

RECOMMANDATION UIT-R BS.1284*

Méthodes d'évaluation subjective de la qualité du son – Conditions générales à respecter

(Question UIT-R 85/10)

(1997)

L'Assemblée des radiocommunications de l'UIT,

considérant

- a) qu'en raison de l'introduction de nouveaux types de traitement du signal audio, tels que le codage numérique et la réduction du débit binaire, de nouveaux types de signal de télévision utilisant le multiplexage temporel des composantes et de nouveaux services tels que la télévision à définition améliorée et la télévision à haute définition (TVHD), il faudra peut-être utiliser les méthodes nouvelles ou modifiées pour l'évaluation subjective de la qualité du son;
- b) que ces techniques produisent leurs propres dégradations du signal;
- c) que les essais d'écoute subjective permettent l'évaluation du degré de gêne causée à l'auditeur par toute dégradation du signal utile au cours de sa transmission entre la source et l'auditeur;
- d) qu'il existe un grand nombre de méthodes d'évaluation subjective;
- e) qu'en particulier, d'autres Questions UIT-R appellent d'urgence l'élaboration de directives applicables à l'évaluation subjective;
- f) qu'il est hautement souhaitable de normaliser les méthodes d'évaluation subjective et l'interprétation des résultats, pour pouvoir comparer aussi valablement que possible les résultats obtenus à différentes dates et/ou en différents lieux;
- g) qu'il est hautement souhaitable que les échelles de notation employées pour décrire la qualité subjective du son permettent l'utilisation de méthodes de traitement statistique plus homogènes et indépendantes de la langue utilisée pour exprimer les opinions;
- h) qu'il serait souhaitable de disposer d'une même échelle d'évaluation du son et des programmes de télévision;
- j) que les caractéristiques géométriques et les propriétés acoustiques des salles de contrôle et des salles d'écoute peuvent avoir une influence considérable sur l'écoute, et qu'il faut par conséquent spécifier de manière précise les conditions d'écoute,

recommande

1 d'utiliser les procédures d'essai et d'évaluation données dans l'Annexe 1 pour l'évaluation subjective de la qualité du son reproduit.

* La Commission d'études 6 des radiocommunications a apporté des modifications rédactionnelles à cette Recommandation en 2002 conformément aux dispositions de la Résolution UIT-R 44.

ANNEXE 1

1 Généralités

La présente Annexe est divisée selon les paragraphes suivants donnant les conditions détaillées à respecter pour les différents aspects des essais:

- 1 Généralités
- 2 Conception des expériences
- 3 Choix des groupes d'auditeurs
- 4 Méthode d'essai
- 5 Caractéristiques
- 6 Eléments de programme
- 7 Dispositifs de reproduction
- 8 Conditions d'écoute
- 9 Traitement statistique des données
- 10 Présentation des résultats
- 11 Contenu des rapports d'essais
- 12 Références bibliographiques
- 13 Bibliographie

La présente Recommandation est destinée à servir de guide à l'évaluation générale de la qualité du son. Elle s'appuie sur la Recommandation UIT-R BS.1116 qui traite des méthodes d'évaluation subjective des dégradations faibles par les systèmes audio, y compris les systèmes sonores multivoies. Toutefois, les conditions à respecter énoncées dans la Recommandation UIT-R BS.1116 sont très strictes, étant essentiellement destinées à l'évaluation des dégradations faibles. Des évaluations de caractère plus général font intervenir habituellement des différences plus importantes et par conséquent n'exigent pas une maîtrise aussi étroite des paramètres d'essai. La Recommandation UIT-R BS.1116 contient un glossaire de termes, dont certains sont utilisés dans la présente Recommandation.

D'autres Recommandations applicables à certains cas spécifiques sont citées dans la Recommandation UIT-R BS.1283 – Evaluation subjective de la qualité du son – Guide des Recommandations existantes.

2 Conception des expériences

Pour la conception des expériences, il faudra tenir compte des indications contenues dans le § 2 de la Recommandation UIT-R BS.1116, étant donné que les dégradations étudiées ne sont pas nécessairement faibles, il n'est pas toujours indispensable d'utiliser une référence. Lorsqu'une référence est utilisée, il n'est pas indispensable qu'elle soit absolument exempte de dégradation.

En général, des compétences en matière de statistique seront nécessaires pour concevoir les expériences. En effet, il faudra déterminer le nombre d'observations nécessaires, les méthodes d'analyse statistique des données et interpréter correctement les résultats de l'analyse statistique et également vérifier la validité des hypothèses sur le modèle.

3 Choix des groupes d'auditeurs

On préfère toujours des auditeurs experts. On a fait remarquer que des auditeurs non experts pouvaient être représentatifs de la population générale et que les auditeurs experts pouvaient être beaucoup trop critiques. Cependant, à la longue, l'exposition aux défauts transforme certains auditeurs non experts en experts. Par conséquent, les tests avec des experts donnent une indication meilleure et plus rapide des résultats probables dans le long terme. En cas de doute, la relation entre l'opinion des experts et celle des non-experts devra être étudiée.

En général, un certain temps sera consacré à familiariser les auditeurs avec la procédure d'essai, les éléments de programme et l'environnement des essais.

4 Méthode d'essai

4.1 Echelles d'évaluation

Pour l'évaluation subjective de la qualité du son ou des dégradations, on recommande l'utilisation des échelles suivantes à cinq notes. La nature et l'objet des essais détermineront laquelle des deux échelles convient le mieux.

Qualité		Dégradation	
5	Excellent	5	Imperceptible
4	Bon	4	Perceptible mais non gênant
3	Moyen	3	Légèrement gênant
2	Médiocre	2	Gênant
1	Mauvais	1	Très gênant

Pour les essais comparatifs, on pourra utiliser soit une méthode utilisant l'échelle de comparaison à sept notes suivante ou une basée sur les différences numériques en utilisant les échelles ci-dessus à cinq notes. En général, ces échelles ne sont pas équivalentes et peuvent donner des résultats différents.

(Il est essentiel d'indiquer le sens voulu des comparaisons).

Comparaison	
3	Bien meilleure
2	Meilleure
1	Légèrement meilleure
0	Identique
-1	Légèrement plus mauvaise
-2	Plus mauvaise
-3	Bien plus mauvaise

NOTE 1 – Les échelles doivent être traitées comme étant continues, avec une résolution recommandée étant la première décimale.

NOTE 2 – On a montré que l'utilisation de points de repère intermédiaires prédéfinis pouvait introduire des distorsions. Il est possible d'utiliser l'échelle de chiffres sans la description des points de repère; dans ce cas, l'orientation voulue de l'échelle doit être indiquée. Cette façon de procéder peut permettre de surmonter les problèmes de traduction rencontrés lors de comparaison d'essais effectués dans des langues différentes.

Si on n'utilise pas des points de repère intermédiaires, il est indispensable que les résultats de chaque participant soient normalisés par rapport à un écart moyen type. On peut appliquer l'équation (1) pour normaliser les résultats tout en conservant l'échelle initiale:

$$Z_i = \frac{(x_i - x_{si})}{s_{si}} \cdot s_s + x_s \quad (1)$$

où:

- Z_i : résultat normalisé:
- x_i : note donnée par le participant i
- x_{si} : note moyenne du participant i pendant la séance s
- x_s : note moyenne de tous les participants pendant la séance s
- s_s : écart type pour tous les participants pendant la séance s
- s_{si} : écart type pour le participant i pendant la séance s .

4.2 Procédure d'essai

Les essais peuvent être constitués par des simples présentations, des comparaisons à deux (dont l'une peut être la référence) ou des comparaisons multiples avec ou sans référence. Les présentations peuvent être répétées en fonction des nécessités. Ces procédures d'essai doivent être utilisées avec les échelles d'évaluation spécifiées au § 4.1.

Compte tenu des limitations de la mémoire humaine à court terme, chaque extrait de programme ne doit pas durer plus de 15 à 20 s; la durée peut être très courte (quelques secondes) pour certains essais. Lorsque la séquence est un extrait musical, la phrase ne doit pas sembler interrompue. Lorsque la séquence d'essai n'est pas placée sous le contrôle du participant, il est nécessaire de donner une indication claire de la présentation en cours.

Une séance ne doit pas durer plus de 15 à 20 min sans interruption. Lorsque les séances sont consécutives, elles doivent être séparées par des périodes de repos, au moins égales à la durée d'une séance.

Le dispositif de commutation ne doit pas introduire de perturbations audibles.

Lorsque les auditeurs effectuent des tests individuellement, il est hautement souhaitable que ces auditeurs commandent la commutation entre les stimuli comme indiqué dans la Recommandation UIT-R BS.1116.

5 Caractéristiques

Le nombre de caractéristiques, ainsi que leur type, utilisées pour décrire la qualité perçue dépendront des objectifs de l'essai.

Toute caractéristique utilisée doit être clairement définie.

5.1 Qualité audio de base

La caractéristique qualité audio de base inclut tous les aspects de la qualité du son en cours d'évaluation. Elle inclut entre autres le timbre, la transparence, l'image stéréophonique, la présentation spatiale, les caractéristiques de réverbération, les échos, la distorsion harmonique, le bruit de quantification, les sons explosifs, les claquements et le bruit de fond. Pour l'évaluation des dégradations faibles, la caractéristique de qualité audio de base est définie différemment dans la Recommandation UIT-R BS.1116.

5.2 Caractéristiques spécifiant la qualité du son stéréophonique à deux voies et du son multivoies dans le détail

Une liste des caractéristiques est donnée aux § 5.2 et 5.3 de la Recommandation UIT-R BS.1116.

5.3 Caractéristiques spécifiant la relation entre le son et l'image d'accompagnement

Une liste de caractéristiques est donnée au § 5 de la Recommandation UIT-R BS.1286 – Méthodes d'évaluation subjective des systèmes audio avec images d'accompagnement.

5.4 Caractéristiques principales pour l'évaluation absolue de la qualité du son dans le détail

Une liste des caractéristiques est donnée par l'Union européenne de radio-télévision dans le document technique 3286, «Méthodes d'évaluation subjective de la qualité des programmes radiophoniques».

5.5 Caractéristiques spécifiant la qualité du son numérique transmis/codé dans le détail

Une liste des principales caractéristiques est donnée dans l'Annexe 2.

6 Eléments de programme

Selon l'objectif précis des essais et notamment de la catégorie du système de transmission et de reproduction du programme sonore testé, les éléments de programme peuvent être choisis de façon délibérée pour leur caractère hautement critique en ce qui concerne les dégradations introduites dans le système étudié. Dans d'autres cas, on peut utiliser des éléments de programme moins critiques.

Le § 6 de la Recommandation UIT-R BS.1116 contient une présentation détaillée des facteurs associés aux éléments de programme critiques d'essai et leur sélection en fonction des objectifs fixés.

Lorsque le système est destiné à acheminer un son de haute qualité, il conviendra d'utiliser des éléments de type critique. Pour pouvoir faire des comparaisons entre les données obtenues en différents lieux et/ou à différentes dates, il faudra utiliser les mêmes séquences de programme.

Dans tous les cas, le contenu d'une séquence de programme ne doit pas être excessivement intéressant ou désagréable ou fatigant afin de ne pas distraire l'auditeur.

7 Dispositifs de reproduction

7.1 Essais pour lesquels les haut-parleurs (ou écouteurs) ne font pas partie intégrante du système étudié

Il conviendra de suivre les indications contenues dans le § 7 de la Recommandation UIT-R BS.1116. Il faut cependant noter, que la mesure des pressions acoustiques avec pondération «A» associée à un signal à large bande ne donnent pas nécessairement une évaluation précise du volume sonore subjectif. Cela est particulièrement vrai lorsque les éléments du système de reproduction ont des bandes passantes différentes.

Il peut être nécessaire d'utiliser d'autres méthodes pour obtenir un réglage de gain correct pour toutes les voies de reproduction.

7.2 Essais pour lesquels les haut-parleurs (ou des écouteurs) font partie intégrante du système étudié

Les essais dans lesquels les dispositifs de reproduction font partie du système étudié doivent être réalisés conformément aux spécifications du système. Dans les essais comparatifs, le niveau sonore des systèmes doit être correctement réglé.

8 Conditions d'écoute

A moins de faire partie intégrante de l'essai, les conditions d'écoute doivent être conformes à celles spécifiées dans le § 8 de la Recommandation UIT-R BS.1116.

Il convient de noter que ces conditions spécifiées peuvent être beaucoup trop strictes pour certains types d'essais.

9 Traitement statistique des données

Les données subjectives doivent être traitées de manière à obtenir les valeurs moyennes et les intervalles de confiance. On peut ainsi décrire les données et, si la différenciation résultante ne correspond pas aux objectifs de l'essai, il conviendra d'effectuer un traitement plus poussé des données. On pourra utiliser les méthodes décrites dans le § 9 de la Recommandation UIT-R BS.1116. En général des connaissances en matière de statistiques seront nécessaires pour analyser les résultats.

La valeur globale de l'essai sera améliorée si l'analyse des données est encore plus poussée pour vérifier les hypothèses sous-jacentes de l'essai et évaluer la fiabilité des sujets.

10 Présentation des résultats

Dans la mesure du possible, la présentation des résultats sera conforme aux indications du § 10 de la Recommandation UIT-R BS.1116.

11 Contenu des rapports d'essais

Dans la mesure du possible, tous les aspects des essais décrits dans la Recommandation UIT-R BS.1116 devront être consignés, même si certains de ces aspects ne sont pas mis en œuvre ou contrôlés.

Le rapport devra par exemple consigner le fait qu'aucune formation n'a été dispensée.

ANNEXE 2

Catégories de défauts qui peuvent apparaître avec les techniques de codage ou de transmission numérique.

Pour l'évaluation des dégradations des signaux audio causées par le codage ou la transmission numérique, on peut utiliser un certain nombre de catégories pour analyser et classer le type de défaut.

<i>Catégorie de défaut</i>	<i>Explication</i>
Défaut de quantification	Défauts associés à une résolution numérique insuffisante (exemples: distorsion granulaire, modification non stationnaire du niveau de bruit)
Distorsion de la caractéristique de fréquence	Insuffisance de fréquences élevées ou basses, excès de fréquences hautes, sifflement, effets de formants, effets de filtre en peigne
Distorsion de la caractéristique de gain	Modification du niveau (gain) ou de la dynamique des signaux source, saut de niveau
Effet de modulation périodique	Variations périodiques de l'amplitude du signal (exemples: gazouillement, pompage ou grésillement)
Effet de modulation non périodique	Effets associés à des transitoires (exemple: claquement ou explosion, déformation des processus transitoires)
Distorsion non linéaire	Distorsion non linéaire harmonique ou non harmonique, distorsion par battement de fréquence
Distorsion temporelle	Pré et post-écho, écrasement (perte de la transparence temporelle du signal source), désynchronisation des signaux ou des voies
Son supplémentaire (bruit)	Bruits parasites non reliés à la séquence source tels des claquements, du bruit, des composantes tonales
Son manquant	Perte des composantes du son de la séquence source, par exemple causée par un défaut masquant
Effet de corrélation (diaphonie)	Diaphonie linéaire ou non linéaire entre les voies, fuite ou corrélation entre voies
Distorsion de la qualité spatiale de l'image	Tous les aspects comprenant l'étalement, le mouvement, la stabilité de localisation, l'équilibre, la précision de localisation, les modifications du caractère spatial.
