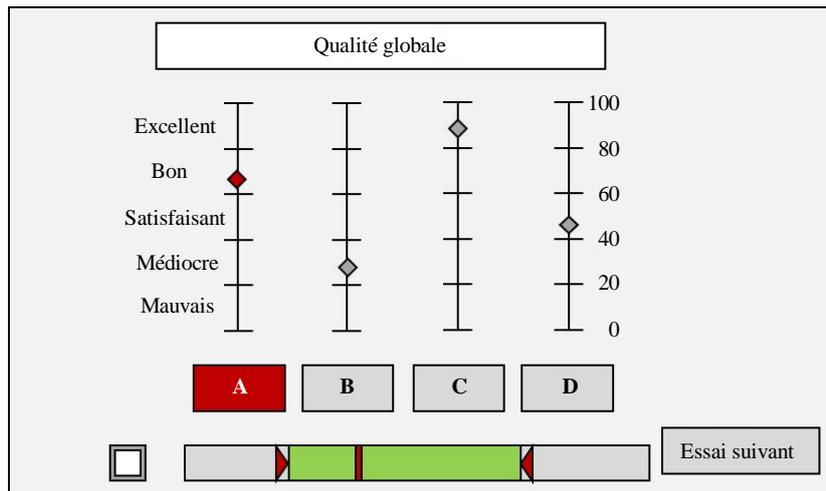
BS.2132-01

Au départ, il est demandé aux estimateurs d'évaluer la qualité sonore subjective globale au moyen de l'échelle de qualité continue¹. Il est demandé aux estimateurs d'évaluer la qualité sonore subjective globale de chaque présentation et de fournir leur évaluation au moyen de l'échelle de qualité continue. Cette échelle est une échelle linéaire à 100 niveaux (généralement >10 cm), qui est divisée en cinq intervalles égaux, délimités par les adjectifs montrés dans la Fig. 1. De multiples systèmes sont présentés dans un seul essai avec un élément de programme commun, chacun doté de sa propre échelle de notation, comme illustré dans la Fig. 2.

¹ Cette échelle est également utilisée pour l'évaluation de la qualité de l'image (Recommandation UIT-R BT.500 – Méthodologie d'évaluation subjective de la qualité des images de télévision et Recommandation UIT-R BS.1534 – Méthode d'évaluation subjective du niveau de qualité intermédiaire des systèmes audio).

FIGURE 2

Exemple d'interface graphique utilisateur pour l'évaluation de la qualité sonore subjective globale



BS.2132-02

4.1.8 Notation des attributs

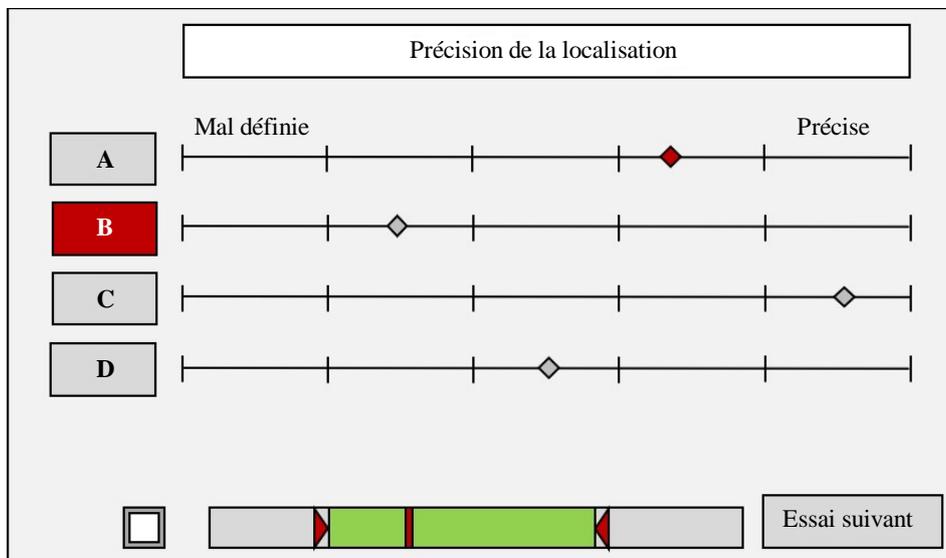
Une fois que les estimateurs ont noté la qualité sonore subjective globale, il leur est également demandé de noter des associations de systèmes et d'éléments de programme au vu de chaque attribut descriptif. Là aussi, de multiples systèmes sont présentés dans un seul essai avec un élément de programme commun, chacun disposant de sa propre échelle de notation.

Les échelles relatives aux attributs sont généralement des échelles continues à 100 niveaux.

La Figure 3 illustre un exemple d'interface de notation des attributs, où l'attribut concerné et sa définition sont la précision de la localisation.

FIGURE 3

Exemple d'interface graphique utilisateur pour la notation d'un attribut (*précision de la localisation*)



BS.2132-03

5 Protocole d'essai

5.1 Plan d'expérience

L'expérience devrait être conçue avec attention pour s'assurer qu'elle produise des données de haute qualité, tout en minimisant les sources d'effets aléatoires ou non contrôlés. Le plan est aussi utile pour estimer l'ampleur et la durée de l'expérience, ainsi que pour fournir la structure de l'analyse statistique. Le plan regroupe deux aspects clés, à savoir la conception du traitement et le plan relatif à l'attribution des stimuli, tel que décrit ci-dessous.

5.1.1 Conception du traitement

La conception du traitement précise quelles variables contrôlées doivent être utilisées au sein de l'expérience, à l'exception de la variable de l'estimateur.

Pour les expériences de taille modérée, il est recommandé d'établir un plan factoriel complet, qui évalue toutes les combinaisons possibles de niveaux de variables contrôlées. Pour une expérience factorielle complète, le nombre de conditions est obtenu en multipliant le nombre de niveaux au sein de chaque variable indépendante.

5.1.2 Plan relatif à l'attribution du stimulus

Le plan d'attribution du stimulus définit comment les conditions doivent être présentées à chaque estimateur.

Un plan «intra-individuel» est recommandé, dans lequel toutes les conditions sont présentées à chaque estimateur. L'ordre de présentation est contrôlé – généralement par randomisation – afin de limiter les effets de biais systématiques. Un ordre de présentation complètement équilibré est souhaitable.

5.1.3 Sous-division de grandes expériences

Dans certaines situations, l'expérience peut devenir trop grande et compliquée avec un plan factoriel intra-individuel complet. De telles situations peuvent se produire lorsque de nombreux systèmes sonores doivent être évalués ou lorsque la durée globale de l'essai par estimateur devient déraisonnablement longue.

On peut alors envisager un plan d'expérience plus avancé.

Cette partie illustre seulement des approches à envisager dans de tels cas. Cependant, l'expérimentateur intéressé devrait consulter la littérature existante sur la conception d'expérience afin d'obtenir les meilleures orientations possible.

Les deux solutions présentées ci-dessous constituent deux exemples pour aborder des expériences de grande ampleur.

5.1.3.1 Plan en blocs subdivisés

Au paragraphe 4.1.1, il est recommandé de limiter le nombre de systèmes à l'étude entre cinq et neuf. Lorsqu'on doit évaluer de nombreux systèmes sonores, un plan en blocs subdivisés peut être envisagé. Par exemple, si l'on doit évaluer 14 systèmes sonores, l'évaluation globale pourrait être divisée en deux essais portant chacun sur sept systèmes sonores. Afin de contrôler tout effet de biais lié à la présentation par bloc, l'attribution des systèmes sonores à chacun des essais devrait être randomisée. Cependant, cette randomisation ne devrait pas affecter le nombre total de conditions présentées à chaque estimateur ou dans le cadre de l'expérience globale (voir le § 5.1.1).

Lorsque l'on utilise un plan en blocs subdivisés, il est important que les facteurs liés à la division en blocs soient intégrés à l'analyse. Dans l'idéal, la division en blocs ne devrait pas être un facteur significatif sur le plan statistique.

5.1.3.2 Plan intersujets

Un nombre élevé de conditions à évaluer par estimateur peut entraîner de trop nombreuses séances d'écoute, surtout lorsque la durée totale de l'essai excède quatre heures par estimateur.

Pour aborder ce problème, une approche envisageable consiste à recourir à un plan intersujets (ou intergroupes). Il s'agit d'une méthode globale dans laquelle différentes conditions sont présentées à différents estimateurs ou groupes d'estimateurs. Un moyen simple de limiter le nombre de conditions présentées à chaque estimateur (ou groupe d'estimateurs) serait de leur attribuer à chacun un sous-ensemble différent d'éléments de programme. Cela doit être fait avec précaution, afin de veiller à ce que le nombre global de conditions présenté soit équilibré entre chaque estimateur (ou groupe d'estimateurs).

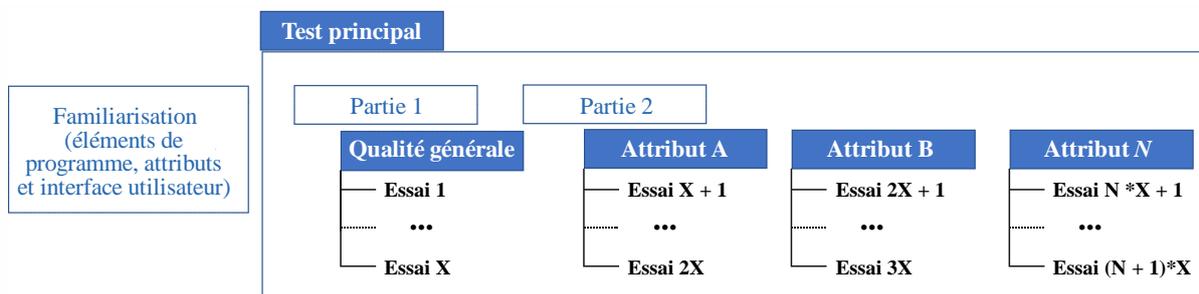
Lorsque l'on recourt à de tels plans, il est important que l'analyse de l'essai tienne compte des facteurs liés au regroupement. Dans l'idéal, le regroupement ne devrait pas être un facteur significatif sur le plan statistique.

5.2 Organisation de l'essai

L'essai doit être organisé en deux étapes, avec en outre une étape de familiarisation qui précède l'essai réel lorsque les attributs sont utilisés. L'ordre de présentation des différents éléments aux estimateurs est illustré dans la Fig. 4. Dans la figure, on part du principe qu'un nombre N d'attributs seront évalués, au moyen de M éléments de programme. Un total de x essais sera effectué pour évaluer la qualité sonore subjective globale, et de y essais pour la notation des attributs.

FIGURE 4

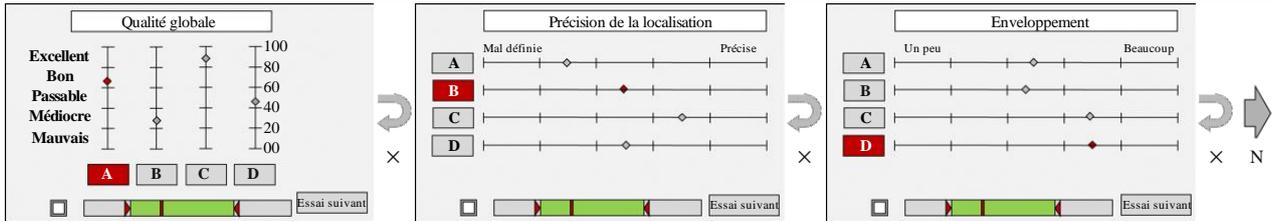
Diagramme de flux général de la structure globale de l'essai, y compris l'étape de familiarisation et l'essai principal



BS.2132-04

Afin de s'assurer de recueillir des données de haute qualité, les estimateurs devraient être familiarisés avec le protocole d'essai, les interfaces utilisateur, les éléments de programme, ainsi qu'avec les attributs de perception. Chaque estimateur doit aussi avoir la possibilité d'écouter les stimuli d'essai, d'examiner les attributs et d'essayer l'interface utilisateur. L'étape de familiarisation peut comprendre l'écoute libre et à l'aveugle des stimuli avec un sous-ensemble de l'essai aux seules fins de la familiarisation. La Figure 5 illustre le processus privilégié décrit ci-dessus. L'estimateur évalue des stimuli individuels pour un attribut à l'aide d'interfaces individuelles dans un essai.

FIGURE 5
Processus d'évaluation de stimuli individuels à l'aide d'interfaces individuelles



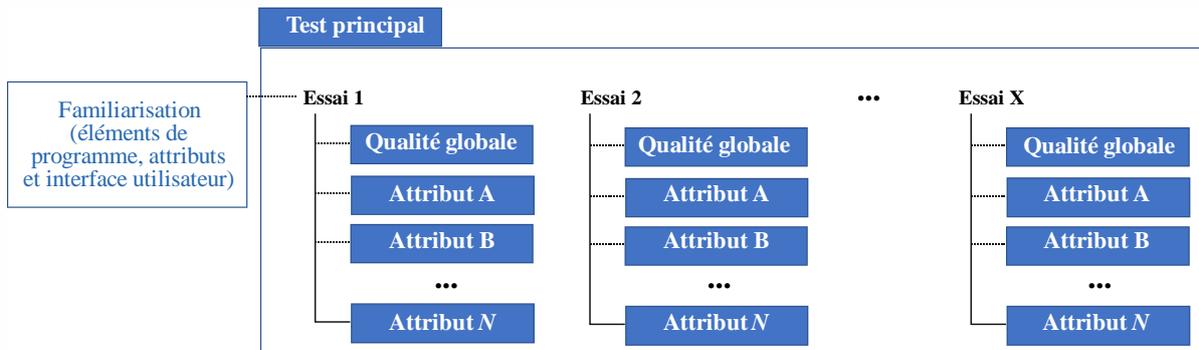
BS.2132-05

5.2.1 Procédure facultative

La Figure 6 illustre un flux facultatif de l'essai global. On peut présenter à l'estimateur des interfaces individuelles pour chaque attribut, ou une interface associant de multiples attributs. La Figure 7 montre un exemple d'interface utilisateur graphique dans laquelle un estimateur évalue chaque stimulus à l'aide d'interfaces individuelles pour chaque attribut. La Figure 8 montre un exemple d'interface utilisateur graphique dans laquelle un estimateur évalue de multiples attributs dans une même interface.

FIGURE 6

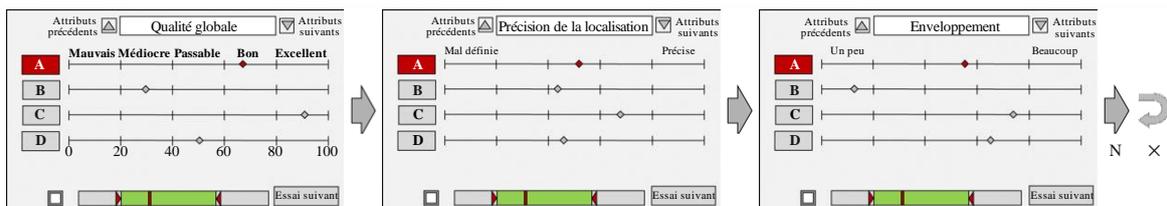
Diagramme de flux facultatif de la structure globale de l'essai, y compris l'étape de familiarisation et l'essai principal



BS.2132-06

FIGURE 7

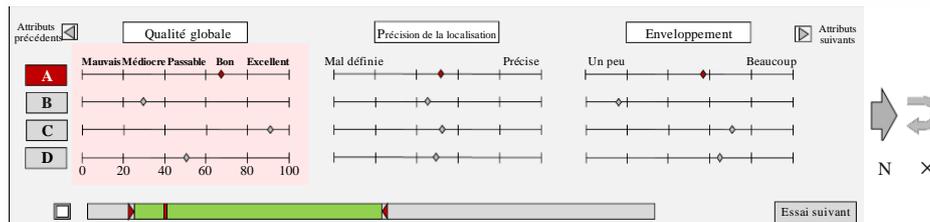
Processus d'évaluation de chaque attribut pour les mêmes stimuli à l'aide d'interfaces individuelles



BS.2132-07

FIGURE 8

Processus d'évaluation de multiples attributs pour les mêmes stimuli à l'aide d'une même interface



BS.2132-08

5.3 Instructions destinées aux estimateurs

Afin de veiller à ce que les estimateurs soient bien informés de la tâche à réaliser, il convient de leur fournir des instructions à la fois écrites et verbales en amont de l'expérience. Les instructions devraient leur permettre d'être bien conscients de la tâche à réaliser, sans introduire de biais excessif, et devraient leur présenter le protocole d'essai, les variables de réponse (qualité sonore subjective globale, définition des attributs) et la manière dont elles doivent être utilisées avec l'interface utilisateur graphique de l'essai. Des exemples de telles instructions sont fournis dans la Pièce jointe 2.

6 Environnement de l'essai

6.1 Environnement d'écoute

Les essais doivent être réalisés dans une salle d'écoute conforme à la Recommandation UIT-R BS.1116.

6.2 Dispositifs de reproduction

Des casques ou des haut-parleurs peuvent être utilisés dans le cadre de l'essai; cependant, il est interdit d'utiliser les deux dispositifs au sein d'une même session d'essai. Tous les estimateurs doivent utiliser le même type de transducteurs. Il convient d'utiliser des haut-parleurs ou des casques de contrôle de référence, comme précisé dans la Recommandation UIT-R BS.1116.

La configuration des haut-parleurs et les exigences qui y sont associées, ainsi que les positions d'écoute de l'estimateur devraient de préférence être configurées conformément aux exigences énoncées dans la Recommandation UIT-R BS.1116.

Les configurations des haut-parleurs devraient de préférence être sélectionnées parmi celles définies dans la Recommandation UIT-R BS.2051.

6.3 Étalonnage

Afin de garantir le caractère répétable et reproductible des résultats, il est important d'être attentif à la configuration des équipements et du matériel à utiliser dans l'essai.

Intensité sonore relative des éléments

L'intensité sonore relative de différents éléments ne devrait être liée d'aucune façon aux mesures d'intensité sonore précisées dans la Recommandation UIT-R BS.1770. De courts extraits audio doivent être ajustés à l'intensité sonore souhaitée. La différence entre un élément fort (fortissimo) et un élément faible (pianissimo), d'une intensité de 15 dB par exemple, doit être préservée. L'intensité sonore relative de chaque élément doit être ajustée de façon subjective, de préférence par un groupe

d'estimateurs expérimentés. Afin de maintenir cette différence dans le niveau de reproduction, il est important de tester les différents éléments de façon adéquate.

Intensité sonore relative des stimuli

De faibles différences d'intensité sonore ont tendance à introduire un biais en faveur du stimulus le plus fort. De telles différences doivent être supprimées entre les différents stimuli d'un élément. On doit recourir à la mesure de l'intensité sonore objective (plutôt que subjective) précisée dans la Recommandation UIT-R BS.1770. S'il n'est pas possible d'utiliser une mesure objective, l'intensité sonore de chaque extrait doit être ajustée de façon subjective par un groupe d'estimateurs expérimentés, avant son intégration dans un essai.

Synchronisation d'éléments

La commutation instantanée entre des stimuli consistant en des versions traitées différemment du même élément de programme ne devrait pas entraîner de décalage temporel perceptible. Pour plus d'informations sur la présentation des stimuli, veuillez-vous reporter à la Recommandation UIT-R BS.1534.

6.3.1 Étalonnage du système de reproduction

Pour les essais menés au moyen de haut-parleurs, la configuration des haut-parleurs utilisée devrait de préférence être l'une de celles décrites dans les Recommandations UIT-R BS.775 et UIT-R BS.2051. Autrement, la notation définie dans la Recommandation appropriée devrait être utilisée pour décrire la configuration utilisée dans le cadre de l'essai.

Le niveau du système de reproduction doit être ajusté conformément à la description qui figure dans la Recommandation UIT-R BS.1116.

Les détails relatifs à l'étalonnage des haut-parleurs d'extrêmes graves et des systèmes de gestion des graves dépassent le champ d'application du présent document. La gestion des graves devrait permettre que la réponse en fréquence de l'association d'un haut-parleur individuel et d'un haut-parleur d'extrêmes graves soit uniforme (dans les limites de la tolérance précisée dans la Recommandation UIT-R BS.1116).

Au cours des séquences d'essai précédentes, on a noté que chaque auditeur peut préférer des niveaux absolus d'écoute différents. Bien que ce ne soit pas la meilleure solution, on ne peut pas toujours empêcher les participants de demander autant de souplesse. On ne sait pas encore si cela aura une influence sur l'audibilité de certains des défauts étudiés. Par conséquent, si les participants eux-mêmes règlent le gain du système, il faudra le mentionner dans le résultat des essais.

La différence de temps de propagation entre les voies d'un système stéréophonique ne doit pas dépasser 20 μ s pour les casques, ou 100 μ s pour les haut-parleurs, comme indiqué dans la Recommandation UIT-R BS.1116.

Dans le cas d'un système avec image associée, le temps de propagation global du casque ou du haut-parleur de contrôle de référence par rapport au ou aux systèmes étudiés ne doit pas dépasser les limites imposées dans la Recommandation UIT-R BS.775.

NOTE – La mesure des paramètres acoustiques des systèmes sonores évolués peut être beaucoup plus complexe que celle des systèmes audio multicanaux antérieurs. Il faut prêter une attention particulière au choix du microphone de mesure et à son orientation lors des mesures (voir le Rapport UIT-R BS.2419). La Recommandation UIT-R BS.1116 donne de plus amples informations sur les exigences électroacoustiques et sur les caractéristiques de réponse de la salle de contrôle.

7 Analyse statistique

L'analyse statistique des données d'évaluation de l'estimateur est réalisée pour donner des informations sur la qualité subjective des systèmes à l'étude. Les évaluations moyennes sont calculées pour donner des indications de cette performance, et on utilise une estimation de la variance des données pour indiquer la fiabilité des différences observées en matière de performance des systèmes.

Lorsque les données sont recueillies, chaque estimateur fournit les notes associées aux attributs des systèmes à l'étude. Ces systèmes sont testés avec différents éléments de programme. L'estimateur évalue chaque système sur une liste d'attributs de perception au moyen d'échelles prédéfinies. Pour chaque élément de programme, chaque estimateur note chaque attribut sur le même ensemble d'échelles d'attribut. L'estimateur évalue aussi chaque association de système et d'élément de programme en termes de qualité subjective globale.

Pour chaque élément de programme, les estimateurs doivent fournir leur évaluation des attributs de chaque système, ainsi que leur évaluation de la qualité sonore subjective globale.

La Recommandation UIT-R BS.1534 donne des précisions concernant l'analyse statistique des données relatives à la qualité sonore subjective globale et des données relatives à chaque attribut descriptif.

En outre, les informations obtenues par l'analyse des données sensorielles découlant de l'utilisation de la présente Recommandation sont comparables à celles obtenues par l'analyse des données sensorielles découlant d'une analyse descriptive quantitative plus classique. De telles analyses sont par exemple l'analyse de la variance (ANOVA) réalisée pour chaque attribut sensoriel, ainsi que l'analyse à plusieurs variables (telle que l'utilisation de l'analyse en composantes principales, ACP).

8 Références

- [1] Recommandation UIT-R BS.2051 – Système sonore évolué pour la production de programmes
- [2] Recommandation UIT-R BS.775 – Système de son stéréophonique multicanal avec ou sans image associée
- [3] Recommandation UIT-R BS.645 – Signaux d'essai et instruments de mesure pour les liaisons radiophoniques internationales
- [4] Recommandation UIT-R BS.1116 – Méthodes d'évaluation subjective des dégradations faibles dans les systèmes audio
- [5] Recommandation UIT-R BS.1534 – Méthode d'évaluation subjective du niveau de qualité intermédiaire des systèmes audio
- [6] Recommandation UIT-R BS.1770 – Algorithmes de mesure de l'intensité sonore des programmes audio et des niveaux de crête vrais des signaux audio
- [7] Recommandation UIT-R BS.1864 – Pratiques d'exploitation concernant l'intensité sonore dans l'échange international de programmes de télévision numérique
- [8] Miller, G.A. (1956), The magical number seven, plus or minus two: Some limits on our capacity for processing information [Le nombre magique sept plus ou moins deux: sur quelques limites de notre capacité à traiter l'information]. *Psychological Review*. 63 (2): 81-9
- [9] Recommandation UIT-R BT.500-13 – Méthodologie d'évaluation subjective de la qualité des images de télévision
- [10] Recommandation UIT-T P.835 – Méthodologie d'évaluation subjective des systèmes de communication vocale avec algorithmes de suppression du bruit
- [11] Recommandation UIT-T P.806 – Méthode d'évaluation subjective de la qualité à l'aide d'échelles de notation multiples

- [12] Rapport UIT-R BS.2419 – Effet de la directivité du microphone sur l'étalonnage du niveau et l'égalisation des systèmes sonores évolués

Pièce jointe 1 à l'Annexe 1 (pour information)

Outil Excel d'estimation de la durée d'une expérience

L'outil Excel fourni ci-dessous sert à estimer la durée d'une expérience d'évaluation audio subjective, à des fins de planification des ressources. Le Tableau 1 montre un exemple d'expérience. Les champs de couleur beige sont les champs d'entrée. Les champs vert clair sont les champs de sortie.



Copy of
R15-WP6C-190715-T

TABLEAU 1

Capture d'écran du tableau Excel fourni, servant à estimer la durée d'une expérience d'évaluation audio subjective

Outil de calcul de plan factoriel complet

v4

Variables expérimentales contrôlées (variables indépendantes)					
Variable	Nom	Description	Nombre de niveaux	Calcul nombre de niveaux	Degrés de liberté
1	SYSTEM	Nombre de systèmes à l'étude	7	7	6
2	PROGRAMME	Nombre d'éléments de programme	3	3	2
3	REPLICATE	Nombre de présentations ou de répétitions	1	1	0
4	ASSESSOR	Nombre d'estimateurs	20	20	19
Total				31	27

Variables de réponse (variables dépendantes)			
Variable	Nom	Description	Nombre de niveaux
Total		Qualité subjective globale	1
Descriptif		Enveloppement, largeur de la source, etc.	6

Champs d'entrée

Champs de sortie

Paramètres d'essai			
Paramètre	Nombre	Unité	Commentaires
Nombre de conditions expérimentales contrôlées (total)	21		Compte non tenu des estimateurs
Nombre de conditions d'essai (par reproduction)	21		Compte non tenu des estimateurs
Nombre total de variables indépendantes	420		Compte tenu des estimateurs
Nombre de blocs	1		1 = plan intra-sujets >1 = plan inter-sujets
Nombre d'éléments de programme par bloc	3		Doit être un nombre entier ≥2
Nombre d'évaluations par condition	20		
Nombre total d'évaluations par estimateur	147		
Estimation du temps d'évaluation moyen par estimateur	20	seconde	
Estimation de la durée totale par estimateur	0.8	heure	
Durée de la session	2	heure	2 heures maximum, pauses comprises
Estimation du nombre de sessions par évaluateur	1	session	
Nombre total de sessions	20	session	
Nombre de points de données par variable de réponse	420		
Nombre total de points de données dans l'expérience	2940		

**Pièce jointe 2
à l'Annexe 1
(pour information)**

Exemple d'instructions destinées aux estimateurs

A2.1 Instructions générales

Chère auditrice, cher auditeur,

Bienvenue dans cet essai, dans lequel vous entendrez et évalueriez différents systèmes sonores ainsi qu'un ensemble d'éléments de programme. L'essai fait appel à la méthode utilisant «plusieurs stimuli sans référence connue».

Vous disposez d'un total d'environ deux heures pour procéder à l'essai, qui comporte une étape de familiarisation suivie de l'essai proprement dit. Pendant l'étape de familiarisation, les éléments de programme, les interfaces utilisateur, ainsi que les attributs utilisés pour évaluer chaque système sonore vous seront présentés. Au terme de cette étape, vous réaliserez un essai en deux parties.

La première partie correspond à l'évaluation des systèmes sonores en termes de *qualité subjective globale*.

Qualité subjective globale – Attribut rendant à lui seul tous les aspects de la qualité sonore évaluée.

La seconde partie correspond à l'évaluation de chaque système sonore en ce qui concerne les attributs suivants:

- Profondeur de la scène
- Enveloppement
- Englobement
- Précision de la localisation
- Brillance
- Distorsion

Les définitions des attributs sont fournies ci-dessous et seront expliquées en amont de l'essai.

À chaque étape de l'essai, veuillez écouter attentivement les éléments de programme, en prenant le temps nécessaire pour réfléchir à chacun d'eux et les évaluer.

N'hésitez pas à poser toute question ou à solliciter toute précision supplémentaire concernant le protocole d'essai, de préférence pendant ou après l'étape de familiarisation.

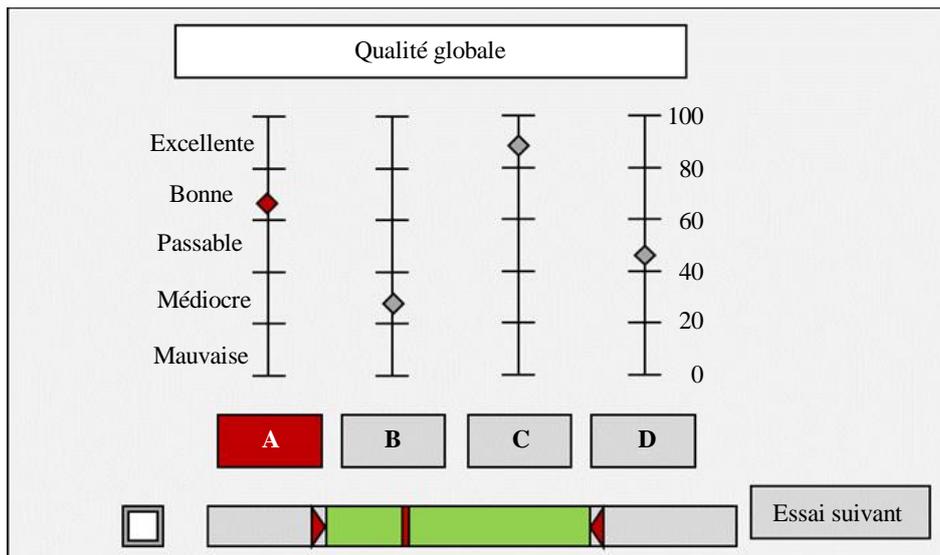
Vous êtes encouragé à faire des pauses toutes les 20 à 30 minutes pour vous dégourdir et vous reposer rapidement.

A2.1.1 Note de qualité globale

Il vous sera demandé d'évaluer la qualité globale de chaque échantillon sonore en termes de *qualité subjective globale*, à l'aide d'une échelle de qualité continue allant de 0 à 100 niveaux, comme illustré dans la Fig. 9. Veuillez écouter attentivement chaque échantillon, autant de fois que nécessaire, avant de l'évaluer lorsque vous êtes prêts. Une fois tous les échantillons évalués, cliquez sur «suivant» pour commencer l'essai suivant.

FIGURE 9

Interface utilisateur pour l'évaluation de la qualité subjective globale



BS.2132-09

A2.1.2 Attributs descriptifs

Pour chaque essai, il vous sera demandé d'évaluer la qualité sonore de chaque système sur l'un des attributs (voir le Tableau 2).

TABLEAU 2

Désignation et définition des attributs

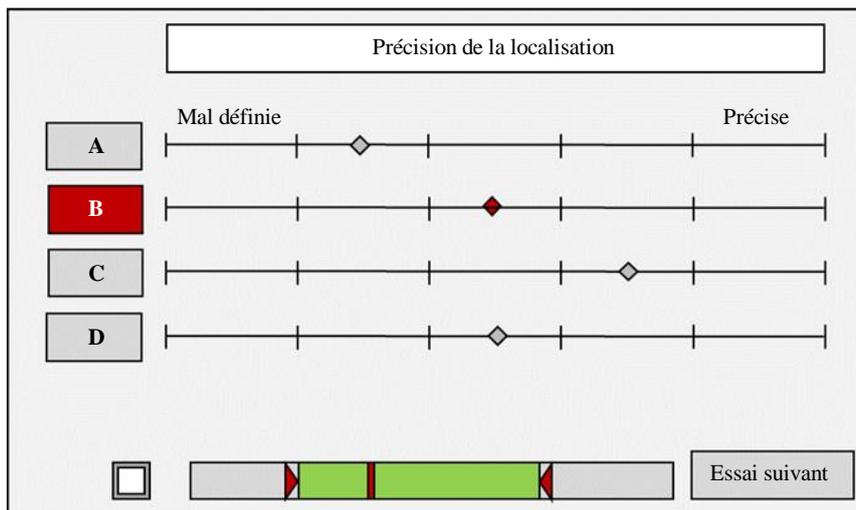
Attribut	Définition	Exemple	Désignation inférieure	Désignation supérieure
Profondeur de scène	Perception de la profondeur de l'image sonore. Tient compte à la fois de la profondeur globale de la scène et de la distance relative des différentes sources sonores.		Uniforme	Profonde
Enveloppement	Êtes-vous entourés par le son reproduit, et cela vous donne-t-il la sensation de l'espace qui vous entoure? Sensation d'être enveloppé par le son.		Un peu	Beaucoup
Englobement	Portée perçue d'une source sonore dans la direction verticale. Sensation d'être transporté, et que le son entoure ou couvre complètement.	Dans quelle mesure avez-vous l'impression que le son se situe au-dessus et en dessous de vous? Dans quelle mesure le son vous englobe-t-il verticalement?	Un peu	Beaucoup

TABLEAU 2 (*fin*)

Attribut	Définition	Exemple	Désignation inférieure	Désignation supérieure
Précision de la localisation	À quel point les différents instruments et les différentes voix sont-ils placés et séparés spatialement dans l'image sonore? À quel point chaque source sonore est-elle précisément positionnée dans la pièce? Si les sources sonores individuelles sont involontairement étendues ou élargies, la précision est faible. Les différents instruments et les différentes voix peuvent-ils être clairement placés et séparés spatialement dans l'image sonore? À quel point les sources sonores individuelles sont-elles précisément positionnées dans la pièce?		Mal définie	Précise
Brillance	Extension aiguë ou haute fréquence: <ul style="list-style-type: none"> – Un peu: comme si vous entendiez la musique à travers une porte, étouffée, brouillée ou sourde. – Beaucoup: légèreté, pureté et clarté, avec de l'espace pour les instruments. Clarté dans les hautes fréquences sans que le son soit perçant ou strident, et sans distorsion. 		Un peu	Beaucoup
Distorsion	Sons supplémentaires et non désirés qui polluent le son.	Peut apparaître sous la forme d'effets temporels ou de timbre, susceptibles de produire un son «perçant», «épillé», ou «entrecoupé», ou une tonalité temporelle, par exemple.	Un peu	Beaucoup

FIGURE 10

Interface utilisateur pour l'évaluation d'attributs descriptifs



BS.2132-10

Veillez commencer l'étape de familiarisation lorsque vous êtes prêt.

Nous vous remercions de votre participation.

Pièce jointe 3 à l'Annexe 1 (pour information)

Exemples de cas d'utilisation pour l'évaluation et la description subjectives de la qualité sonore de systèmes audio sans référence connue

L'introduction de systèmes sonores avancés, telle que décrite dans la Recommandation UIT-R BS.2051, donne des outils pour l'expression créative en production. L'évaluation de ces systèmes peut comprendre des conditions dans lesquelles on ne dispose pas de références ni de repères cachés connus. Les méthodes décrites dans la présente Recommandation sont applicables dans ces conditions.

En outre, d'autres cas d'utilisation peuvent se produire, dans lesquels les expérimentateurs peuvent bénéficier du recours à cette méthodologie pour la caractérisation de la qualité subjective de leurs systèmes et signaux.

Les exemples de cas d'utilisation applicables pour la présente méthode comprennent notamment des évaluations subjectives de la qualité des éléments suivants:

- Comportement de la restitution de production d'un système sonore avancé en l'absence de référence disponible ou appropriée à l'intention du producteur.
- Reproduction d'un programme sonore avancé sur différentes configurations de haut-parleurs par une seule restitution de production.
- Systèmes de reproduction de contenus de systèmes sonores avancés en home cinéma, lorsqu'il n'existe pas de système donnant *a priori* un objectif connu de qualité optimale.

- Algorithmes d'augmentation ou de réduction par mixage.
 - Microphones pour l'enregistrement et la production du son dans l'espace.
 - Processeurs de réverbération pour la production audio dans l'espace.
 - Systèmes et paramètres de traitement dynamique multibande pour la distribution radio.
 - Techniques de rétablissement du son.
-