|  |  |  |
| --- | --- | --- |
| Recommandations | | Secteur des radiocommunications |
| Recommandation UIT-R BS.1387-2 | |
| (05/2023) | |
|  | Série BS: Service de radiodiffusion sonore | |
|  | Méthode de mesure objective de la qualité du son perçu | |

|  |  |
| --- | --- |
| **ITUPublications** | **Union internationale des télécommunications** |

Logo, icon

Description automatically generated

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, 2024

© UIT 2024

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.1387-2

Méthode de mesure objective de la qualité du son perçu

(1998-2001-2023)

Domaine d'application

Cette Recommandation indique une méthode de mesure objective de la qualité du son perçu.

Mots clés

Qualité de son perçue, mesure objective, codage à faible débit binaire

L'Assemblée des radiocommunications de l'UIT,

considérant

*a)* que les méthodes objectives traditionnelles (par exemple de mesure du rapport signal/bruit et de la distorsion) ne sont plus adaptées à la mesure de la qualité du son perçu de systèmes qui utilisent des schémas de codage à faible débit binaire ou un traitement des signaux analogiques ou numériques;

*b)* que les schémas de codage à faible débit binaire sont rapidement mis en service;

*c)* que les réalisations conformes à une spécification ou une norme ne garantissent pas toutes la meilleure qualité possible avec cette spécification ou cette norme;

*d)* que les méthodes formelles d'évaluation subjective ne sont pas adaptées au contrôle continu de la qualité du son, par exemple en conditions d'exploitation;

*e)* que la mesure objective de la qualité du son perçu peut éventuellement compléter ou remplacer les méthodes d'essai objectives traditionnelles dans tous les domaines de mesure;

*f)* que la mesure objective de la qualité du son perçu peut utilement compléter les méthodes d'évaluation subjective;

*g)* que, pour certaines applications, une méthode qui puisse être appliquée en temps réel est nécessaire,

recommande

1 que pour les applications dont on trouvera la liste en Annexe 1, la méthode expliquée dans l'Annexe 2 soit utilisée pour la mesure objective de la qualité du son perçu.

Avant-propos

Cette Recommandation définit une méthode de mesure objective de la qualité du son perçu d'un système testé, par exemple un codec à faible débit binaire. Elle est composée de deux annexes. L'Annexe 1 offre à l'utilisateur un aperçu général de la méthode et comporte quatre Pièces jointes. La Pièce jointe 1 décrit les applications et les signaux de test. La Pièce jointe 2 est une liste des variables de sortie de modèle; on y trouvera également les limites d'utilisation et de précision. La Pièce jointe 3 est une description rapide du modèle et la Pièce jointe 4 décrit les principes et les caractéristiques générales des méthodes de mesure objective de la qualité du son perçu.

L'Annexe 2 offre au réalisateur une description détaillée de la méthode avec les deux versions du modèle psychoacoustique développées pendant la phase d'intégration lors de laquelle six modèles ont été associés. On trouvera dans la Pièce jointe 1 de l'Annexe 2 la description du processus de validation de la méthode de mesure objective. La Pièce jointe 2 de l'Annexe 2 survole toutes les bases de données utilisées pour le développement et la validation de la méthode.

TABLE DES MATIÈRES

Page

[Politique en matière de droits de propriété intellectuelle (IPR) ii](#_Toc160633128)

[Annexe 1 – Aperçu général 8](#_Toc160633129)

[1 Introduction 8](#_Toc160633130)

[2 Applications 8](#_Toc160633131)

[3 Versions 9](#_Toc160633132)

[4 Le domaine subjectif 9](#_Toc160633133)

[5 Résolution et précision 11](#_Toc160633134)

[6 Exigences et limitations 11](#_Toc160633135)

[Pièce jointe 1 de l'Annexe 1 – Applications 12](#_Toc160633136)

[1 Généralités 12](#_Toc160633137)

[2 Applications principales 12](#_Toc160633138)

[2.1 Évaluation des réalisations 12](#_Toc160633139)

[2.2 Alignement qualité perceptuelle 12](#_Toc160633140)

[2.3 Contrôle en ligne 12](#_Toc160633141)

[2.4 État des équipements ou des connexions 13](#_Toc160633142)

[2.5 Identification des codecs 13](#_Toc160633143)

[2.6 Développement des codecs 13](#_Toc160633144)

[2.7 Planification des réseaux 13](#_Toc160633145)

[2.8 Aides à l'évaluation subjective 14](#_Toc160633146)

[2.9 Résumé des applications 14](#_Toc160633147)

[3 Les signaux d'essais 14](#_Toc160633148)

[3.1 Sélection de signaux d'essais naturels 15](#_Toc160633149)

[3.2 Durée 16](#_Toc160633150)

[4 Synchronisation 16](#_Toc160633151)

[5 Droits de reproduction 16](#_Toc160633152)

[Pièce jointe 2 de l'Annexe 1 – Variables de sortie 16](#_Toc160633153)

[1 Introduction 16](#_Toc160633154)

[2 Les variables de sortie de modèle 17](#_Toc160633155)

[3 Qualité audio de base 17](#_Toc160633156)

[4 Marge de codage 18](#_Toc160633157)

[5 Exigences des utilisateurs 18](#_Toc160633158)

[Pièce jointe 3 de l'Annexe 1 – Présentation du modèle 19](#_Toc160633159)

[1 Traitement du son 20](#_Toc160633160)

[1.1 Paramètres définis par l'utilisateur 20](#_Toc160633161)

[1.2 Modèle psychoacoustique 20](#_Toc160633162)

[1.3 Modèle cognitif 20](#_Toc160633163)

[Pièce jointe 4 de l'Annexe 1 – Principes et caractéristiques des méthodes de mesure objective de la qualité perceptuelle du son 21](#_Toc160633164)

[1 Introduction et historique 21](#_Toc160633165)

[2 Structure générale des méthodes de mesure objective de la qualité   
perceptuelle du son 22](#_Toc160633166)

[3 Considérations psychoacoustiques et cognitives 22](#_Toc160633167)

[3.1 Caractéristique de transfert par l'oreille externe et par l'oreille moyenne 23](#_Toc160633168)

[3.2 Échelles perceptuelles des fréquences 23](#_Toc160633169)

[3.3 Excitation 24](#_Toc160633170)

[3.4 Détection 25](#_Toc160633171)

[3.5 Masquage 25](#_Toc160633172)

[3.6 Intensité acoustique et masquage partiel 26](#_Toc160633173)

[3.7 Acuité 26](#_Toc160633174)

[3.8 Traitement cognitif 27](#_Toc160633175)

[4 Les modèles intégrés 28](#_Toc160633176)

[4.1 DIX 28](#_Toc160633177)

[4.2 NMR 28](#_Toc160633178)

[4.3 OASE 29](#_Toc160633179)

[4.4 PAQM (perceptual audio quality measure) 29](#_Toc160633180)

[4.5 PERCEVAL 30](#_Toc160633181)

[4.6 POM 31](#_Toc160633182)

[4.7 L'approche Toolbox 31](#_Toc160633183)

[Annexe 2 – Description du modèle 32](#_Toc160633184)

[1 Généralités 32](#_Toc160633185)

[1.1 Version de base 33](#_Toc160633186)

[1.2 Version avancée 33](#_Toc160633187)

[2 Modèle auditif périphérique 33](#_Toc160633188)

[2.1 Modèle auditif TFR 33](#_Toc160633189)

[2.2 Modèle auditif fondé sur un banc de filtres 45](#_Toc160633190)

[3 Prétraitement des caractéristiques d'excitation 54](#_Toc160633191)

[3.1 Adaptation des niveaux et des caractéristiques 54](#_Toc160633192)

[3.2 Modulation 56](#_Toc160633193)

[3.3 Intensité acoustique 56](#_Toc160633194)

[3.4 Calcul du signal d'erreur 57](#_Toc160633195)

[4 Calcul des variables de sortie de modèle 57](#_Toc160633196)

[4.1 Aperçu général 57](#_Toc160633197)

[4.2 Différence de modulation 58](#_Toc160633198)

[4.3 Intensité acoustique du bruit 59](#_Toc160633199)

[4.4 Largeur de bande 61](#_Toc160633200)

[4.5 Rapport bruit/masque (NMR) 62](#_Toc160633201)

[4.6 Relative Disturbed FramesB 63](#_Toc160633202)

[4.7 Probabilité de détection 63](#_Toc160633203)

[4.8 Structure harmonique de l'erreur 65](#_Toc160633204)

[5 Moyennes 66](#_Toc160633205)

[5.1 Calcul de la moyenne spectrale 66](#_Toc160633206)

[5.2 Calcul de la moyenne temporelle 66](#_Toc160633207)

[5.3 Moyenne sur les canaux audio 68](#_Toc160633208)

[6 Estimation de la qualité audio de base perçue 68](#_Toc160633209)

[6.1 Réseau neuronal artificiel 68](#_Toc160633210)

[6.2 Version de base 69](#_Toc160633211)

[6.3 Version avancée 71](#_Toc160633212)

[7 Conformité des mises en œuvre 72](#_Toc160633213)

[7.1 Généralités 72](#_Toc160633214)

[7.2 Sélection 72](#_Toc160633215)

[7.3 Réglages pour les tests de conformité 72](#_Toc160633216)

[7.4 Intervalle de tolérance acceptable 73](#_Toc160633217)

[7.5 Éléments de tests 73](#_Toc160633218)

[Pièce jointe 1 de l'Annexe 2 – Processus de validation 74](#_Toc160633219)

[1 Généralités 74](#_Toc160633220)

[2 Phase de compétition 75](#_Toc160633221)

[3 Phase de collaboration 76](#_Toc160633222)

[4 Vérification 76](#_Toc160633223)

[4.1 Comparaison des valeurs de SDG et de ODG 77](#_Toc160633224)

[4.2 Corrélation 77](#_Toc160633225)

[4.3 Valeurs de l'erreur absolue (AES) 79](#_Toc160633226)

[4.4 Comparaison des ODG et de l'intervalle de confiance 80](#_Toc160633227)

[4.5 Comparaison des ODG et de l'intervalle de confiance (tolérance) 84](#_Toc160633228)

[5 Sélection des versions optimales du modèle 86](#_Toc160633229)

[5.1 Critères de présélection fondés sur la corrélation 86](#_Toc160633230)

[5.2 Analyse du nombre d'éléments non conformes 87](#_Toc160633231)

[5.3 Analyse de la gravité des non-conformités 87](#_Toc160633232)

[6 Conclusion 88](#_Toc160633233)

[Pièce jointe 2 de l'Annexe 2 – Description des bases de données de référence 89](#_Toc160633234)

[1 Introduction 89](#_Toc160633235)

[2 Éléments par base de données 91](#_Toc160633236)

[3 Conditions expérimentales 91](#_Toc160633237)

[3.1 MPEG90 92](#_Toc160633238)

[3.2 MPEG91 92](#_Toc160633239)

[3.3 ITU92DI 92](#_Toc160633240)

[3.4 ITU92CO 92](#_Toc160633241)

[3.5 ITU93 92](#_Toc160633242)

[3.6 MPEG95 93](#_Toc160633243)

[3.7 EIA95 93](#_Toc160633244)

[3.8 DB2 93](#_Toc160633245)

[3.9 DB3 93](#_Toc160633246)

[3.10 CRC97 94](#_Toc160633247)

[4 Eléments par condition pour DB2 et DB3 94](#_Toc160633248)

[4.1 DB2 94](#_Toc160633249)

[4.2 DB3 97](#_Toc160633250)

[Références 99](#_Toc160633251)

[Bibliographie 100](#_Toc160633252)

Annexe 1  
  
Aperçu général

# 1 Introduction

La qualité du son est l'un des facteurs clefs à prendre en compte lors de la conception d'un système numérique de radiodiffusion. L'introduction rapide de divers schémas de réduction du débit binaire a conduit à des efforts significatifs pour définir et affiner des procédures d'évaluations subjectives, tout simplement parce que les tests d'écoute formels étaient la seule méthode pertinente pour juger de la qualité du son. L'expérience acquise a servi de base à la Recommandation UIT‑R BS.1116, qui à son tour a servi de fondement à la plupart des tests d'écoute de ce type.

Les évaluations subjectives de la qualité étant à la fois longues et coûteuses, il est nécessaire de développer une méthode de mesure objective pour obtenir une estimation de la qualité du son. Les méthodes de mesure objective traditionnelles, comme le rapport signal/bruit (*S*/*N*) ou la distorsion harmonique totale (THD) n'ont jamais apporté la preuve de leur fiabilité pour ce qui est de leur rapport à la qualité du son perçu. Les problèmes sont encore plus évidents lorsque les méthodes sont appliquées aux codecs modernes qui sont à la fois non linéaires et non stationnaires.

Ces dix dernières années ont vu l'apparition d'un certain nombre de méthodes de mesure objective de la qualité du son perçu. Mais aucune de ces méthodes n'a été totalement validée, ni, en conséquence, normalisée ou acceptée sur une grande échelle. En 1994, l'UIT-R a identifié le besoin urgent d'établir une norme dans ce domaine, et les travaux ont été lancés. Un appel à propositions a été passé et les six méthodes de mesure proposées suivantes ont été retenues; Indice de perturbation (DIX), Rapport bruit/masque (NMR), Mesure perceptuelle de la qualité du son (PAQM), Évaluation perceptuelle PERCEVAL, Mesure perceptuelle objective (POM) et la Toolbox Approach. Ces méthodes sont décrites dans la Pièce jointe 4 à l'Annexe 1.

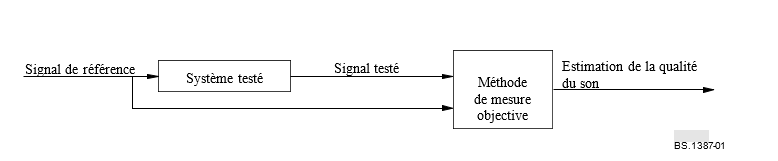
La méthode de mesure proposée dans cette Recommandation est le résultat d'un processus au cours duquel on a étudié les performances des six méthodes ci-dessus, et extrait les outils les plus prometteurs pour les intégrer à une méthode unique. La méthode recommandée a été soigneusement validée sur un certain nombre de sites d'essais. Elle s'est avérée générer des informations fiables et utiles pour de nombreuses applications. Il ne faut cependant pas oublier que la méthode de mesure objective proposée dans cette Recommandation ne remplace pas, de façon générale, un test d'écoute formel.

# 2 Applications

Le concept fondamental des mesures objectives avec la méthode recommandée est illustré par la Fig. 1 ci-dessous.

FIGURE 1

Concept fondamental des mesures objectives



BS.1387-01

La méthode de mesure proposée dans cette Recommandation est applicable à la plupart des types d'équipement de traitement de signaux audio, numériques et analogiques. Néanmoins, il est probable que beaucoup d'applications seront fondées sur des codecs audio.

On a identifié les huit classes d'applications décrites dans le Tableau 1:

TABLEAU 1

Applications

|  |  |  |  |
| --- | --- | --- | --- |
|  | Application | Description rapide | Version |
| 1 | Évaluation de réalisations | Procédure qui permet de caractériser différentes réalisations d'équipements de traitement audio, dans de nombreux cas des codecs audio | Base/Avancée |
| 2 | Alignement qualité perceptuelle | Procédure rapide précédant la mise en service d'une partie d'équipement ou d'un circuit | Base |
| 3 | Contrôle en ligne | Processus en continu pour surveiller une transmission audio en service | Base |
| 4 | État des équipements ou des connexions | Analyse détaillée d'une partie d'équipement ou d'un circuit | Avancée |
| 5 | Identification de codec | Procédure permettant d'identifier le type et la réalisation d'un codec donné | Avancée |
| 6 | Développement de codec | Procédure qui permet de caractériser les performances du codec aussi précisément que possible | Base/Avancée |
| 7 | Planification de réseau | Procédure qui permet d'optimiser le coût et les performances d'un réseau de transmission selon des contraintes données | Base/Avancée |
| 8 | Aide à l'évaluation subjective | Outil de sélection des matériaux critiques à intégrer à un test d'écoute | Base/Avancée |

# 3 Versions

Afin de parvenir à la meilleure adéquation possible avec les différentes exigences de coût et de performances, la présente Recommandation propose deux versions de la méthode de mesure objective recommandée. La version de base permet une réalisation en temps réel d'un bon rapport coût/performances, la version avancée vise à obtenir la plus grande précision possible. Selon la réalisation, cette précision supplémentaire augmente la complexité selon un facteur d'environ quatre par rapport à la version de base.

Le Tableau 1 aidera au choix de la version à utiliser pour chacune des applications.

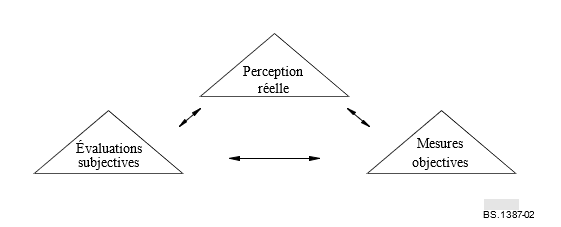
# 4 Le domaine subjectif

Les essais formels d'écoute subjective, par exemple ceux fondés sur la Recommandation UIT‑R BS.1116, sont précisément conçus pour parvenir à une estimation aussi fiable que possible de la qualité du son. On ne peut cependant pas espérer que le résultat d'un essai d'écoute subjectif soit le parfait reflet de la perception réelle. La Figure 2 illustre les imperfections inévitables, aussi bien dans le domaine subjectif que dans le domaine objectif.

Il est évidemment impossible de valider directement une méthode objective. Les méthodes de mesure objectives sont donc validées par rapport à des essais d'écoute subjective.

FIGURE 2

Concepts de validation



BS.1387-02

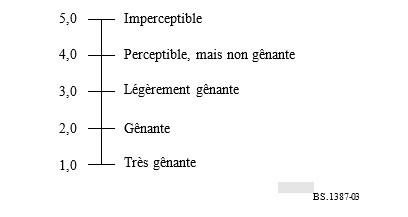
La méthode de mesure objective proposée dans la présente Recommandation est centrée sur des applications qui sont normalement évaluées dans le domaine subjectif en appliquant la Recommandation UIT-R BS.1116. On peut rapidement décrire le principe fondamental de cette méthode d'essai particulière de la façon suivante: l'auditeur a le choix entre trois sources («A», «B», «C»). La source «A» est toujours le Signal de référence connu. Le Signal de référence masqué et le Signal testé sont disponibles simultanément mais sont, selon l'essai, affectés de façon aléatoire à «B» ou à «C».

On demande à l'auditeur d'évaluer les dégradations en «B» par rapport à «A», et en «C» par rapport à «A», selon l'échelle de dégradations continue à cinq notes. Une des sources, «B» ou «C», ne doit pas pouvoir être discernée de la source «A»; l'autre peut révéler des dégradations. Toute différence perçue entre la référence et l'autre source doit être interprétée comme étant une dégradation. On n'utilise normalement qu'un seul attribut, la «Qualité audio de base», défini comme un attribut global qui intègre toutes les différences détectées entre le Signal de référence et le Signal testé.

Cette échelle d'évaluation est continue et présente des «points de référence» qui proviennent de l'échelle de dégradation à cinq notes proposée dans la Recommandation UIT‑R BS.1284 et représentée ci-dessous.

FIGURE 3

Échelle de dégradation à cinq notes de l'UIT-R



L'analyse des résultats d'un essai d'écoute subjective est généralement fondée sur la note différentielle subjective (SDG), définie par l'équation:

*SDG*  *GradeSignal Under Test* – *GradeReference Signal*

Idéalement, les valeurs de SDG doivent être situées dans une fourchette entre 0 et −4, où 0 correspond à une dégradation imperceptible et −4 à une dégradation jugée comme très gênante.

# 5 Résolution et précision

La note différentielle objective (ODG) est la variable de sortie de la méthode de mesure objective; elle correspond à la SDG dans le domaine subjectif. La résolution de l'ODG est limitée à une décimale. Il faut néanmoins être prudent et ne pas croire, en règle générale, qu'une différence d'un dixième entre deux paires quelconques d'ODG soit significative. Il en va de même lorsque l'on étudie les résultats d'un essai d'écoute subjective.

La précision de la méthode de mesure objective ne peut être décrite par un chiffre unique. Il faut prendre en compte un certain nombre d'indices de qualité, dont la corrélation entre les SDG et les ODG. Il est important de savoir qu'il n'y a pas de garantie que cette corrélation dépassera une valeur prédéfinie. Les performances de la méthode de mesure seront susceptibles de varier selon, par exemple, le type et le niveau de la dégradation introduite.

Le nombre de valeurs non conformes est également intéressant. Une valeur non conforme est une valeur mesurée qui n'entre pas dans une fourchette de tolérance prédéfinie. Selon les exigences de l'utilisateur, la méthode de mesure devra fournir la plus grande précision possible dans le haut de l'échelle (c'est-à-dire une qualité audio élevée). En conséquence, on acceptera que la précision soit moindre dans les parties médiane et inférieure de l'échelle.

La corrélation offre normalement une bonne estimation de la précision de la méthode de mesure objective, mais il faut garder à l'esprit qu'une valeur de corrélation relativement élevée peut cacher une performance inacceptable (du point de vue des valeurs non conformes) d'une méthode de mesure.

Le troisième chiffre intéressant utilisé lors du processus de validation est la valeur de l'erreur absolue (AES), qui est le reflet de la moyenne de la relation entre la taille de l'intervalle de confiance de la SDG et l'écart entre SDG et ODG.

On trouvera dans la Pièce jointe 1 de l'Annexe 2 plus de détails concernant les performances attendues de la méthode de mesure et les performances lors du processus de validation.

# 6 Exigences et limitations

Le signal provenant du Système testé et le Signal de référence doivent être synchronisés avec une précision de 24 échantillons sur toute la durée de la mesure. Le mécanisme de synchronisation ne fait pas partie de cette Recommandation parce qu'il variera selon la réalisation.

Pièce jointe 1  
de l'Annexe 1  
  
Applications

# 1 Généralités

Cette Pièce jointe contient les définitions et exigences spécifiques des principales applications auxquelles est destinée cette Recommandation d'une méthode de mesure objective de la qualité du son perçu.

Certaines applications demandent une mise en œuvre en temps réel de la méthode de mesure objective; pour d'autres applications, une mesure en temps réel n'est pas nécessaire. Pour les mises en œuvre en temps réel, nous recommandons que le retard maximum dans l'appareil de mesure n'excède pas 200 ms; un retard de plus de 1 s n'est pas acceptable.

De plus, il convient de distinguer entre les mesures en ligne et les mesures hors ligne. Lors des mesures hors ligne, la procédure de mesure a un accès total à l'appareil ou à la connexion; les mesures en ligne impliquent qu'un programme est en cours et qu'il ne doit pas être interrompu par la mesure.

# 2 Applications principales

## 2.1 Évaluation des réalisations

Les radiodiffuseurs, opérateurs de réseaux, et autres, ont besoin d'évaluer diverses réalisations d'équipements, particulièrement de codecs audio, lors du choix de l'équipement à acheter ou lors d'essais de réception.

Pour ces types d'applications, une haute précision est nécessaire, particulièrement pour évaluer les petites dégradations et classer correctement différentes réalisations. En ce qui concerne les variables de sortie, une sortie simple comme l'ODG suffit aux utilisateurs, mais les développeurs de codecs audio pourront effectuer une analyse plus profonde avec un ensemble adapté de variables de sortie de modèle (MOV).

On peut utiliser les deux versions du modèle, mais nous recommandons la version avancée.

## 2.2 Alignement qualité perceptuelle

Il s'agit d'une procédure rapide lancée avant la mise en service d'une partie d'équipement ou d'un circuit. Le but en est de contrôler la fonctionnalité et la qualité. L'équipement de mesure sera géré par le personnel d'exploitation. Tous types de distorsion peuvent être présents.

La mesure doit se faire en temps réel. On peut utiliser des signaux d'essai ou des signaux audio prédéfinis. Les ODG doivent être correctement affichées et données au moins deux fois par seconde ou, si l'on utilise un signal d'essai spécial, directement à la fin du signal d'essai.

La version de base suffit.

## 2.3 Contrôle en ligne

Il s'agit d'une procédure en continu, pendant une transmission radiophonique. Le programme ne doit pas être interrompu par la procédure de mesure. On doit donc utiliser pour la mesure soit le signal du programme lui-même, soit un fragment audio prédéfini. Ce dernier peut être un signal de la station ou un jingle. L'équipement de mesure sera géré par le personnel d'exploitation.

La mesure doit se faire en temps réel. Les ODG doivent être correctement affichées et données au moins deux fois par seconde, ou directement à la fin du signal prédéfini. L'affichage des MOV n'est pas nécessaire.

La version de base suffit.

## 2.4 État des équipements ou des connexions

Pour assurer la fonctionnalité des connexions ou des équipements audio, un contrôle complet de la qualité est nécessaire de temps à autre. Contrairement au contrôle en ligne ou à l'alignement perceptuel, cette application nécessite le contrôle de plusieurs paramètres techniques.

Le système de mesure devra fournir des informations détaillées sur l'influence des états des équipements ou des connexions sur la qualité du son perçu en affichant le jeu complet de MOV en plus des ODG. La mesure en temps réel n'est pas nécessaire.

La version avancée est recommandée.

## 2.5 Identification des codecs

Pour identifier les codecs (différents algorithmes ou différentes mises en œuvre du même algorithme), le système de mesure doit pouvoir mémoriser, récupérer et comparer différentes caractéristiques. Une similarité entre les caractéristiques peut être interprétée comme une mesure de la similarité de différentes mises en œuvre de codecs. On utilise cette procédure pour identifier le type et la mise en œuvre d'un codec particulier.

Le système de mesure doit enregistrer autant d'informations que possible sur les caractéristiques. Les ODG à elles seules peuvent ne pas fournir assez d'informations.

La version de base suffit, même si une mesure en temps réel n'est pas nécessaire.

NOTE – On est peu familiarisé avec la méthode recommandée. De plus, aucune mesure de similarité entre les caractéristiques n'est encore définie.

## 2.6 Développement des codecs

Pour cette application, la méthode de mesure doit déterminer la performance du codec testé de manière aussi précise et détaillée que possible, particulièrement pour les faibles dégradations.

Le contrôle continu nécessite un traitement en temps réel qui n'est pas forcément supporté par la version avancée. Mais, en ce qui concerne les faibles dégradations et pour obtenir des informations détaillées, il faut utiliser la version avancée. Le système de mesure doit pouvoir afficher les valeurs de sortie à la même vitesse que celle à laquelle elles sont calculées. Il faut avoir un accès direct à l'historique des sorties sur une durée de 4 s.

Nous recommandons d'utiliser la version avancée. Néanmoins, pour des mesures en temps réel, la version de base suffit. Il faut effectuer une analyse aussi bien en temps réel, qu'en hors ligne et trame par trame. Toute distorsion grave devra être indiquée, par exemple par affichage des valeurs de crêtes. Il faut avoir accès à l'ensemble des MOV.

## 2.7 Planification des réseaux

La planification des réseaux nécessite une évaluation de la qualité espérée à différents moments du processus de planification. Pour étudier différentes configurations afin d'optimiser la qualité du son, on peut procéder à une simulation logicielle des composants du réseau qui permettra d'associer différentes étapes de traitement du son. Ultérieurement, les composants réels du traitement du son pourront être testés dans la configuration choisie.

La planification des réseaux est effectuée par des ingénieurs de système qui doivent pouvoir récupérer des informations détaillées sur l'influence des caractéristiques du réseau sur la qualité du son. La classification des différentes configurations réseaux possibles devra être fondée sur un ensemble de MOV adapté à l'application spécifique du réseau. Ainsi, un affichage des seules ODG n'est pas suffisant. Une évaluation dans cette application ne nécessite pas une mesure en temps réel.

On peut utiliser les deux versions du modèle, mais la version avancée est recommandée.

## 2.8 Aides à l'évaluation subjective

La méthode de mesure objective fournit un outil qui permet de choisir le matériau sonore critique à utiliser pour les essais d'écoute subjectifs. On peut utiliser l'ensemble des MOV pour définir ce matériau.

Il faut obtenir la précision la plus haute possible; la version avancée est donc recommandée. Mais il faut effectuer des mesures en temps réel afin de réduire le temps nécessaire au choix des matériaux critiques.

## 2.9 Résumé des applications

Le Tableau 2 résume les exigences relatives à la méthode de mesure pour les principales applications.

TABLEAU 2

Exigences relatives à la méthode de mesure

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Application | Catégorie | Temps réel | Min,  ROV(1) (Hz) | En ligne/ Hors ligne | Version du modèle |
| 1 | Évaluation de réalisations | Diagnostic | Non | – | Hors ligne | Les deux |
| 2 | Alignement qualité perceptuelle | Exploitation | Oui/Non | 2 | Hors ligne | Base |
| 3 | Contrôle en ligne | Exploitation | Oui | 2 | En ligne | Base |
| 4 | État des équipements ou des connexions | Diagnostic | Oui/Non | – | En ligne/ Hors ligne | Avancée |
| 5 | Identification de codec | Diagnostic | Non | – | Hors ligne | Les deux |
| 6 | Développement de codec | Développement | Oui/Non | – | Hors ligne | Les deux |
| 7 | Planification de réseau | Développement | Oui/Non | – | Hors ligne | Les deux |
| 8 | Aide à l'évaluation subjective | Développement | Oui/Non | – | Hors ligne | Avancée |
| (1) Débit des valeurs de sorties (par seconde). | | | | | | |

# 3 Les signaux d'essais

On distingue deux groupes de signaux d'essais: les signaux naturels et les signaux synthétisés. La liste des signaux d'essais naturels fournie ici est constituée de séquences sonores critiques déjà utilisées, par l'UIT‑R et d'autres organisations, lors de tests d'écoute, pour l'évaluation de la qualité du son. Ces signaux doivent être disponibles sur le site de transmission et sur le site de mesure. Le système de mesure doit donc avoir une mémoire.

Les signaux synthétisés sont définis mathématiquement et peuvent être transformés de façon contrôlée. Ces signaux peuvent être générés sur le site de transmission et sur le site de mesure. Il n'y a pas besoin de mémoire supplémentaire dans le système de mesure. Du fait de la nature de ces signaux, il est difficile, voire impossible, de dériver pour eux des classifications subjectives. En conséquence, pour ces signaux, la méthode de mesure n'a pas été validée par comparaison avec les résultats subjectifs.

## 3.1 Sélection de signaux d'essais naturels

On trouvera dans le Tableau 3 une liste de sous-ensembles de signaux d'essais utilisés lors de la procédure de vérification qui a conduit à la présente Recommandation. Le type d'artefacts révélés par ces signaux codec à faible débit binaire est également indiqué.

TABLEAU 3

Liste de sous-ensembles de signaux d'essais

| N° | Élément | Nom fichier | Remarques |
| --- | --- | --- | --- |
| 1 | Castagnettes | cas | (1) |
| 2 | Clarinette | cla | (2) |
| 3 | Claves | clv | (1) |
| 4 | Flûte | flu | (2) |
| 5 | Glockenspiel | glo | (1), (2), (5) |
| 6 | Clavecin | hrp | (1), (2), (4) |
| 7 | Timbale | ket | (1) |
| 8 | Marimba | mar | (1) |
| 9 | Piano Schubert | pia | (2) |
| 10 | Diapason | pip | (4) |
| 11 | Ry Cooder | ryc | (2), (4) |
| 12 | Saxophone | sax | (2) |
| 13 | Cornemuse | sb1 | (2), (4), (5) |
| 14 | Voix de femme Angl. | sfe | (3) |
| 15 | Voix d'homme Angl. | sme | (3) |
| 16 | Voix d'homme Allem. | smg | (3) |
| 17 | Caisse claire | sna | (1) |
| 18 | Soprano Mozart | sop | (4) |
| 19 | Tambourin | tam | (1) |
| 20 | Trompette | tpt | (2) |
| 21 | Triangle | tri | (1), (2), (5) |
| 22 | Tuba | tub | (2) |
| 23 | Susanne Vega | veg | (3), (4) |

TABLEAU 3 (*fin*)

| N° | Élément | Nom fichier | Remarques |
| --- | --- | --- | --- |
| 24 | Xylophone | xyl | (1), (2) |
| (1) Transitoires: sensibilité au préécho, rémanence du bruit dans le domaine temporel.  (2) Structure tonale: sensibilité au bruit, rugosité.  (3) Voix naturelle (association critique de parties tonales et d'attaques): sensibilité à la distorsion, dégradation des attaques.  (4) Son complexe: soumet le Système testé à de fortes contraintes.  (5) Grande largeur de bande: soumet le Système testé à de fortes contraintes, perte des fréquences élevées, bruits de fréquence élevée modulés par le programme. | | | |

## 3.2 Durée

La durée d'un signal d'essai naturel doit être à peu près la même que lors d'un test d'écoute. Cette durée est typiquement de l'ordre de 10 à 20 s. Il est très probable que la partie critique du signal d'essai, qui dévoile la plupart des artefacts, est limitée à une brève fraction de cette durée.

La durée des signaux d'essais synthétisés doit être suffisante pour soumettre le codec testé à de fortes contraintes; ce codec peut comporter un buffer de signal sonore codé. Si l'on considère la longueur de ces tampons et les constantes temporelles de la méthode de mesure, la durée de chaque séquence d'un élément de test doit être supérieure à 500 ms. La durée peut être limitée à cette valeur faible parce que ces signaux ne seront pas utilisés lors de tests d'écoute subjectifs.

# 4 Synchronisation

Pour la procédure de mesure, le Signal testé et le Signal de référence devront être synchronisés dans le temps. Cette remarque s'applique aux signaux d'essai naturels et synthétisés.

# 5 Droits de reproduction

Les signaux d'essai du Tableau 3 ne peuvent être utilisés hors droits de reproduction qu'aux fins de mesure en association avec la méthode de mesure objective, décrite dans l'Annexe 2 de cette Recommandation.

NOTE – L'exonération des droits de reproduction doit être obtenue pour toutes les séquences, essentiellement auprès de l'UER (disque EBU SQAM).

Pièce jointe 2  
de l'Annexe 1  
  
Variables de sortie

# 1 Introduction

La méthode de mesure objective décrite dans cette Recommandation mesure la qualité du son et propose une valeur censée correspondre à la qualité du son perçu. Cette méthode de mesure modélise les propriétés fondamentales du système auditif. De nombreux stades intermédiaires modélisent les effets physiologiques et psychoacoustiques.

Ces valeurs intermédiaires peuvent servir à déterminer les artefacts. Ces paramètres sont les variables de sortie de modèle (MOV). La dernière étape du modèle de mesure combine les valeurs de MOV pour obtenir une valeur de sortie unique qui correspond directement au résultat espéré d'une évaluation de la qualité subjective.

# 2 Les variables de sortie de modèle

Le Tableau 4 présente une description des MOV utilisées pour prédire les notes différentielles objectives. Les indicesA proviennent du banc de filtres du modèle, les indicesB du modèle de transformation de Fourier rapide (TFR). Les notes différentielles objectives sont prédites, soit uniquement à partir de la partie TFR (version de base), soit d'une association de la partie TFR et du banc de filtres (version avancée). La moyenne est toujours calculée par rapport au temps.

# 3 Qualité audio de base

Le paramètre le mieux connu à partir des essais d'écoute subjective est la qualité audio de base (BAQ). La BAQ est la mesure d'une note différentielle subjective (SDG) qui est la différence entre le niveau de la référence et le niveau du Signal testé dans un test subjectif[[1]](#footnote-1)1. La SDG a normalement une valeur négative. On appelle note différentielle objective (ODG) le paramètre de sortie du modèle correspondant. La fonction de passage des MOV à une ODG est fondée sur un grand nombre d'éléments d'essai fiables, voir l'Annexe 2, Pièce jointe 2.

TABLEAU 4

Description des variables de sortie de modèle

|  |  |
| --- | --- |
| Variables de sortie de modèle | Description |
| *WinModDiffB* | *Différence moyenne de modulation après fenêtrage (enveloppes) entre le Signal de référence et le Signal testé* |
| *AvgModDiff1B* | *Différence de modulation moyenne* |
| *AvgModDiff2B* | *Différence de modulation moyenne avec prise en compte particulière des modulations introduites et des variations de modulation quand la référence contient peu ou pas de modulations* |
| *RmsModDiffA* | *Valeur Rms de la différence de modulation* |
| *RmsMissingComponentsA* | *Valeur Rms de l'intensité du bruit des composantes de fréquences manquantes, (valeur utilisée dans RmsNoiseLoudAsymA)* |
| *RmsNoiseLoudA* | *Valeur Rms de l'intensité moyenne du bruit avec prise en compte particulière des composantes introduites* |
| *RmsNoiseLoudAsymA* | *RmsNoiseLoudA**0,5RmsMissingComponentsA* |
| *AvgLinDistA* | *Mesure des distorsions linéaires moyennes avec une pondération plus importante dans les bandes de fréquences élevées* |
| *BandwidthRefB* | *Largeur de bande du Signal de référence* |

TABLEAU 4 (*fin*)

|  |  |
| --- | --- |
| Variables de sortie de modèle | Description |
| *BandwidthTestB* | *Largeur de bande du signal de sortie du Système testé* |
| *TotNMRB* | *Logarithme de la moyenne du rapport bruit/masque total* |
| *RelDistFramesB* | *Fraction relative des trames pour lesquelles au moins une bande de fréquences contient une composante de bruit significative* |
| *AvgSegmNMRB* | *Moyenne par segmentation du logarithme du rapport bruit/masque* |
| *MFPDB* | *Probabilité maximum de détection après filtre passe-bas* |
| *ADBB* | *Moyenne de blocs distordus, logarithme du rapport de la distorsion totale/nombre total de trames fortement distordues* |
| *EHSB* | *Structure harmonique de l'erreur dans le temps* |

L'ODG est le paramètre mesuré de façon objective qui correspond à la qualité perçue de façon subjective. La tâche d'un auditeur lors d'un essai d'écoute étant d'évaluer la BAQ d'un élément d'essai, l'ODG est également une mesure de la BAQ.

# 4 Marge de codage

La marge de codage (CM), description des artefacts inaudibles, est un paramètre qui pourrait s'avérer intéressant dans le futur. On évalue la marge de codage subjective (SCM) en amplifiant les artefacts jusqu'à ce qu'ils deviennent audibles par l'auditeur. La SCM définit l'écart jusqu'au seuil d'audibilité des artefacts.

Pour déterminer ce seuil, on doit amplifier ou atténuer les artefacts lors du test d'écoute. On utilise la méthode des différences. Le signal de différence entre l'original synchrone dans le temps et le signal codé est amplifié et ajouté au signal original. La meilleure méthode de détection du seuil d'audibilité est la méthode du choix forcé. On obtient la SCM en faisant la moyenne des valeurs de seuil pour l'amplification ou l'atténuation obtenues auprès des auditeurs. Des valeurs négatives de CM indiquent des artefacts audibles, des valeurs positives indiquent des artefacts inaudibles. Contrairement à la BAQ, la marge de codage est la mesure du moment (du niveau) où les artefacts deviennent audibles et non de la gêne qu'ils génèrent. On trouvera en [Feiten, 1997] la définition et la validation de la méthode de mesure de la SCM.

La marge de codage objective (OCM) provient elle aussi des MOV. A l'heure actuelle, on n'a évalué que peu d'éléments de test pour la marge de codage objective. La schématisation des OCM du modèle de la présente Recommandation n'a pas encore fait l'objet de recherches.

# 5 Exigences des utilisateurs

Les exigences des utilisateurs en ce qui concerne les variables de sortie de la méthode de mesure varient selon l'application. Pour certaines applications, par exemple, les numéros 2 et 3 (voir la Pièce jointe 1 de l'Annexe 1), la mesure fait partie d'une procédure d'exploitation. Il est très important dans ce cas que le résultat de la méthode soit facile à lire et à interpréter par des personnes n'ayant pas de connaissances approfondies de la technique de mesure. Le meilleur moyen est que la méthode ne donne qu'**une** **seule valeur** qui corresponde à la qualité du son perçu.

Cette remarque peut concerner d'autres applications, par exemple, les Applications 1 et 4. Néanmoins, pour ces applications, ainsi que pour les Applications 5-8, des variables de sortie plus sophistiquées pourront représenter un avantage pour les utilisateurs ayant une connaissance plus approfondie des mécanismes de la méthode de mesure.

Pièce jointe 3  
de l'Annexe 1  
  
Présentation du modèle

Selon la Recommandation UIT‑R BS.1116, on obtient une SDG pour un élément de test sonore lors d'un test d'écoute, et la SDG moyenne sur un certain nombre d'auditeurs représente la qualité subjective de l'élément. L'élément peut contenir divers types de distorsions du son, les variations de la qualité sont donc intégrées dans le temps. Ainsi, pour prédire les SDG sur la base de mesures physiques, il faut un modèle précis du système auditif périphérique ainsi que des aspects cognitifs du jugement de la qualité sonore.

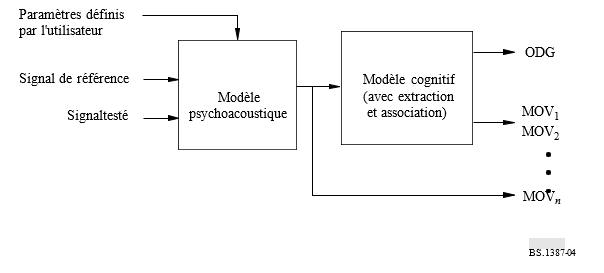
Le modèle recommandé pour la mesure objective produit un certain nombre de variables de sortie de modèle (MOV) fondées sur des comparaisons entre le Signal de référence et le Signal testé. Ces MOV sont transformées en une ODG avec une technique d'optimisation qui réduit le carré de la différence entre la distribution des ODG et la distribution correspondante des moyennes de SDG pour un ensemble de données suffisamment important.

Il y a deux variations sur le modèle – une version fondée sur la transformation de Fourier discrète (TFD) que l'on peut utiliser pour le contrôle en temps réel, et une version, fondée sur un banc de filtres et la TFD, qui donnent des résultats plus précis. La version TFD est la version de base, la version avancée y associe le banc de filtres.

La Figure 4 représente la structure haute des deux versions.

FIGURE 4

Étapes du traitement mises en œuvre dans le modèle



# 1 Traitement du son

Comme lors des essais d'écoute subjective, la qualité du signal testé est jugée par rapport au Signal de référence. Le Signal de référence et le Signal testé (signaux mono ou stéréo) sont transformés en une représentation psychoacoustique. On compare ces représentations afin d'obtenir une ODG. Ces opérations sont effectuées lors des étapes du traitement indiquées sur la Fig. 4.

## 1.1 Paramètres définis par l'utilisateur

La méthode de mesure requiert que le niveau d'écoute soit considéré comme étant un paramètre. Ainsi, l'utilisateur doit préciser un niveau de pression acoustique en dB SPL produit par une onde sinusoïdale pleine échelle de 1 019,5 Hz. Si le niveau d'écoute exact est inconnu, il est recommandé de supposer un niveau d'écoute de 92 dB SPL.

## 1.2 Modèle psychoacoustique

Le modèle psychoacoustique transforme les trames successives du signal en une représentation sur la membrane basilaire. Ce processus commence avec une TFD et un banc de filtres. La TFD transforme les données dans le domaine des fréquences, et le résultat est transformé de l'échelle des fréquences en échelle des hauteurs sonores, équivalent psychoacoustique de la fréquence. Dans le modèle à banc de filtres, le passage des fréquences aux hauteurs sonores est pris directement en compte par les largeurs de bandes et l'espacement des filtres passe-bande.

On utilise deux concepts différents pour parvenir à un masquage simultané. Certaines MOV sont calculées avec le *concept de seuil de masquage*, d'autres sont fondées sur une *comparaison des* *représentations internes*. Le premier concept calcule directement un seuil de masquage par l'intermédiaire de fonctions de masquage psychophysiques. Les MOV sont fondées sur l'écart entre le signal d'erreur physique et ce seuil de masquage. Dans le concept de comparaison des représentations internes, les énergies du Signal testé et du Signal de référence sont étendues aux hauteurs sonores adjacentes afin d'obtenir des schémas d'excitation. Les MOV sont fondées sur la comparaison de ces schémas d'excitation. Le masquage non simultané est mis en œuvre par rémanence des représentations du signal dans le temps.

Le seuil absolu est modélisé par application d'une fonction de pondération dépendant de la fréquence et par ajout, aux schémas d'excitation, d'un décalage dépendant de la fréquence. Ce seuil est une approximation de la pression audible minimum [ISO 389-7, Acoustics – Reference zero for the calibration of audiometric equipment – Part 7: Reference threshold of hearing under free-field and diffuse-field listening conditions, 1996].

Les résultats principaux du modèle psychoacoustique sont l'excitation et le seuil de masquage en fonction du temps et de la fréquence. Les résultats du modèle à plusieurs niveaux sont disponibles pour la suite du processus.

## 1.3 Modèle cognitif

Le modèle cognitif condense les informations à partir d'une séquence de trames produite par le modèle psychoacoustique. Les sources d'informations les plus importantes pour des mesures de qualité sont les différences entre le Signal de référence et le Signal testé dans le domaine fréquentiel et celui des hauteurs sonores. Dans le domaine fréquentiel, on mesure les largeurs de bandes spectrales des deux signaux, ainsi que la structure harmonique de l'erreur. Dans le domaine des hauteurs sonores, les mesures de l'erreur sont dérivées de la modulation de l'enveloppe de l'excitation, ainsi que de l'amplitude de l'excitation.

Les caractéristiques calculées sont pondérées, de façon à ce que leur association donne une ODG suffisamment proche de la SDG pour une certaine distorsion du son étudiée. La version de base utilise 11 caractéristiques pour produire une ODG, la version avancée en utilise cinq. L'optimisation a été obtenue grâce à l'algorithme d'apprentissage d'un réseau neuronal à propagation rétrograde (voir l'Annexe 2, § 6). Les données d'apprentissage comprennent la totalité des bases de données 1 et 2, et une partie de la base de données 3. Les données d'essai de généralisation ont été obtenues à partir de la fin de la base de données 3 et de toutes les données CRC97 (voir la Pièce jointe 2 de l'Annexe 2).

Pièce jointe 4  
de l'Annexe 1  
  
Principes et caractéristiques des méthodes de mesure  
objective de la qualité perceptuelle du son

# 1 Introduction et historique

La transmission numérique et le stockage des signaux audio sont de plus en plus fondés sur des algorithmes de réduction des données, qui sont adaptés aux propriétés du système auditif humain et reposent tout particulièrement sur des effets de masquage. Ces algorithmes ne visent pas essentiellement à réduire les distorsions, mais plutôt à gérer ces distorsions de façon à ce qu'elles soient le moins audibles possible. La qualité de ces encodeurs perceptuels ne peut plus être évaluée par les méthodes de mesure conventionnelles, qui déterminent normalement la valeur générale de la distorsion. Pour illustrer ces limitations, on cite souvent ce que l'on appelle le miracle des 13 dB: un bruit ajouté ayant une structure spectrale adaptée à celle du signal audio est quasiment inaudible même si le rapport *S*/*N* non pondéré qui en résulte descend jusqu'à 13 dB.

C'est pour cette raison que les évaluations des codecs perceptuels nécessitent des essais d'écoute afin d'évaluer la qualité du son. Une bonne fiabilité et une bonne répétabilité des essais d'écoute demandent beaucoup de temps et de travail.

Ces problèmes peuvent être surmontés grâce à des schémas de mesure objective qui intègrent le système auditif humain. C'est en 1979 que Schroeder, Atal et Hall publient cette idée [Schroeder et autres, 1979]. Leur article, qui concerne essentiellement le codage de la parole, décrit la procédure de la méthode de mesure «Intensité acoustique du bruit (NL, *noise loudness*)».

Dans cet article, les auteurs évaluent l'intensité perçue du signal de bruit du codec de la parole, c'est‑à‑dire la différence entre le signal d'entrée et le signal de sortie, pour chaque trame temporelle d'environ 20 ms. Si le signal de bruit est totalement masqué, l'intensité perçue est zéro. Un masquage partiel réduit l'intensité du signal de bruit non masqué. Le seuil de masquage utilisé est optimisé pour le bruit qui masque le son et on calcule pour chaque trame la dégradation finale de la parole. Il n'y a pas de calcul d'un résumé de la qualité totale d'un échantillon de parole.

En 1985, Karjalainen publie sa méthode de mesure «Différence auditive spectrale (ASD, *auditory spectral difference*)» [Karjalainen, 1985]. Il reprend plusieurs idées de Schroeder, Atal et Hall mais remplace l'analyse fondée sur la trame par un banc de filtres qui se recouvrent partiellement, change la méthode d'intégration du seuil absolu et ajoute un modèle de masquage temporel. Les deux signaux d'entrée dans le système de mesure sont traités exactement de la même façon, afin de produire une sorte de représentation interne. Ces représentations internes sont comparées les unes aux autres pour expliquer les différences perçues entre le signal d'entrée et le signal de sortie d'un schéma de codage de la parole. Il n'y a pas de calcul d'un résumé de la qualité totale d'un échantillon de parole. La résolution temporelle de ASD est mieux adaptée aux propriétés du système auditif humain, mais augmente la complexité de l'algorithme.

En 1987, Brandenburg publie la méthode de mesure «Rapport bruit/masque (NMR)» [Brandenburg, 1987], destiné à être un outil de développement de schémas de codage du son. La complexité de ce schéma est moindre que celle de NL: on calcule l'étalement sur les bandes perceptuelles avec une fonction d'étalement conçue comme une courbe du pire cas. Le seuil utilisé après masquage est optimisé pour le son qui masque le bruit. Il ajoute un schéma simple de modélisation du postmasquage et plusieurs méthodes d'évaluation de la qualité perçue d'extraits sonores plus longs. Ce schéma sera le premier à être mis en application dans des matériels de mesure en temps réel.

En 1989, Moore et Glasberg [Moore, 1989] présentent un modèle perceptuel mais ne proposent pas de méthode pour juger de la qualité perçue de signaux sonores dégradés.

# 2 Structure générale des méthodes de mesure objective de la qualité perceptuelle du son

Tous les schémas de mesure perceptuelle fonctionnent avec deux signaux d'entrée: l'un est le Signal de référence (REF), l'autre le Signal testé (SUT). Dans les cas où la référence ne peut être transmise à l'équipement de mesure, mais si ce signal est connu, le Signal de référence peut être une référence interne mémorisée dans l'équipement de mesure. Il est essentiel que les deux signaux soient temporellement alignés.

Il existe deux manières d'intégrer la psychoacoustique dans les schémas de mesure. La première est très proche de la structure des schémas de codage du son: on utilise le Signal de référence pour calculer une estimation du seuil de masquage réel (voir ci-dessous). La différence entre le Signal testé et le Signal de référence est alors comparée à ce seuil de masquage. On appelle cette méthode «concept de seuil de masquage»; elle est utilisée dans NL et NMR. On peut calculer la différence entre les signaux d'entrée soit dans le domaine temporel, soit comme étant la différence entre les spectres d'énergie de courte durée. Cette dernière méthode offre une meilleure protection contre les erreurs d'alignement temporel mais réduit la résolution temporelle. La différence dans le domaine temporel est généralement trop sensible aux distorsions de phase; c'est pourquoi on ne l'utilise plus.

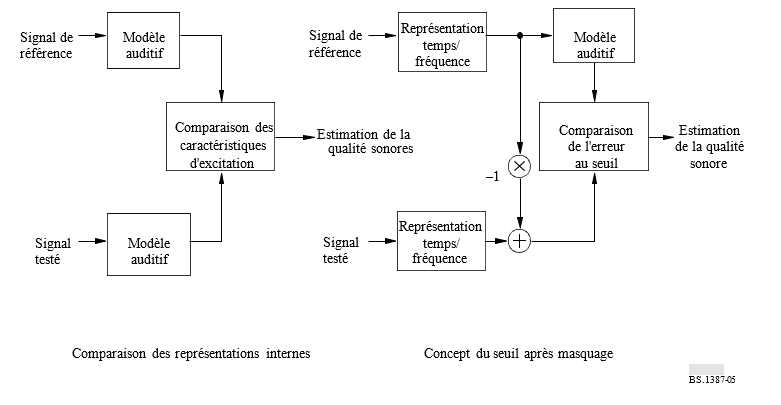
La seconde méthode est plus proche des processus physiologiques du système auditif humain: on calcule une représentation interne du Signal de référence et du Signal testé. Cette représentation interne est une estimation de l'information mise à la disposition du cerveau humain pour comparer les signaux. On appelle cette méthode «comparaison des représentations internes»; elle est utilisée dans ASD.

# 3 Considérations psychoacoustiques et cognitives

Ce paragraphe présente les propriétés du système auditif humain les plus importantes en ce qui concerne l'évaluation de la qualité perçue des signaux sonores. Le point le plus important est la méthode de modélisation de ces propriétés.

FIGURE 5

Concepts psychoacoustiques utilisés dans les différentes approches   
des schémas de mesure perceptuelle



## 3.1 Caractéristique de transfert par l'oreille externe et par l'oreille moyenne

Les signaux sonores doivent en général passer par l'oreille externe et par l'oreille moyenne avant d'atteindre l'oreille interne où ont lieu les processus de détection et d'analyse du son. L'oreille externe et l'oreille moyenne ont une fonction de filtrage passe-bande du signal d'entrée. Le bruit présent dans le nerf auditif, et le bruit causé par la circulation sanguine, sont ajoutés au signal d'entrée. L'amplitude de ce bruit décroît dans les basses fréquences. La fonction de transfert par l'oreille externe et par l'oreille moyenne et les bruits internes limitent la capacité à détecter les signaux sonores ténus, et ont une influence essentielle sur le seuil auditif absolu.

## 3.2 Échelles perceptuelles des fréquences

Dans l'oreille humaine, les récepteurs de la pression sonore sont les cils, situés dans l'oreille interne, plus précisément dans la cochlée. Dans la cochlée a lieu une transformation des fréquences en positions. La position de l'excitation maximum dépend de la fréquence du signal d'entrée. Chaque cil à une position donnée dans la cochlée est responsable d'une partie de l'échelle des fréquences; les différentes parties se recouvrent partiellement. L'impression perceptuelle de la hauteur du son est liée à la distance constante entre les cils.

Selon les expériences psychoacoustiques, on a découvert différentes fonctions de transformation de la fréquence en hauteur sonore:

On trouvera en [Zwicker et Feldtkeller, 1967] un tableau qui partage l'échelle des fréquences en Hz en 24 bandes qui ne se recouvrent pas, les bandes critiques. On trouvera dans le Tableau 5 les fréquences limites supérieures de ces bandes. Ce Tableau contient également une définition de l'échelle de Bark: 1 Bark correspond à 100 Hz, 24 Bark correspondent à 15 500 Hz.

TABLEAU 5

Échelle des bandes critiques définie par Zwicker

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Bande critique | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Fréquence limite supérieure (Hz) | 100 | 200 | 300 | 400 | 510 | 630 | 770 | 920 | 1 080 | 1 270 | 1 480 | 1 720 |
|  | | | | | | | | | | | | |
| Bande critique | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 |
| Fréquence limite supérieure (Hz) | 2 000 | 2 320 | 2 700 | 3 150 | 3 700 | 4 400 | 5 300 | 6 400 | 7 700 | 9 500 | 12 000 | 15 500 |

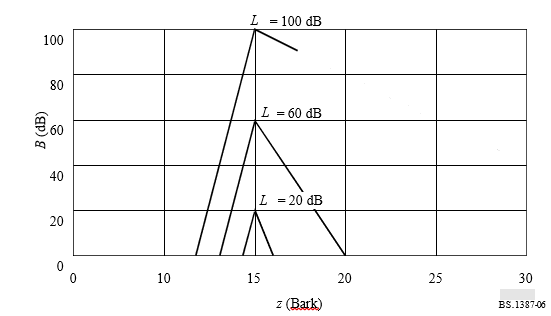
On a dans le passé trouvé plusieurs approximations de l'échelle de Bark. On trouvera en [Cohen et Fielder, 1992] une étude détaillée des différentes échelles. Dans le contexte de la mesure objective de la qualité du son perçu, c'est avec l'échelle de Bark que l'on obtient les meilleurs résultats.

## 3.3 Excitation

Chaque cil réagit à une gamme de fréquences que l'on peut décrire comme une caractéristique de filtre. C'est sur une échelle perceptuelle telle que décrite ci-dessus que l'on peut le mieux exprimer la pente des filtres. Sur une telle échelle, la forme des filtres est quasiment indépendante de la fréquence centrale. La pente inférieure de l'excitation est indépendante du niveau L du signal d'entrée (environ 27 dB/Bark). La pente supérieure est plus accentuée pour les niveaux bas que pour les niveaux élevés du signal d'entrée (*–*5 à *–*30 dB/Bark). Cette caractéristique de pente provient d'un mécanisme de feed-back entre deux types différents de cils, et il faut un certain temps pour qu'elle se stabilise. La meilleure résolution de fréquences auditives est donc atteinte avec des signaux stationnaires plusieurs millisecondes après le début du signal. Les caractéristiques d'excitation de signaux constitués de plusieurs composantes s'ajoutent de façon non linéaire.

FIGURE 6

Dépendances de l'excitation aux niveaux selon Terhardt [1979]



Après exposition à un signal, les cils et le traitement neuronal ont besoin d'un certain temps pour recouvrer toute leur sensibilité. La durée de ce processus de récupération dépend du niveau et de la durée du signal et peut durer jusqu'à plusieurs centaines de millisecondes. Les signaux de forte intensité sont traités plus rapidement que les signaux de faible intensité entre les cils et le cerveau. Le début d'un signal fort peut donc masquer un signal antérieur plus faible.

L'échelle ERB [Moore, 1986] propose une autre approche d'un modèle d'excitation qui utilise les filtres ROEX [Moore, 1986]. Dans le contexte de la mesure objective de la qualité du son perçu, on a obtenu les meilleurs résultats avec les modèles fondés sur [Zwicker y Feldtkeller, 1967] et [Terhardt, 1979].

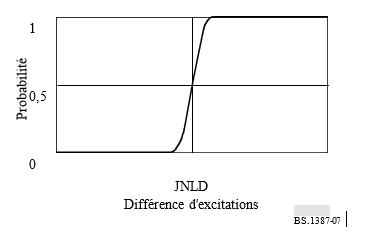
## 3.4 Détection

Les excitations des différents signaux sonores sont transmises au cerveau. Il existe trois types de mémoire qui diffèrent selon le degré de précision des détails et la durée pendant laquelle l'information est présente: une mémoire à long terme, une mémoire à court terme et une mémoire à ultra court terme. Dans le contexte des essais d'écoute, ce sont les mémoires à ultra court terme qui ont le rôle le plus important. La plupart des détails d'un signal sont conservés en mémoire si la durée de l'extrait sonore est inférieure à 5 à 8 s, selon l'auditeur et l'extrait sonore. Ce phénomène est pris en compte dans la procédure d'évaluation définie dans la Recommandation UIT‑R BS.1116 où les sujets des essais peuvent choisir d'écouter de façon plus précise de toutes petites parties d'un extrait sonore. Au niveau du seuil de détection, la probabilité de détection est de 50%. De part et d'autre du seuil, la probabilité de détection des différences passe lentement de 0% à 100%.

La différence de niveau tout juste audible (JNLD, *just-noticeable level difference*) est le seuil de détection des différences de niveau. La JNLD subit l'influence du niveau des signaux d'entrée. Pour des signaux faibles, la détection demande de grandes différences (niveau: 20 dBSPL, JNLD: 0,75 dB). Pour les signaux de forte intensité la sensibilité aux légères différences est beaucoup plus importante (niveau: 80 dBSPL, JNLD: 0,2 dB). Ces chiffres sont fondés sur des expériences en modulation d'amplitude.

FIGURE 7

Principe de la probabilité de détection



## 3.5 Masquage

Un signal parfaitement audible lorsqu'il est isolé peut être complètement inaudible en présence d'un autre signal, le masque. On appelle cet effet le masquage et le signal inaudible est le masqué. Il faut distinguer deux cas:

*– Masquage simultané*

Dans ce cas, le masque et le masqué sont présentés en même temps et sont quasi stationnaires. Si le masque a une largeur de bande discrète, le seuil auditif est élevé même pour des fréquences inférieures ou supérieures à celles du masque. La quantité de masquage dépend de la structure du masque et du masqué. Dans le cas où un signal semblable à un bruit masque un signal tonal, la quantité de masquage est presque indépendante de la fréquence. Si le niveau de pression sonore du masqué est d'environ 5 dB inférieure au niveau du masque, il devient inaudible. Dans le cas où un signal tonal masque un signal semblable à un bruit, la quantité de masquage dépend de la fréquence du masque. La formule  où *z* est la bande critique du masque, peut en donner une estimation. En outre, pour des signaux de haut niveau, des effets non linéaires réduisent le seuil de masquage aux alentours du masque. On retrouve des effets similaires si un signal tonal masque un autre signal tonal. Les seuils de masquage de plusieurs signaux s'ajoutent de façon non linéaire. Le seuil de masquage qui en résulte est généralement supérieur au seuil de masquage de chaque signal pris individuellement.

*– Masquage temporel*

Dans ce cas, le masque et le masqué sont présentés à des moments différents. Peu après la fin du son masquant, le seuil de masquage est plus proche du masquage simultané de ce masque que du seuil absolu. Selon la durée du masque, le temps d'affaiblissement du seuil peut être compris entre 5 ms (masque: impulsion gaussienne d'une durée d'environ 0,05 ms) et plus de 150 ms (masque: bruit rose d'une durée de 1 s). Les signaux faibles émis juste avant des signaux plus puissants sont masqués. La durée de cet effet de prémasquage est d'environ 5 ms. Si le masqué est juste au-dessus du seuil, il n'est pas perçu avant le masque mais comme une modification du masque. Le prémasquage montre de fortes déviations d'un auditeur à un autre.

## 3.6 Intensité acoustique et masquage partiel

L'intensité acoustique perçue de signaux sonores dépend de leur fréquence, de leur durée, et du niveau de pression sonore. Du fait de l'auto-masquage, l'intensité acoustique d'un signal complexe est moindre que la somme de l'intensité acoustique de toutes ses composantes. Dans le contexte de la mesure de la qualité d'un son, l'intensité acoustique de la distorsion indésirable ajoutée au Signal de référence, l'intensité acoustique du bruit, est réduite par le masquage partiel causé par le Signal de référence.

## 3.7 Acuité

L'acuité, une des principales composantes de la sensation, est liée au timbre. Un son est perçu comme étant aigu s'il contient principalement des composantes de hautes fréquences. Par exemple, un son sinusoïdal ou un bruit de bande limitée de hautes fréquences, ou un bruit passe-haut de fréquence limite supérieure à environ 3 kHz est dit aigu. Les détails de la structure fréquentielle du signal sonore, n'a cependant pas une influence majeure sur l'acuité. C'est Von Bismarck [1974] qui a mené la recherche fondamentale sur l'acuité.

En ce qui concerne l'acuité, des recherches complémentaires ont été effectuées en 1984 par W. Auras [Auras, 1984]. Le résultat de ces recherches est une légère modification de la fonction de pondération par rapport à la fonction de pondération définie par Bismark. En ce qui concerne l'évaluation de l'acuité, elle joue un rôle moins important dans les bandes critiques très basses et très hautes, et un rôle plus important dans les bandes critiques comprises entre 14 et 20 Bark. En outre, ces recherches ont montré que l'acuité des signaux sonores présentant une forte variation du niveau de pression sonore et un contenu important en fréquences élevées ne peut être fondée uniquement sur l'intensité acoustique, mais aussi sur une fonction de pondération, qui dépend de l'intensité acoustique globale.

## 3.8 Traitement cognitif

Il est évident que la qualité du son perçu est fortement influencée par les effets cognitifs. Une expérience simple permet de le démontrer.

Un Signal de référence comportant un bruit de fond clairement audible est traité par un équipement audio quelconque qui ne peut pas transmettre ce bruit de fond. Le bruit étant une distorsion indésirable, le Signal de référence serait moins bien noté que le signal traité lors d'un essai d'écoute. À l'opposé, c'est ce même signal traité qui serait moins bien noté si la partie la plus importante du Signal de référence était ce léger bruit de fond.

Donner une liste de tous les effets cognitifs possibles n'est pas du domaine de cette Pièce jointe, mais on peut en donner quelques exemples:

Exemple 1: Séparation des distorsions linéaires des distorsions non linéaires

Les distorsions linéaires sont moins gênantes que les distorsions non linéaires. La séparation des distorsions linéaires des distorsions non linéaires peut relativement facilement être mise en œuvre par un filtrage rétrograde adaptable du signal de sortie. La méthode définie dans cette Recommandation fait usage d'une séparation des distorsions linéaires des distorsions non linéaires.

Exemple 2: Analyse du paysage auditif

L'analyse du paysage auditif [Bregman, 1990] est un processus cognitif qui permet aux auditeurs de séparer différents événements auditifs et de les regrouper en différents objets. On trouvera en [Beerends et Stemerdink, 1994] une approche pragmatique utile pour quantifier un effet d'analyse du paysage auditif. Si une composante temps-fréquence n'est pas codée par un codec, le signal restant constitue toujours un paysage auditif cohérent, alors que l'introduction d'une nouvelle composante temps-fréquence sans relation avec le signal conduit à deux perceptions différentes. Du fait de ce partage en deux perceptions différentes, la distorsion sera plus gênante que ce qu'on aurait pu attendre sur la base de l'intensité acoustique de la nouvelle composante de distorsion introduite. Cela conduit à une asymétrie de perception entre la dégradation apportée par une distorsion causée par le non codage d'une composante temps‑fréquence et la dégradation causée par l'introduction d'une nouvelle composante temps‑fréquence.

Exemple 3: Masquage informationnel

On peut modéliser le masquage informationnel en définissant une mesure de la complexité spectro‑temporelle semblable à l'entropie. Cet effet dépend très probablement de la formation que les sujets ont reçue avant l'évaluation subjective. On trouvera une première tentative de modélisation de cet effet en [Beerends et autres, 1996], où est calculée une estimation de la complexité locale sur une fenêtre temporelle d'environ 100 ms. Si cette complexité locale est élevée, les distorsions dans cette fenêtre temporelle sont plus difficiles à entendre que lorsque la complexité locale est peu élevée. La formation peut réduire le seuil de masquage de plusieurs dizaines de dB [Leek et Watson, 1984].

Exemple 4: Pondération spectro-temporelle

Certaines régions spectro-temporelles du signal sonore sont porteuses de plus d'informations que d'autres, et peuvent, de ce fait, être plus importantes. On a découvert que la pondération spectro‑temporelle était importante pour des jugements de qualité de codecs de la parole. Dans le langage parlé, certaines composantes spectro-temporelles, par exemple les sons voyelles, sont de façon évidente porteuses de plus d'informations que d'autres [Beerends et Stemerdink, 1994]. À l'inverse, en musique, toutes les composantes spectro-temporelles du signal, même les silences, peuvent véhiculer de l'information.

# 4 Les modèles intégrés

## 4.1 DIX

La méthode de mesure perceptuelle DIX (*disturbance index* – indice de perturbation) [Thiede et Kabot, 1996] est fondée sur un banc de filtres auditifs qui offre une haute résolution temporelle et permet ainsi (par rapport aux approches fondées sur TFR) une modélisation plus précise d'effets temporels comme le prémasquage et le postmasquage. La structure temporelle fine des enveloppes de chaque filtre auditif est préservée et peut être utilisée pour obtenir des informations complémentaires sur les signaux et les distorsions introduites.

Les fréquences centrales de chaque filtre sont également réparties sur une échelle perceptuelle des hauteurs sonores. La partie supérieure du filtre est légèrement arrondie pour assurer que le nombre de filtres choisi couvre toute la gamme de fréquences sans ondulations dans la réponse en fréquence. Afin de modéliser les seuils de masquage, la pente des filtres décroît de façon exponentielle sur l'échelle de Bark. La pente des filtres dépend du niveau des signaux d'entrée. Dans la première version de DIX, la gamme des fréquences audibles était couverte par 80 filtres; ce chiffre sera ultérieurement réduit à 40, ce qui signifie que la résolution de fréquences correspond à environ 0,6 Bark. L'algorithme du banc de filtres est plutôt rapide par rapport à d'autres bancs de filtres comportant des filtres individuels, mais toujours plus gourmand en temps que des transformations fondées sur des blocs comme TFR et des transformations par ondelettes.

DIX adapte de façon dynamique les niveaux et les spectres entre le Signal testé et le Signal de référence afin de séparer les distorsions linéaires des distorsions non linéaires. Le système évalue la structure des enveloppes temporelles en sortie de filtres afin de modéliser l'augmentation du masquage due aux masques modulés et semblables à des bruits par rapport aux sons purs.

En comparant les représentations internes du Signal testé et du Signal de référence, on calcule de nombreux paramètres de sortie, dont l'intensité acoustique partielle des distorsions non linéaires, des indicateurs de la quantité de distorsions linéaires et des mesures des effets temporels et bilatéraux. On peut néanmoins parvenir à une bonne estimation de la qualité du son de base en n'utilisant que deux des paramètres de sortie: l'intensité acoustique partielle des distorsions non linéaires et un des indicateurs de quantité de distorsions linéaires permettent d'obtenir une estimation de la qualité sonore de base du Signal testé.

## 4.2 NMR

La méthode de mesure NMR (*noise-to-masked-ratio* *–* rapport bruit/masque) [Brandenburg, 1987] évalue la différence de niveaux entre le seuil de masquage et le signal de bruit. Pour analyser le contenu fréquentiel du signal, on utilise une TFD et une fenêtre de Hann d'environ 20 ms. Les coefficients de transformation sont associés aux bandes conformément à l'échelle de Bark. On estime le seuil de masquage pour chaque bande. La pente du seuil de masquage est obtenue par une approche du pire cas en prenant en compte le fait que les pentes sont plus accentuées pour les signaux faibles mais atteignent le seuil absolu pour des niveaux plus élevés. Le seuil absolu est adapté à la résolution du signal d'entrée (généralement 16 bits), mais pas aux exigences psychoacoustiques. Grâce à cela, NMR n'est pas affectée par les changements du niveau de reproduction. La résolution de l'échelle des hauteurs sonores est d'environ 1 Bark. La puissance de calcul nécessaire étant faible, il a été possible de mettre NMR en service comme système en temps réel dès les premières étapes de son développement.

Ce modèle, utilisé depuis 1987, a apporté la preuve de sa fiabilité de base.

Les valeurs de sortie les plus importantes fournies par NMR sont le taux d'indicateurs de masquage, le pourcentage de trames présentant des distorsions audibles, ainsi que le NMR total et moyen, qui sont autant de méthodes pour obtenir l'écart moyen entre l'énergie d'erreur et le seuil après masquage.

## 4.3 OASE

Le système de mesure OASE (*objective audio signal evaluation* – Evaluation objective du signal audio) [Sporer, 1997] utilise un banc de filtres composé de 241 filtres pour analyser les signaux d'entrée. Les fréquences centrales sont également espacées sur l'échelle de Bark avec un écart de 0,1 Bark. Les filtres se recouvrent partiellement. Chaque filtre est adapté à la réponse en fréquence d'un point de la membrane basilaire. La dépendance des pentes au niveau est intégrée par une approche du pire cas, comme dans NMR. Les filtres des fréquences centrales basses nécessitent un calcul au taux d'échantillonnage complet, alors que les filtres des fréquences centrales hautes peuvent être calculés avec un taux d'échantillonnage réduit. Après les filtres, on calcule un modèle des effets temporels du système auditif humain comme dans ASD. Ensuite, il est possible de réduire le taux d'échantillonnage pour toutes les bandes de filtres. Cela conduit à une résolution temporelle du banc de filtres de 0,66 ms pour un taux d'échantillonnage de 48 kHz. Les sorties des filtres pour le signal référence et le signal testé sont comparées avec une fonction de probabilité de détection. Pour calculer la JNLD, cette fonction utilise comme entrée l'intensité acoustique des signaux d'entrée. La probabilité de détection totale est dérivée de la probabilité de détection de chaque bande. On effectue cette opération pour les deux canaux d'entrée, ainsi que pour le canal dit central. La probabilité de détection dans le canal central pour chaque bande est le pire cas de la probabilité de détection du canal de gauche et du canal de droite. Pour chaque trame de 0,66 ms la somme des pas au‑dessus du seuil est également calculée.

Il existe plusieurs méthodes pour obtenir la moyenne temporelle de la probabilité de détection et les pas au-dessus du seuil:

– la moyenne temporelle de la probabilité de détection;

– la fréquence des trames de probabilité de détection supérieure à 0,5;

– la probabilité de détection maximum avec filtres passe-bas;

– la probabilité de détection maximum avec filtres passe-bas après oubli;

– le nombre moyen de pas au-dessus du seuil pour les trames de probabilité de détection supérieure à 0,5;

– le nombre moyen de pas au-dessus du seuil;

– le nombre maximum de pas au-dessus du seuil;

– le nombre moyen de pas au-dessus du seuil de 10% des moins bonnes trames.

## 4.4 PAQM (perceptual audio quality measure)

Le principe de base de PAQM [Beerends et Stemerdink, 1992] est de soustraire les représentations internes (les représentations dans le cerveau du sujet) du signal de référence et du signal dégradé et de représenter cette différence sur un schéma cognitif de la qualité du son perçu de façon subjective. Le passage du domaine physique externe au domaine psychophysique interne se fait par l'intermédiaire de quatre opérations:

– une représentation temps-fréquence par l'intermédiaire d'une TFD avec une fenêtre de Hann d'une durée d'environ 40 ms;

– une déformation des fréquences avec l'échelle de Bark;

– un étalement temps-fréquence (convolution non linéaire);

– une déformation d'intensité (compression).

L'association de la dégradation et de la compression permet de modéliser le comportement de masquage du système auditif humain au niveau du seuil de masquage et au-dessus de celui-ci. La compression est optimisée avec les résultats subjectifs de la première évaluation du codec audio MPEG (ISO/IEC JTC 1/SC 2/WG 11 MPEG/Audio test report, Document MPEG90/N0030, octobre 1990) (ISO/IEC JTC 1/SC 2/WG 11 MPEG/Audio test report, Document MPEG91/N0010, juin 1991). La différence dans la représentation interne est exprimée en termes de gêne créée par le bruit. Dans les dernières versions de PAQM, soumises à l'UIT-R, deux effets cognitifs sont intégrés à la représentation de la dégradation créée par le bruit sur la qualité subjective, le flux perceptuel [Beerends et Stemerdink, 1994] et le masquage informationnel [Beerends et autres, 1996].

Une version simplifiée de PAQM, la mesure perceptuelle de la qualité de la voix, PSQM [Beerends et Stemerdink, 1994] fut développée avec un modèle cognitif présenté en [Beerends et Stemerdink, 1994] mais étendue par une pondération des intervalles de silence. Lors du développement de PSQM, il apparut que lors du jugement de la qualité de la voix dans un contexte de téléphonie, le bruit qui apparaît lors des silences est moins important que le bruit qui apparaît dans les intervalles parlés. Lors d'un test de performance effectué par l'UIT-T, la proposition PSQM a montré la plus forte corrélation entre la qualité objective et la qualité subjective (Commission d'études 12 du Secteur de la normalisation des télécommunications, COM 12-74 – Examen des essais de validation pour la mesure objective de la qualité de la voix). La Recommandation UIT-T P.862 décrit la mesure perceptuelle de la qualité de la voix et contient un algorithme d'évaluation de la qualité vocale amélioré.

## 4.5 PERCEVAL

PERCEVAL (*PERCeptual EVALuation* – Evaluation Perceptuelle) [Paillard et autres, 1992] modélise les caractéristiques de transfert par l'oreille moyenne et par l'oreille interne pour constituer une représentation interne du signal. Le signal d'entrée est décomposé en une représentation temps‑fréquence avec une TFD. Typiquement, une fenêtre de Hann d'environ 40 ms est appliquée aux données d'entrée, avec un recouvrement de 50% des fenêtres successives. Le spectre d'énergie est multiplié par une fonction dépendant de la fréquence qui modélise l'effet du canal auditif et de l'oreille moyenne. Les valeurs atténuées de l'énergie spectrale sont transformées de l'échelle des fréquences sur l'échelle de hauteurs sonores, qui est plus linéaire en ce qui concerne les propriétés physiques de l'oreille interne et les effets psychophysiques observés. Les composantes transformées de l'énergie sont alors convoluées avec une fonction d'étalement pour simuler la dispersion de l'énergie sur la membrane basilaire. Enfin, une énergie intrinsèque dépendant de la fréquence est ajoutée à chaque composante de la hauteur sonore pour prendre en compte le seuil d'audition absolu. La conversion de l'énergie en décibels a pour résultat la représentation du signal sur la membrane basilaire.

Lors de simulations d'expériences de masquage auditif, une représentation sur membrane basilaire est créée pour chaque stimulus, et la différence entre les représentations est l'information disponible pour effectuer le travail. Une des représentations est celle du masque seul, l'autre est celle de l'association du masque et du signal d'essai. La différence est la composante du signal qui n'est pas masquée. PERCEVAL calcule la probabilité de détection de cette différence. La probabilité de non‑détection de la différence pour chaque détecteur le long de la membrane basilaire stimulée est estimée par une fonction de probabilité sigmoïde. En supposant que les détecteurs sont statistiquement indépendants, la probabilité globale de détection pour l'ensemble des détecteurs est calculée en prenant le complément du produit des probabilités individuelles de non‑détection. Plusieurs expériences de masquage ont pu être simulées avec succès grâce à cette approche, et ce modèle a servi à évaluer la faisabilité de la modélisation individuelle d'auditeurs [Treurniet, 1996].

Outil d'estimation de la qualité du son, PERCEVAL calcule la différence entre les représentations du Signal de référence et du Signal testé. En appliquant des hypothèses raisonnables concernant les processus perceptuels et cognitifs de haut niveau, un certain nombre de variables pertinentes quant à la perception sont calculées et représentées par rapport à une mesure de la qualité objective du Signal testé. Cette représentation a été optimisée en minimisant la différence entre la distribution de la qualité objective et la distribution correspondante des taux moyens de qualité subjective pour l'ensemble des données disponibles.

## 4.6 POM

Le but de la mesure perceptuelle objective (POM) [Colomes et autres, 1995] est de quantifier la dégradation qui peut apparaître entre un Signal de référence et sa version dégradée. On y parvient en comparant la représentation basilaire interne des deux signaux, quelle que soit la cause de la dégradation. La représentation basilaire modélise les divers processus auxquels est soumis un signal sonore lorsqu'il est véhiculé dans l'oreille humaine. Ainsi, la première étape de POM est le calcul de la représentation interne d'un signal sonore. Les caractéristiques d'excitation (données en dB), sur la membrane basilaire, ont été choisies pour modéliser le taux de décharge des potentiels d'action dans les neurones le long de la membrane basilaire.

On appelle oreille artificielle ce processus de calcul des caractéristiques de l'excitation. Ensuite, lorsque l'on a obtenu les deux représentations internes des signaux à comparer, POM doit déterminer si la différence entre leurs représentations internes est audible ou non, et si oui, de quelle façon. C'est ce que l'on appelle le processus de détection.

POM utilise une TFD avec une fenêtre de Hann d'une durée d'environ 40 ms (avec un recouvrement de 50% entre deux fenêtres de Hann). Le nombre de canaux basilaires d'analyse est de 620. Les autres parties du modèle auditif sont presque identiques à celles utilisées soit par PAQM soit par PERCEVAL.

La fonction d'étalement est très précisément décrite par une approximation plus précise qui prend en compte la dépendance aux niveaux selon [Terhardt, 1979] et la forme arrondie selon [Schroeder et autres, 1979].

Ce modèle permet d'obtenir la probabilité de détection d'une distorsion entre les deux signaux comparés, ainsi que la distance dite basilaire qui représente l'écart de perception entre les deux excitations comparées.

## 4.7 L'approche Toolbox

Toolbox est une approche en trois étapes de la mesure de la distance perçue en qualité sonore d'un signal sonore d'essai par rapport à une Signal sonore de référence, permettant ainsi d'obtenir une indication du niveau de qualité sonore subjective du signal d'essai. Cette méthode est fondée sur des modèles perceptuels bien connus qui sont utilisés pour décrire la représentation perceptuelle des différences entre les deux signaux sonores. En outre, elle intègre une procédure de pondération pour la qualité sonore perçue d'un signal d'essai stéréo, qui prend en compte les résultats du canal de gauche et du canal de droite. Une corrélation stricte sur une base échantillon par échantillon du Signal de référence et du Signal testé n'est pas nécessaire.

La principale fonctionnalité de Toolbox, étape 1, est fondée sur le calcul de l'intensité acoustique spécifique, selon [Zwicker et Feldtkeller, 1967], avec un TFR de 2 048 points, une fenêtre de Hann, qui correspond à une durée d'environ 40 ms. La fenêtre est déplacée par incréments de 10 ms. En outre, des effets de masquage temporel, postmasquage et prémasquage, selon Zwicker, sont appliqués. À partir de ces valeurs de base de la sensation, d'autres paramètres perceptuels, comme l'intensité intégrée, l'intensité partiellement masquée, l'acuité, selon [von Bismarck, 1974] et [Aures, 1984] et la quantité de prééchos sont calculés; il s'agit d'une étape de prétraitement dont les résultats sont utilisés pour les étapes suivantes.

La seconde étape de Toolbox intègre des procédures de pondération qui dépendent essentiellement de la quantité de différences perçues dans l'intensité et de la variation de l'intensité dans le temps.

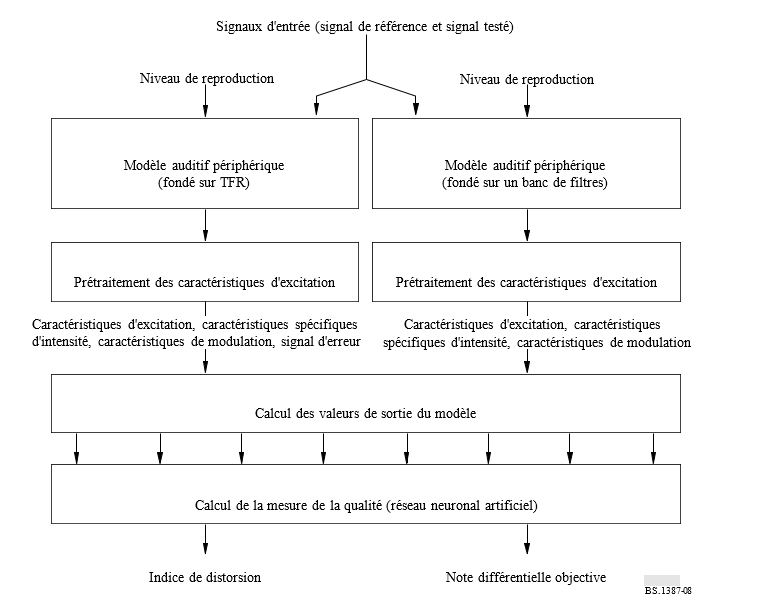
La troisième étape de Toolbox intègre la génération d'un ensemble de valeurs de sortie qui sont des outils intermédiaires et sont fondées sur une analyse statistique des valeurs obtenues lors des étapes 1 et 2. Les résultats de cette analyse statistique comportent les valeurs moyenne, maximum et de r.m.s., ainsi que l'écart type des valeurs moyennes. Une somme pondérée de ces valeurs intermédiaires est utilisée pour la détermination finale de l'écart perceptuel entre le Signal testé et le Signal de référence. Si nécessaire, cette valeur de sortie unique peut être mise en concordance avec la note différentielle subjective, obtenue généralement lors d'essais d'écoute subjectifs, en faisant correspondre de façon séquentielle les données de sortie pour chaque incrément temporel avec des fonctions linéaires ou polynomiales d'ordre plus élevé.

Annexe 2  
  
Description du modèle[[2]](#footnote-2)2

# 1 Généralités

FIGURE 8

Diagramme générique du système de mesure



Cette *méthode de mesure objective de la qualité du son perçu* est composée d'un *modèle auditif périphérique*, de plusieurs étapes intermédiaires (appelées ici «*prétraitement des caractéristiques d'excitation*»), du calcul (principalement) des *variables de sortie de modèle* («MOV») et d'une représentation d'un ensemble de variables de sortie de modèle par une valeur unique représentant la qualité audio de base du Signal testé. Elle comprend deux modèles auditifs, l'un fondé sur TFR et l'autre sur un banc de filtres. À l'exception du calcul du signal d'erreur (qui n'est utilisé qu'avec la partie du modèle fondée sur TFR), la structure générale des deux modèles est semblable.

Les entrées pour le calcul des MOV sont:

– Les caractéristiques d'excitation du Signal testé et du Signal de référence.

– Les caractéristiques d'excitation adaptés spectralement pour le Signal testé et le Signal de référence.

– Les caractéristiques spécifiques d'intensité pour le Signal testé et le Signal de référence.

– Les caractéristiques de modulation pour le Signal testé et le Signal de référence.

– Le signal d'erreur calculé comme étant la différence spectrale entre le Signal testé et le Signal de référence (uniquement pour le modèle fondé sur TFR).

Sauf indications contraires, dans le cas des signaux stéréo, tous les calculs sont effectués indépendamment et de la même façon pour le canal de gauche et le canal de droite.

Cette description définit deux ensembles, l'un appelé «*version de base*» et l'autre «*version avancée*».

Dans toutes les équations, l'indice «*Ref.*» représente toutes les caractéristiques calculées à partir du Signal de référence, et l'indice «*Test*» toutes les caractéristiques calculées à partir du Signal testé. L'indice «*k*» représente la variable de fréquence discrète (la bande de fréquences) et «*n*» la variable temporelle discrète (soit le compteur de trames, soit le compteur d'échantillons). Si les valeurs de *k* ou *n* ne sont pas explicitement définies, les calculs doivent être effectués pour toutes les valeurs possibles de *k* et *n*. Toutes les autres abréviations sont expliquées à leur première occurrence.

Dans les noms des MOV, l'indice «*A*» représente toutes les variables calculées avec la partie du modèle auditif fondé sur le banc de filtres, et l'indice «*B*» toutes les variables calculées avec la partie du modèle auditif fondé sur TFR.

## 1.1 Version de base

La *version de base* n'intègre que les MOV calculées à partir du modèle TFR. On n'utilise pas la partie fondée sur le banc de filtres. La *version de base* utilise un total de 11 MOV pour la prédiction de la *qualité audio de base* perçue.

## 1.2 Version avancée

La *version avancée* intègre des MOV calculées à partir du modèle fondé sur le banc de filtres ainsi que des MOV calculées à partir du modèle TFR. Les caractéristiques d'excitation adaptées spectralement et les caractéristiques de modulation ne sont calculées qu'à partir de la partie fondée sur le banc de filtres du modèle. La *version avancée* utilise 5 MOV pour la prédiction de la *qualité audio de base* perçue.

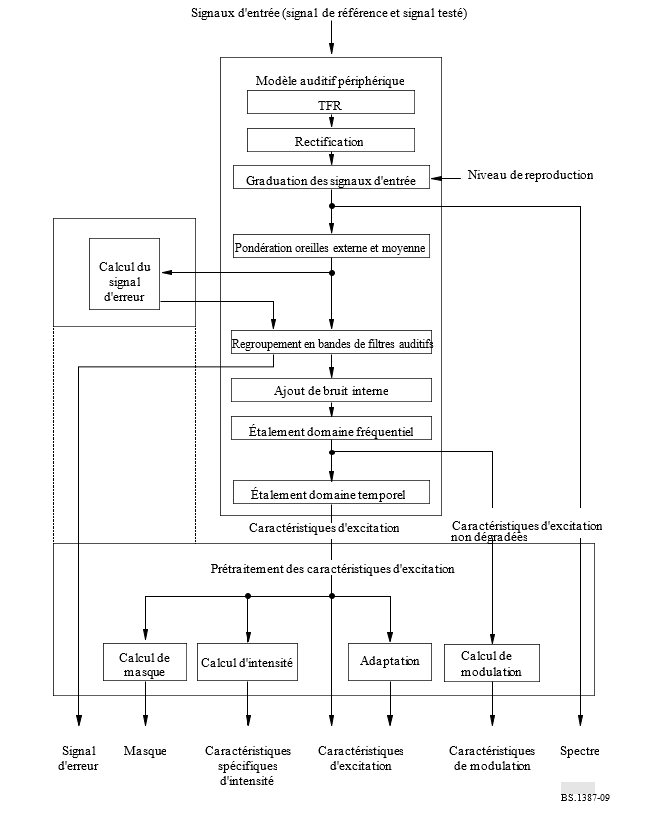
# 2 Modèle auditif périphérique

## 2.1 Modèle auditif TFR

### 2.1.1 Aperçu général

FIGURE 9

Modèle auditif périphérique et prétraitement des caractéristiques  
d'excitation pour la partie du modèle TFR



L'entrée du modèle auditif TFR, Signal de référence et signal d'essai échantillonnés à 48 kHz et alignés temporellement, est découpée en trames d'environ 0,042 s avec un recouvrement de 50%. Chaque trame est transformée dans le domaine fréquentiel avec une fenêtre de Hann et une TFR à court terme, et ajuste au niveau de reproduction. Une fonction de pondération est appliquée aux coefficients du spectre, laquelle modélise la réponse en fréquence de l'oreille externe et de l'oreille moyenne. La transformation en une représentation dans le domaine des hauteurs sonores est obtenue en regroupant les coefficients pondérés du spectre en bandes critiques. Un décalage dépendant de la fréquence est ajouté pour simuler le bruit interne du système auditif. Une fonction d'étalement dépendante du niveau est utilisée pour modéliser les filtres auditifs de spectre dans le domaine fréquentiel. Ensuite, un étalement dans le domaine temporel rend compte des effets de prémasquage.

Les *caractéristiques* *d'excitation* obtenues sont utilisées pour calculer les *caractéristiques* *spécifiques d'intensité* et les *caractéristiques* *de masquage.* Les caractéristiques précédant l'étalement final dans le domaine temporel («*caractéristiques* *d'excitation non dégradée*») sont utilisées pour calculer *les caractéristiques* *de modulation.*

Pour modéliser le signal d'erreur, les caractéristiques des signaux de référence et d'essai de sortie de filtre de l'oreille externe et moyenne sont associées et représentées sur l'échelle des hauteurs sonores en les regroupant en bandes critiques.

Ces sorties sont utilisées avec les caractéristiques d'excitation pour calculer les valeurs des variables de sortie de modèle.

### 2.1.2 Traitement temporel

L'entrée du modèle auditif TFR, signal d'essai et Signal de référence, est découpée en trames de 2 048 échantillons avec un recouvrement de 1 024 échantillons:

 (1)

où *n* est le numéro de la trame temporelle et *kt* un compteur de temps dans une trame.

### 2.1.3 TFR

Le passage du domaine temporel au domaine fréquentiel se fait avec une fenêtre de Hann:

 (2)

 (3)

suivi par une transformation de Fourier à court terme:

 (4)

Le facteur d'échelle de la TFR est calculé à partir du niveau de pression sonore supposé *Lp* d'une onde sinusoïdale à pleine échelle par l'équation:

 (5)

 (6)

où le facteur de normalisation *Norm* est calculé en prenant pour signal d'entrée une onde sinusoïdale de 1 019,5 Hz et 0 dB à pleine échelle et en calculant la valeur absolue maximum des coefficients de spectre sur 10 trames.

Si le niveau de pression sonore est inconnu, nous recommandons *Lp*  92 dBSPL.

### 2.1.4 Oreille externe et oreille moyenne

La réponse en fréquence de l'oreille externe et de l'oreille moyenne est modélisée par une fonction de pondération dépendante de la fréquence:

 (7)

où:

 (8)

est la représentation de la fréquence sur la ligne *k* appliquée aux sorties TFR (équation (9)).

 (9)

*Fe*[*kf*] sont les «*sorties TFR pondérées de l'oreille externe*».

### 2.1.5 Regroupement en bandes critiques

L'échelle auditive des hauteurs sonores est calculée à partir d'une approximation donnée par [Schroeder et autres, 1979].

 (10)

Les unités de hauteur sonore sont appelées *Bark* (bien que cette échelle ne représente pas exactement l'échelle de Bark définie par [Zwicker et Feldtkeller, 1967]).

Les limites de fréquence des filtres sont dans une fourchette située entre 80 Hz et 18 000 Hz. La largeur et l'espacement des bandes de filtres correspondent à une résolution de *res**0,25* Bark pour la version de base et *res**0,5* Bark pour la version avancée.

Ce qui conduit à un nombre de bandes de fréquences *Z**109* pour la version de base et *Z**55* pour la version avancée.

TABLEAU 6

Bandes de fréquences du modèle auditif TFR dans la version de base

| Groupe | Fréquence basse/Hz | Fréquence centrale/Hz | Fréquence haute/Hz | Largeur de fréquences/Hz |
| --- | --- | --- | --- | --- |

| *k* | *fl[k]* | *fc[k]* | *fu[k]* | *fw[k]* |
| --- | --- | --- | --- | --- |
| 0 | 80 | 91,708 | 103,445 | 23,445 |
| 1 | 103,445 | 115,216 | 127,023 | 23,577 |
| 2 | 127,023 | 138,87 | 150,762 | 23,739 |
| 3 | 150,762 | 162,702 | 174,694 | 23,932 |
| 4 | 174,694 | 186,742 | 198,849 | 24,155 |
| 5 | 198,849 | 211,019 | 223,257 | 24,408 |
| 6 | 223,257 | 235,566 | 247,95 | 24,693 |
| 7 | 247,95 | 260,413 | 272,959 | 25,009 |
| 8 | 272,959 | 285,593 | 298,317 | 25,358 |
| 9 | 298,317 | 311,136 | 324,055 | 25,738 |
| 10 | 324,055 | 337,077 | 350,207 | 26,151 |
| 11 | 350,207 | 363,448 | 376,805 | 26,598 |
| 12 | 376,805 | 390,282 | 403,884 | 27,079 |
| 13 | 403,884 | 417,614 | 431,478 | 27,594 |
| 14 | 431,478 | 445,479 | 459,622 | 28,145 |
| 15 | 459,622 | 473,912 | 488,353 | 28,731 |
| 16 | 488,353 | 502,95 | 517,707 | 29,354 |
| 17 | 517,707 | 532,629 | 547,721 | 30,014 |
| 18 | 547,721 | 562,988 | 578,434 | 30,713 |
| 19 | 578,434 | 594,065 | 609,885 | 31,451 |
| 20 | 609,885 | 625,899 | 642,114 | 32,229 |
| 21 | 642,114 | 658,533 | 675,161 | 33,048 |
| 22 | 675,161 | 692,006 | 709,071 | 33,909 |
| 23 | 709,071 | 726,362 | 743,884 | 34,814 |
| 24 | 743,884 | 761,644 | 779,647 | 35,763 |
| 25 | 779,647 | 797,898 | 816,404 | 36,757 |
| 26 | 816,404 | 835,17 | 854,203 | 37,799 |
| 27 | 854,203 | 873,508 | 893,091 | 38,888 |
| 28 | 893,091 | 912,959 | 933,119 | 40,028 |
| 29 | 933,119 | 953,576 | 974,336 | 41,218 |
| 30 | 974,336 | 995,408 | 1016,797 | 42,461 |
| 31 | 1016,797 | 1038,511 | 1060,555 | 43,758 |
| 32 | 1060,555 | 1082,938 | 1105,666 | 45,111 |

TABLEAU 6 (*suite*)

| Groupe | Fréquence basse/Hz | Fréquence centrale/Hz | Fréquence haute/Hz | Largeur de fréquences/Hz |
| --- | --- | --- | --- | --- |

| *k* | *fl[k]* | *fc[k]* | *fu[k]* | *fw[k]* |
| --- | --- | --- | --- | --- |
| 33 | 1105,666 | 1128,746 | 1152,187 | 46,521 |
| 34 | 1152,187 | 1175,995 | 1200,178 | 47,991 |
| 35 | 1200,178 | 1224,744 | 1249,7 | 49,522 |
| 36 | 1249,7 | 1275,055 | 1300,816 | 51,116 |
| 37 | 1300,816 | 1326,992 | 1353,592 | 52,776 |
| 38 | 1353,592 | 1380,623 | 1408,094 | 54,502 |
| 39 | 1408,094 | 1436,014 | 1464,392 | 56,298 |
| 40 | 1464,392 | 1493,237 | 1522,559 | 58,167 |
| 41 | 1522,559 | 1552,366 | 1582,668 | 60,109 |
| 42 | 1582,668 | 1613,474 | 1644,795 | 62,128 |
| 43 | 1644,795 | 1676,641 | 1709,021 | 64,226 |
| 44 | 1709,021 | 1741,946 | 1775,427 | 66,406 |
| 45 | 1775,427 | 1809,474 | 1844,098 | 68,671 |
| 46 | 1844,098 | 1879,31 | 1915,121 | 71,023 |
| 47 | 1915,121 | 1951,543 | 1988,587 | 73,466 |
| 48 | 1988,587 | 2026,266 | 2064,59 | 76,003 |
| 49 | 2064,59 | 2103,573 | 2143,227 | 78,637 |
| 50 | 2143,227 | 2183,564 | 2224,597 | 81,371 |
| 51 | 2224,597 | 2266,34 | 2308,806 | 84,208 |
| 52 | 2308,806 | 2352,008 | 2395,959 | 87,154 |
| 53 | 2395,959 | 2440,675 | 2486,169 | 90,21 |
| 54 | 2486,169 | 2532,456 | 2579,551 | 93,382 |
| 55 | 2579,551 | 2627,468 | 2676,223 | 96,672 |
| 56 | 2676,223 | 2725,832 | 2776,309 | 100,086 |
| 57 | 2776,309 | 2827,672 | 2879,937 | 103,627 |
| 58 | 2879,937 | 2933,12 | 2987,238 | 107,302 |
| 59 | 2987,238 | 3042,309 | 3098,35 | 111,112 |
| 60 | 3098,35 | 3155,379 | 3213,415 | 115,065 |
| 61 | 3213,415 | 3272,475 | 3332,579 | 119,164 |
| 62 | 3332,579 | 3393,745 | 3455,993 | 123,415 |
| 63 | 3455,993 | 3519,344 | 3583,817 | 127,823 |
| 64 | 3583,817 | 3649,432 | 3716,212 | 132,395 |
| 65 | 3716,212 | 3784,176 | 3853,348 | 137,136 |
| 66 | 3853,348 | 3923,748 | 3995,399 | 142,051 |
| 67 | 3995,399 | 4068,324 | 4142,547 | 147,148 |
| 68 | 4142,547 | 4218,09 | 4294,979 | 152,432 |
| 69 | 4294,979 | 4373,237 | 4452,89 | 157,911 |
| 70 | 4452,89 | 4533,963 | 4616,482 | 163,592 |
| 71 | 4616,482 | 4700,473 | 4785,962 | 169,48 |
| 72 | 4785,962 | 4872,978 | 4961,548 | 175,585 |
| 73 | 4961,548 | 5051,7 | 5143,463 | 181,915 |

TABLEAU 6 (*fin*)

| Groupe | Fréquence basse/Hz | Fréquence centrale/Hz | Fréquence haute/Hz | Largeur de fréquences/Hz |
| --- | --- | --- | --- | --- |

|  | *fl[k]* | *fc[k]* | *fu[k]* | *fw[k]* |
| --- | --- | --- | --- | --- |
| 74 | 5143,463 | 5236,866 | 5331,939 | 188,476 |
| 75 | 5331,939 | 5428,712 | 5527,217 | 195,278 |
| 76 | 5527,217 | 5627,484 | 5729,545 | 202,329 |
| 77 | 5729,545 | 5833,434 | 5939,183 | 209,637 |
| 78 | 5939,183 | 6046,825 | 6156,396 | 217,214 |
| 79 | 6156,396 | 6267,931 | 6381,463 | 225,067 |
| 80 | 6381,463 | 6497,031 | 6614,671 | 233,208 |
| 81 | 6614,671 | 6734,42 | 6856,316 | 241,646 |
| 82 | 6856,316 | 6980,399 | 7106,708 | 250,392 |
| 83 | 7106,708 | 7235,284 | 7366,166 | 259,458 |
| 84 | 7366,166 | 7499,397 | 7635,02 | 268,854 |
| 85 | 7635,02 | 7773,077 | 7913,614 | 278,594 |
| 86 | 7913,614 | 8056,673 | 8202,302 | 288,688 |
| 87 | 8202,302 | 8350,547 | 8501,454 | 299,152 |
| 88 | 8501,454 | 8655,072 | 8811,45 | 309,996 |
| 89 | 8811,45 | 8970,639 | 9132,688 | 321,237 |
| 90 | 9132,688 | 9297,648 | 9465,574 | 332,887 |
| 91 | 9465,574 | 9636,52 | 9810,536 | 344,962 |
| 92 | 9810,536 | 9987,683 | 10168,013 | 357,477 |
| 93 | 10168,013 | 10351,586 | 10538,46 | 370,447 |
| 94 | 10538,46 | 10728,695 | 10922,351 | 383,891 |
| 95 | 10922,351 | 11119,49 | 11320,175 | 397,824 |
| 96 | 11320,175 | 11524,47 | 11732,438 | 412,264 |
| 97 | 11732,438 | 11944,149 | 12159,67 | 427,231 |
| 98 | 12159,67 | 12379,066 | 12602,412 | 442,742 |
| 99 | 12602,412 | 12829,775 | 13061,229 | 458,817 |
| 100 | 13061,229 | 13296,85 | 13536,71 | 475,48 |
| 101 | 13536,71 | 13780,887 | 14029,458 | 492,748 |
| 102 | 14029,458 | 14282,503 | 14540,103 | 510,645 |
| 103 | 14540,103 | 14802,338 | 15069,295 | 529,192 |
| 104 | 15069,295 | 15341,057 | 15617,71 | 548,415 |
| 105 | 15617,71 | 15899,345 | 16186,049 | 568,339 |
| 106 | 16186,049 | 16477,914 | 16775,035 | 588,986 |
| 107 | 16775,035 | 17077,504 | 17385,42 | 610,385 |
| 108 | 17385,42 | 17690,045 | 18000 | 614,58 |

TABLEAU 7

Bandes de fréquences du modèle auditif TFR dans la version avancée

| Groupe | Fréquence basse/Hz | Fréquence centrale/Hz | Fréquence haute/Hz | Largeur de fréquences/Hz |
| --- | --- | --- | --- | --- |

| *k* | *fl[k]* | *fc[k]* | *fu[k]* | *fw[k]* |
| --- | --- | --- | --- | --- |
| 0 | 80 | 103,445 | 127,023 | 47,023 |
| 1 | 127,023 | 150,762 | 174,694 | 47,671 |
| 2 | 174,694 | 198,849 | 223,257 | 48,563 |
| 3 | 223,257 | 247,95 | 272,959 | 49,702 |
| 4 | 272,959 | 298,317 | 324,055 | 51,096 |
| 5 | 324,055 | 350,207 | 376,805 | 52,75 |
| 6 | 376,805 | 403,884 | 431,478 | 54,673 |
| 7 | 431,478 | 459,622 | 488,353 | 56,875 |
| 8 | 488,353 | 517,707 | 547,721 | 59,368 |
| 9 | 547,721 | 578,434 | 609,885 | 62,164 |
| 10 | 609,885 | 642,114 | 675,161 | 65,277 |
| 11 | 675,161 | 709,071 | 743,884 | 68,723 |
| 12 | 743,884 | 779,647 | 816,404 | 72,52 |
| 13 | 816,404 | 854,203 | 893,091 | 76,687 |
| 14 | 893,091 | 933,119 | 974,336 | 81,245 |
| 15 | 974,336 | 1016,797 | 1060,555 | 86,219 |
| 16 | 1060,555 | 1105,666 | 1152,187 | 91,632 |
| 17 | 1152,187 | 1200,178 | 1249,7 | 97,513 |
| 18 | 1249,7 | 1300,816 | 1353,592 | 103,892 |
| 19 | 1353,592 | 1408,094 | 1464,392 | 110,801 |
| 20 | 1464,392 | 1522,559 | 1582,668 | 118,275 |
| 21 | 1582,668 | 1644,795 | 1709,021 | 126,354 |
| 22 | 1709,021 | 1775,427 | 1844,098 | 135,077 |
| 23 | 1844,098 | 1915,121 | 1988,587 | 144,489 |
| 24 | 1988,587 | 2064,59 | 2143,227 | 154,64 |
| 25 | 2143,227 | 2224,597 | 2308,806 | 165,579 |
| 26 | 2308,806 | 2395,959 | 2486,169 | 177,364 |
| 27 | 2486,169 | 2579,551 | 2676,223 | 190,054 |
| 28 | 2676,223 | 2776,309 | 2879,937 | 203,713 |
| 29 | 2879,937 | 2987,238 | 3098,35 | 218,414 |
| 30 | 3098,35 | 3213,415 | 3332,579 | 234,229 |
| 31 | 3332,579 | 3455,993 | 3583,817 | 251,238 |
| 32 | 3583,817 | 3716,212 | 3853,348 | 269,531 |
| 33 | 3853,348 | 3995,399 | 4142,547 | 289,199 |

TABLEAU 7 (*fin*)

| Groupe | Fréquence basse/Hz | Fréquence centrale/Hz | Fréquence haute/Hz | Largeur de fréquences/Hz |
| --- | --- | --- | --- | --- |

| *k* | *fl[k]* | *fc[k]* | *fu[k]* | *fw[k]* |
| --- | --- | --- | --- | --- |
| 34 | 4142,547 | 4294,979 | 4452,89 | 310,343 |
| 35 | 4452,89 | 4616,482 | 4785,962 | 333,072 |
| 36 | 4785,962 | 4961,548 | 5143,463 | 357,5 |
| 37 | 5143,463 | 5331,939 | 5527,217 | 383,754 |
| 38 | 5527,217 | 5729,545 | 5939,183 | 411,966 |
| 39 | 5939,183 | 6156,396 | 6381,463 | 442,281 |
| 40 | 6381,463 | 6614,671 | 6856,316 | 474,853 |
| 41 | 6856,316 | 7106,708 | 7366,166 | 509,85 |
| 42 | 7366,166 | 7635,02 | 7913,614 | 547,448 |
| 43 | 7913,614 | 8202,302 | 8501,454 | 587,84 |
| 44 | 8501,454 | 8811,45 | 9132,688 | 631,233 |
| 45 | 9132,688 | 9465,574 | 9810,536 | 677,849 |
| 46 | 9810,536 | 10168,013 | 10538,46 | 727,924 |
| 47 | 10538,46 | 10922,351 | 11320,175 | 781,715 |
| 48 | 11320,175 | 11732,438 | 12159,67 | 839,495 |
| 49 | 12159,67 | 12602,412 | 13061,229 | 901,56 |
| 50 | 13061,229 | 13536,71 | 14029,458 | 968,229 |
| 51 | 14029,458 | 14540,103 | 15069,295 | 1039,837 |
| 52 | 15069,295 | 15617,71 | 16186,049 | 1116,754 |
| 53 | 16186,049 | 16775,035 | 17385,42 | 1199,371 |
| 54 | 17385,42 | 17690,045 | 18000 | 614,58 |

La représentation de la fréquence par rapport à la hauteur sonore est obtenue par l'algorithme décrit dans l'alinéa suivant, où *Fsp[ kf* *]* est la représentation de l'énergie des «*Sorties TFR pondérées de l'oreille externe»*:

 (11)

ou la représentation de l'énergie du signal d'erreur:

 (12)

respectivement. Voir le calcul du signal d'erreur dans le § 3.4.

Les résultats de cette étape du traitement sont les énergies des groupes de fréquences, *Pe*[ *k, n*].

#### 2.1.5.1 Pseudo-code

|  |  |
| --- | --- |
| /\* inputs \*/ |  |
| Fsp[ ] | : énergies d'entrée |
| /\* outputs \*/ |  |
| Pe[ ] | : énergies dans le domaine des hauteurs sonores |
| /\* intermediate values \*/ |  |
| i | : indice des groupes de fréquences |
| k | : \ indice de la ligne fft |
| Z | Nombre de groupes de fréquences:  109 pour la version de base  55 pour la version avancée |
| fl[] | : fréquence basse du groupe de fréquences |
| fu[] | : fréquence haute du groupe de fréquences |
| Fres | : constante pour résolution en fréquences |

Fres = 48000/2048;

for(i=0; i<Z; i++ )

{

Pe[i]=0;

for(k=0;k<1024;k++)

{

/\* line inside frequency group \*/

if( (( k-0.5)\*Fres >= fl[i]) && ((k+0.5)\*Fres <= fu[i]))

{

Pe[i] += Fsp[k];

}

/\* frequency group inside\*/

else if( (( k-0.5)\*Fres < fl[i]) && ((k+0.5)\*Fres > fu[i]))

{

Pe[i] += Fsp[k]\*(fu[i]-fl[i])/Fres;

}

/\* left border \*/

else if( ((k-0.5)\*Fres < fl[i]) && ((k+0.5)\*Fres > fl[i]))

{

Pe[i] += Fsp[k]\*( (k+0.5)\*Fres - fl[i])/Fres;

}

/\* right border

else if( ((k-0.5)\*Fres < fu[i]) && ((k+0.5)\*Fres > fu[i]);

{

Pe[i] += Fsp[k]\*(fu[i]- (k-0.5)\*Fres)/Fres;

}

/\* line outside frequency group \*/

else

{

Pe[i] += 0;

}

}

/\* limit result \*/

Pe[i]=max(Pe[i],0.000000000001);

}

### 2.1.6 Ajout de bruit interne

Un écart dépendant de la fréquence *PThres* est ajouté aux énergies dans chaque groupe de fréquences:

 (13)

 (14)

Les résultats de cette étape du traitement, *Pp[* *k, n]* sont les «*caractéristiques* *des hauteurs sonores»*.

### 2.1.7 Étalement

Les *caractéristiques* *de hauteur sonore* *Pp[* *k* *n]* sont dégradés en fréquence par une fonction d'étalement dépendante du niveau. La fonction d'étalement est une fonction exponentielle double. La pente inférieure est toujours 27 dB/Bark et la pente supérieure dépend de la fréquence et de l'énergie.

Les pentes sont calculées de la façon suivante:

 (15)

 (16)

avec:



L'étalement est effectué indépendamment pour chaque groupe de fréquences *k:*

 (17)

où *Eline* est donné par:

 (18)

*NormSP*[*k*] est calculé de la façon suivante:

 (19)

où:

 (20)

et *res* est la résolution de l'échelle des hauteurs sonores en Bark (0,25 pour la version de base et 0,5 pour la version avancée).

Les caractéristiques obtenues à ce stade du traitement, *E*2*[* *k, n]*, sont utilisées ultérieurement pour le calcul des caractéristiques de modulation et appelées «caractéristiques d'excitation non dégradées».

### 2.1.8 Étalement dans le domaine temporel

Afin de modéliser le postmasquage, les énergies de chaque groupe de fréquences sont dégradées temporellement par des filtres passe-bas de premier ordre. Les constantes temporelles dépendent de la fréquence centrale de chaque groupe (voir l'équation (10) et le Tableau 6) et sont calculées de la façon suivante:

 (21)

Les filtres passe-bas de premier ordre sont calculés de la façon suivante:

****** (22)

 (23)

où *a* est calculé à partir des constantes temporelles ci-dessus par:

****** (24)

*n* est le numéro de trame, *k* est l'indice du groupe et *Ef* [*k,* 0] = 0.

Les caractéristiques obtenues à ce stade du traitement, *E*[ *k, n*], sont dénommées «*caractéristiques* *d'excitation»*.

### 2.1.9 Seuil de masquage

Le masquage est l'effet par lequel un signal faible mais distinctement audible devient inaudible lorsqu'apparaît un signal comparativement plus puissant. Le seuil est calculé en pondérant les caractéristiques d'excitation avec la fonction de pondération *m*[*k*].

 (25)

 (26)

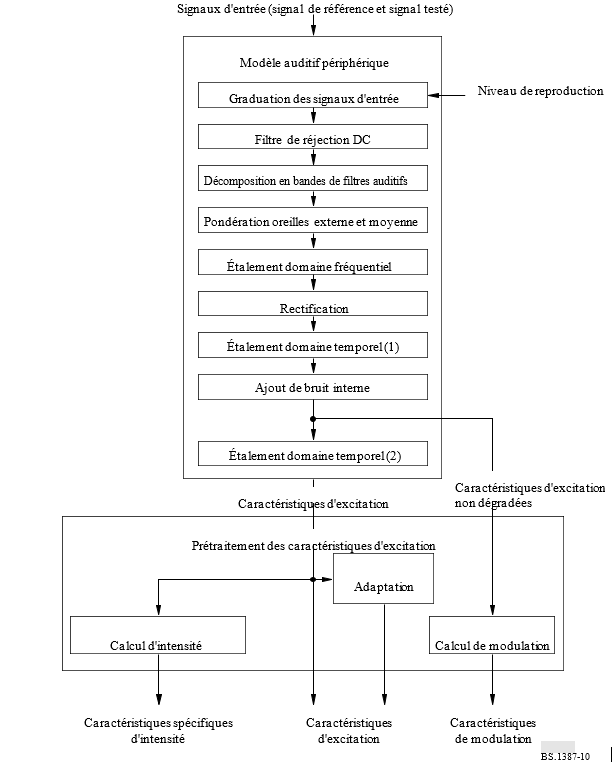
À ce stade du traitement, les caractéristiques *M*[ *k, n*], sont dénommées «*caractéristiques* *de masques»*.

## 2.2 Modèle auditif fondé sur un banc de filtres

### 2.2.1 Aperçu général

FIGURE 10

Modèle auditif périphérique et prétraitement des caractéristiques d'excitation   
pour la partie du modèle fondé sur le banc de filtres



À l'entrée du modèle auditif fondé sur un banc de filtres, le Signal testé et le Signal de référence sont ajustés en fonction du niveau de reproduction supposé et envoyés à travers un filtre passe-haut afin de supprimer les composantes DC et subsoniques des signaux. Les signaux sont ensuite décomposés en signaux de bande passante par des filtres à phase linéaires également distribués sur une échelle perceptuelle des hauteurs sonores. Une pondération dépendante de la fréquence est appliquée aux signaux de bande passante afin de modéliser les caractéristiques spectrales de l'oreille externe et de l'oreille moyenne. La résolution spectrale dépendante du niveau des filtres auditifs est modélisée par une convolution en domaine fréquentiel des sorties avec une fonction d'étalement dépendante du niveau.

Les enveloppes des signaux sont calculées selon la transformation de Hilbert des signaux passe‑bande («rectification») et une convolution dans le domaine temporel avec une fonction de fenêtrage est appliquée afin de modéliser le prémasquage. Ensuite, un écart dépendant de la fréquence est ajouté pour prendre en compte le bruit interne du système auditif et modéliser le seuil au repos. Enfin, une deuxième convolution dans le domaine temporel est effectuée avec une fonction d'étalement exponentielle qui modélise le prémasquage.

Les *caractéristiques* *d'excitation* obtenues sont utilisées pour calculer les *caractéristiques* *spécifiques d'intensité,* et les caractéristiques avant l'étalement final dans le domaine temporel («*caractéristiques* *d'excitation non dégradées»*) sont utilisées pour calculer les *caractéristiques* *de modulation*. Celles‑ci, avec les caractéristiques d'excitation, sont la base sur laquelle les valeurs du modèle sont calculées. Afin de séparer l'influence de la réponse en fréquence du système testé des autres distorsions, les caractéristiques d'excitation du signal d'essai et du Signal de référence sont également adaptées spectralement l'une à l'autre («adaptation*»*). Les caractéristiques de modulation et les caractéristiques spécifiques d'intensité sont calculées à partir des caractéristiques d'excitation adaptées et non adaptées.

### 2.2.2 Sous-échantillonnage

À la sortie du banc de filtres, les signaux sont sous-échantillonnés d'un facteur 32 et après le premier étalement dans le domaine temporel, les signaux sont sous-échantillonnés d'un facteur 6 (voir la Fig. 11).

### 2.2.3 Réglage au niveau de reproduction

Le facteur de graduation pour l'entrée est calculé à partir du niveau de reproduction supposé d'un signal d'entrée pleine échelle par la formule:

 (27)

Si le niveau de reproduction exact n'est pas connu, il est conseillé de régler *Lmax* sur 92 dBSPL.

### 2.2.4 Filtre de réjection DC

Le banc de filtres étant sensible aux infrasons présents dans les signaux d'entrée, on applique aux signaux d'entrée un filtre de rejection DC. On utilise un filtre passe-haut de Butterworth de quatrième ordre d'une fréquence limite de 20 Hz. Ce filtre est une cascade de deux filtres IIR de second ordre:

*yn**xn – 2xn–1**xn–2**b1 yn–1**b2 yn–2* (28)

où les coefficients du premier bloc sont:

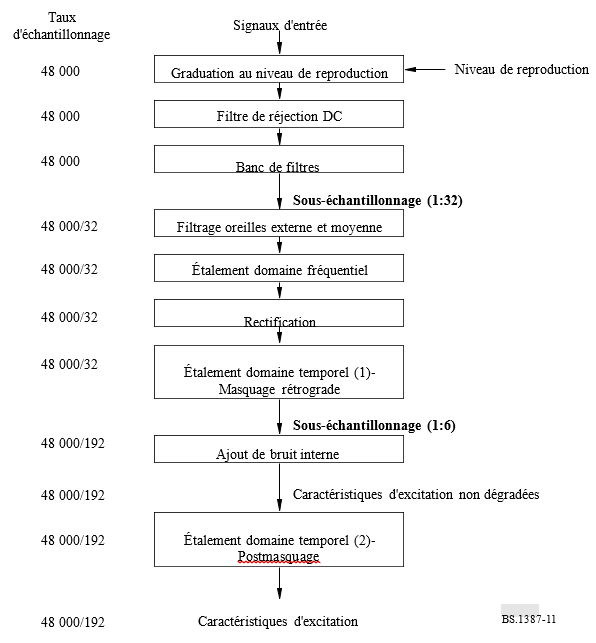
*b1, 2*  *1,99517, –0,995174*

et les coefficients du second bloc sont:

*b1, 2*  *1,99799, –0,997998*

FIGURE 11

Sous-échantillonnage dans le modèle auditif fondé sur un banc de filtres



### 2.2.5 Banc de filtres

Le banc de filtres est composé de 40 paires de filtres pour chaque canal du signal d'essai et du Signal de référence. Les filtres sont espacés régulièrement et ont une largeur de bande absolue constante quand ils sont liés à une échelle auditive des hauteurs sonores. Chaque paire de filtres est composée de deux filtres à réponse en fréquence égale mais avec une différence de 90° en réponse en phase. Ainsi, la sortie du second filtre représente la transformation de Hilbert de la sortie du premier filtre (ou la partie imaginaire, si le premier filtre est supposé représenter la partie réelle d'un signal complexe). Les enveloppes de leurs réponses impulsionnelles ont une forme en cos2. Les filtres sont définis dans le Tableau 8 et par l'équation (29) (où *k* est l'indice du filtre, *n* l'indice de l'échantillon temporel et *T* la durée entre deux échantillons: *T  1/48 000*). Ils peuvent être réalisés comme des filtres FIR en prenant pour coefficients les valeurs *hre(k, n)* et *him(k, n)*. Si les signaux d'entrée sont temporellement limités, les sorties de filtres peuvent aussi être calculées par des algorithmes récurrents très rapides.

 (29)

TABLEAU 8

Fréquence centrale, durée de la réponse en impulsion et retard   
complémentaire pour chaque filtre

| Indice du filtre | Fréquence centrale/Hz | Durée de la réponse en impulsion/échantillons | Retard complémentaire/ échantillons |
| --- | --- | --- | --- |
| *k* | *fc[k]* | *N[k]* | *D[k]* |
| 0 | 50,00 | 1456 | 1 |
| 1 | 116,19 | 1438 | 10 |
|  | 183,57 | 1406 | 26 |
| 3 | 252,82 | 1362 | 48 |
| 4 | 324,64 | 1308 | 75 |
| 5 | 399,79 | 1244 | 107 |
| 6 | 479,01 | 1176 | 141 |
| 7 | 563,11 | 1104 | 177 |
| 8 | 652,97 | 1030 | 214 |
| 9 | 749,48 | 956 | 251 |
| 10 | 853,65 | 884 | 287 |
| 11 | 966,52 | 814 | 322 |
| 12 | 1089,25 | 748 | 355 |
| 13 | 1223,10 | 686 | 386 |
| 14 | 1369,43 | 626 | 416 |
| 15 | 1529,73 | 570 | 444 |
| 16 | 1705,64 | 520 | 469 |
| 17 | 1898,95 | 472 | 493 |
| 18 | 2111,64 | 430 | 514 |
| 19 | 2345,88 | 390 | 534 |
| 20 | 2604,05 | 354 | 552 |
| 21 | 2888,79 | 320 | 569 |
| 22 | 3203,01 | 290 | 584 |
| 23 | 3549,90 | 262 | 598 |
| 24 | 3933,02 | 238 | 610 |
| 25 | 4356,27 | 214 | 622 |
| 26 | 4823,97 | 194 | 632 |
| 27 | 5340,88 | 176 | 641 |

TABLEAU 8 (*fin*)

| **Indice du filtre** | **Fréquence centrale/Hz** | **Durée de la réponse en impulsion/échantillons** | **Retard complémentaire/ échantillons** |
| --- | --- | --- | --- |
| *k* | *fc[k]* | *N[k]* | *D[k]* |
| 28 | 5912,30 | 158 | 650 |
| 29 | 6544,03 | 144 | 657 |
| 30 | 7242,54 | 130 | 664 |
| 31 | 8014,95 | 118 | 670 |
| 32 | 8869,13 | 106 | 676 |
| 33 | 9813,82 | 96 | 681 |
| 34 | 10858,63 | 86 | 686 |
| 35 | 12014,24 | 78 | 690 |
| 36 | 13292,44 | 70 | 694 |
| 37 | 14706,26 | 64 | 697 |
| 38 | 16270,13 | 58 | 700 |
| 39 | 18000,02 | 52 | 703 |

Les fréquences centrales sont situées entre 50 Hz et 18 000 Hz. L'échelle auditive des hauteurs sonores est calculée à partir de l'approximation donnée par [Schroeder et autres, 1979]:

 (30)

Les unités de hauteur sonore sont dénommées *Bark* (bien que cette échelle ne représente pas exactement l'échelle de Bark définie par [Zwicker et Feldtkeller, 1967]).

Afin d'avoir les mêmes retards pour tous les filtres, l'entrée de chaque filtre est retardée de *D* échantillons, où *D* est la moitié de la différence entre la durée de sa réponse en impulsion et la durée de la réponse en impulsion du filtre de réponse la plus longue[[3]](#footnote-3)3:

 (31)

Les sorties de filtres sont sous-échantillonnées selon un facteur 32, ce qui signifie que les valeurs de sortie sont calculées tous les 32 échantillons d'entrée pour tous les filtres[[4]](#footnote-4)4.

### 2.2.6 Filtrage par l'oreille externe et par l'oreille moyenne

La réponse en fréquence de l'oreille externe et de l'oreille moyenne est modélisée par une fonction de pondération dépendante de la fréquence appliquée aux sorties de filtres (l'équation (32)).

 (32)

#### 2.2.6.1 Pseudo-code

|  |  |
| --- | --- |
| /\* inputs \*/ |  |
| out\_re,out\_im | : sorties du banc de filtres (parties réelle et imaginaire) |
| W | : fonction de pondération (voir l'équation (32)) |
| /\* outputs \*/ |  |
| out\_re,out\_im | : sorties du banc de filtres |
| /\* intermediate values \*/ |  |
| k | : indice du filtre |
| Wt | : facteur de pondération |

*/\* outer and middle ear filtering \*/*

for(k=0..39)

{

Wt = pow(10,W[k]/20)

out\_re[k] \*= Wt;

out\_im[k] \*= Wt;

}

### 2.2.7 Étalement dans le domaine fréquentiel

Les valeurs de sortie du banc de filtres sont dégradées en fréquence par une fonction d'étalement dépendante du niveau. Cette fonction d'étalement est une fonction exponentielle double. La pente inférieure est toujours 31 dB/Bark et la pente supérieure varie entre ***−***24 et ***−***4 dB/Bark.

La pente supérieure, *s*[*k*], est calculée par la formule:

 (33)

Le niveau *L*[ *k*] est calculé indépendamment pour chaque canal de filtre en prenant le carré de la valeur absolue de la sortie du filtre et en le transformant sur l'échelle dB. Les fréquences centrales, *fc* [ *k*] proviennent du Tableau 8. Les représentations linéaires des pentes sont temporellement amorties par un filtre passe-bas de premier ordre ayant une constante temporelle de 100 ms.

L'étalement est effectué indépendamment pour les filtres représentant la partie réelle des signaux et les filtres représentant les parties imaginaires des signaux (équation (29)). L'étalement de fréquence est d'abord effectué pour la pente supérieure (dépendante du niveau) puis pour la pente inférieure avec un algorithme de filtre IIR de premier ordre.

#### 2.2.7.1 Pseudo-code

|  |  |
| --- | --- |
| /\* inputs \*/ |  |
| out\_re,out\_im | : sorties du banc de filtres (parties réelle et imaginaire) |
| z[ ] | : bandes critiques pour les fréquences centrales des bandes de filtres en Bark (conformément au Tableau 8 et à l'équation (30)) |
| /\* outputs \*/ |  |
| A\_re,A\_im | : caractéristiques de sorties |
| /\* intermediate values \*/ |  |
| j,k | : indice du filtre |
| a,b | : coefficients d'amortissement temporel |
| dist | : constante de calcul des parasites |
| L[ ] | : niveau à chaque sortie de filtre |
| s[ ] | : pente locale d'étalement vers le haut |
| d1,d2 | : tampons |
| /\* static \*/ | (les valeurs de la trame précédente sont conservées; les valeurs sont initialisées avec des zéros au début de la mesure) |
| cl, cu[ ] | : fraction étalée du signal |

*/\* level dependent upward spreading \*/*

dist = pow(0.1,(z[39]-z[0])/(39\*20));

*/\* (z[39]-z[0])/39 is the distance in Bark between two adjacent filter bands \*/*

a = exp(-32/(48000\*0.1));

b = 1 - a;

for(k=0..39)

{

A\_re[k] = out\_re[k];

A\_im[k] = out\_im[k];

}

for(k=0..39)

{

*/\* calculate level dependent slope \*/*

L[k] = 10\*log10(out\_re[k]\* out\_re[k] + out\_im[k]\* out\_im[k];

s[k] = max(4,(24 + 230/fcentre[k] - 0.2\*L[k]));

*/\* calculate spread fraction and smooth it over time\*/*

cu[k] = a\*pow(dist,s[k])+b\*cu[k];

*/\* spreading of band k \*/*

d1 = out\_re[k]

d2 = out\_im[k]

for(j=k+1..39)

{

d1 \*= cu[k];

d2 \*= cu[k];

A\_re[j] += d1;

A\_im[j] += d2;

}

}

*/\* downward spreading \*/*

cl = pow(dist,31);

d1 = 0;

d2 = 0;

for(k=39..0)

{

/\* spreading of band k \*/

d1 = d1 \* cl + A\_re[k];

d2 = d2 \* cl + A\_im[k];

A\_re[k] = d1;

A\_im[k] = d2;

}

### 2.2.8 Rectification

On calcule les énergies en sortie de filtre en ajoutant le carré des valeurs du filtre représentant la partie réelle du signal à celui représentant la partie imaginaire du signal.

 (34)

Toutes les opérations suivantes sont effectuées sur ces énergies.

### 2.2.9 Dégradation dans le domaine temporel (1) – Prémasquage

Pour modéliser le prémasquage, les énergies en sortie de filtres sont temporellement dégradées par un filtre FIR de réponse impulsionnelle en cos2 avec 12 dérivations (ce qui correspond à une réponse de filtre de 384 échantillons au taux d'échantillonnage d'entrée du banc de filtres). Après dégradation temporelle, les sorties sont sous-échantillonnées selon un facteur 6. Les valeurs qui en résultent sont multipliées par un facteur d'étalonnage *cal1**0,9761* afin d'obtenir les niveaux de sortie appropriés pour le niveau de reproduction donné.

 (35)

### 2.2.10 Ajout de bruit interne

Après la première dégradation dans le domaine temporel, un écart dépendant de la fréquence *EThres* est ajouté aux énergies dans chaque canal de filtre.

 (36)

 (37)

À cette étape du traitement, les caractéristiques *E2*[*k, n*] sont utilisées ultérieurement pour le calcul des caractéristiques de modulation et sont dénommées «*caractéristiques d'excitation non dégradées»*.

### 2.2.11 Dégradation dans le domaine temporel (2) – Postmasquage

Pour modéliser le postmasquage, les énergies dans chaque canal de filtre sont dégradées temporellement par des filtres passe-bas de premier ordre. Les constantes temporelles dépendent de la fréquence centrale de chaque filtre (donnée dans le Tableau 5) et sont obtenues par l'équation:

 (38)

Les filtres passe-bas de premier ordre sont calculés selon l'équation:

 (39)

où *a* est calculé à partir des constantes temporelles ci-dessus par:

 (40)

À ce stade du traitement, les caractéristiques *E*[*k, n*] sont appelées «*caractéristiques d'excitation»*.

# 3 Prétraitement des caractéristiques d'excitation

La plupart des calculs décrits dans ce paragraphe sont utilisés aussi bien avec le modèle auditif fondé sur le banc de filtres qu'avec le modèle TFR. Le facteur de sous-échantillonnage et le nombre de bandes de fréquences étant différents entre les deux modèles auditifs, les constantes qui dépendent de ce facteur sont décrites par les variables dépendantes du modèle auditif *StepSize* et *Z*. Pour le modèle auditif TFR, la valeur de *StepSize* est *1 024* et celle de *Z* est soit *55* (version avancée), soit *109* (version de base). Pour le modèle auditif fondé sur le banc de filtres la valeur de *StepSize* est *192* et celle de *Z* est *40*. Sauf indication contraire, toutes les variables et les filtres récursifs sont initialisés à zéro.

## 3.1 Adaptation des niveaux et des caractéristiques

Pour compenser les différences de niveau et les distorsions linéaires entre le signal d'essai et le Signal de référence, on adapte les uns aux autres les niveaux moyens du Signal testé et du Signal de référence.

Les énergies dans chaque canal de filtre sont d'abord amorties par des filtres passe-bas de premier ordre. Les constantes temporelles dépendent des fréquences centrales des filtres et sont choisies de la façon suivante:

 (41)

Les filtres passe-bas de premier ordre sont calculés de la façon suivante:

 (42)

 (43)

où *ETest* et *ERef* sont les caractéristiques d'excitation qui doivent être adaptées les unes aux autres et *a* est calculé à partir des constantes temporelles selon l'équation:

 (44)

### 3.1.1 Adaptation des niveaux

À partir des caractéristiques d'entrée filtrées par les filtres passe-bas ** et ** on calcule un facteur de correction temporaire *LevCorr* avec l'équation:

 (45)

Si ce facteur de correction est supérieur à 1, le Signal de référence est divisé par le facteur de correction, sinon le signal d'essai est multiplié par le facteur de correction.

 (46)

### (47)

### 3.1.2 Adaptation des caractéristiques

On calcule les facteurs de correction pour chaque canal en comparant les enveloppes temporelles des sorties de filtre du signal d'essai et du Signal de référence:

 (48)

Les valeurs de *a* sont calculées conformément à l'équation ci-dessus (équation (44)) à partir des constantes temporelles données par l'équation (41). Si *R*[*k, n*] est supérieur à 1, le facteur de correction pour le signal testé est *R*[*k, n*]*–1* et le facteur de correction pour le Signal de référence est 1. Dans le cas contraire, le facteur de correction pour le Signal de référence est *R*[*k, n*] et le facteur de correction pour le signal d'essai est 1.

 (49)

Si le dénominateur de (48) est nul (*R*[*k, n*] serait alors indéfini) et si le numérateur est plus grand que zéro, *RTest*[*k, n*] vaut zéro et *RRef*[*k, n*] vaut un. Si le numérateur de (48) est également nul, les rapports *RTest*[*k, n*] et *RRef*[*k, n*] sont copiés à partir de la bande de fréquences inférieure. S'il n'y a pas de bande de fréquences inférieure (à savoir si *k* *0*) les rapports *RTest*[*k, n*] et *RRef*[*k, n*] sont égaux à 1.

On fait la moyenne des facteurs de correction sur *M* canaux de filtres et on les amortit temporellement (équation (50)) avec les mêmes constantes temporelles que ci-dessus (équations (41) à (44)). La largeur de la fenêtre de fréquences *M* vaut 3 pour le modèle auditif fondé sur le banc de filtres, et 4 (version avancée) ou 8 (version de base) pour le modèle TFR.

 (50)



Aux limites de l'échelle des fréquences où la fenêtre fréquentielle excéderait la gamme des bandes de filtres, la largeur de la fenêtre fréquentielle est réduite en conséquence:

*M*1  min(*M*1, *k*), *M*2  min(*M*2, *z*– *k*– 1), *M*  *M*1  *M*2  1 (51)

Les caractéristiques d'entrée adaptées au niveau sont pondérées avec les facteurs de correction correspondants *PattCorrTest/Ref*[*k, n*] afin d'obtenir les caractéristiques adaptées spectralement:

** (52)

** (53)

## 3.2 Modulation

À partir des *caractéristiques d'excitation non dégradées*, *E*2[*k, n*], on calcule une intensité acoustique en élevant l'excitation à la puissance 0,3. Cette valeur et la valeur absolue de sa déviation temporelle sont temporellement dégradées.

 (54)

 (55)

Les valeurs de *a* sont calculées comme en (44) à partir des constantes temporelles données par:

 (56)

À partir des valeurs qui en résultent, on calcule ** et ** la mesure de la modulation de l'enveloppe à chaque sortie de filtre:

 (57)

Ultérieurement, on utilise également les valeurs de ** dans le calcul de la différence de modulation.

## 3.3 Intensité acoustique

Les caractéristiques spécifiques d'intensité acoustique du Signal testé et du Signal de référence sont calculées grâce à la formule:

 (58)

donnée en [Zwicker et Feldtkeller, 1967]. L'intensité acoustique totale du Signal testé et du Signal de référence est la somme pour tous les canaux de filtre de toutes les valeurs d'intensité acoustique supérieures à zéro.

 (59)

La constante de graduation est *const*1,07664 pour le modèle auditif périphérique TFR et *const*1,26539 pour le modèle auditif périphérique fondé sur le banc de filtres, afin d'obtenir une intensité acoustique globale de une sonie pour un son sinusoïdal de 40 dBSPL à 1 kHz. L'indice de seuil *s* et l'excitation au seuil *EThres* sont calculés par les équations:

 (60)

et

 (61)

respectivement.

NOTE – Du fait de l'existence de différents modèles auditifs périphériques, l'intensité acoustique calculée ici n'est pas identique à l'intensité acoustique telle qu'elle est définie dans ISO 532 (Acoustics – Method for calculating loudness level, 1975).

## 3.4 Calcul du signal d'erreur

Le signal d'erreur n'est calculé que dans le modèle TFR. On le calcule dans le domaine fréquentiel en prenant la différence entre les spectres en amplitude filtrés de l'oreille externe et de l'oreille moyenne du signal de référence et du signal d'essai (voir le § 2.1.4).

 (62)

*Fnoise* est représenté dans le domaine des hauteurs sonores par l'algorithme décrit au § 2.1.5.

Les sorties de cet algorithme, *Pnoise*[*n, k*], sont dénommées «*Caractéristiques du bruit»*.

# 4 Calcul des variables de sortie de modèle

## 4.1 Aperçu général

TABLEAU 9

Aperçu des variables de sortie de modèle utilisées   
pour la prévision de la qualité audio de base

| Variable de sortie de modèle (MOV) | Calculée dans le modèle auditif … | | Utilisée dans la version … | |
| --- | --- | --- | --- | --- |
|  | TFR | Banc de filtres | De base | Avancée |
| WinModDiff1B | oui | non | oui | non |
| AvgModDiff1B | oui | non | oui | non |
| AvgModDiff2B | oui | non | oui | non |
| RmsModDiffA | non | oui | non | oui |
| AvgNoiseLoudB | oui | non | oui | non |
| RmsNoiseLoudB | oui | non | oui | non |
| RmsNoiseLoudAsymA | non | oui | non | oui |
| AvgLinDistA | non | oui | non | oui |
| BandwidthRefB | oui | non | oui | non |
| BandwidthTestB | oui | non | oui | non |
| Total NMRB | oui | non | oui | non |
| RelDistFramesB | oui | non | oui | non |
| Segmental NMRB | oui | non | non | oui |
| MFPDB | oui | non | oui | non |
| ADBB | oui | non | oui | non |
| EHSB | oui | non | oui | oui |

## 4.2 Différence de modulation

Les différences de modulation des enveloppes temporelles du Signal testé et du Signal de référence sont mesurées en calculant une différence de modulation locale pour chaque canal de filtre (équation (63)): où *Modtest* et *ModRef* sont déterminés par application de l'équation (57) au signal de référence *Rtest*.

 (63)

La différence de modulation temporaire est la somme des différences de modulation locales sur tous les canaux de filtre (équation (64)).

 (64)

Le seuil au repos est pris en compte par un facteur de pondération dépendant du niveau (équation (65)) calculé à partir des caractéristiques d'excitation modifiées pour le signal de référence données par l'équation (55) et de la fonction de bruit interne définie par l'équation (36) pour le modèle auditif fondé sur le banc de filtres et l'équation (13) pour le modèle TFR.

 (65)

La moyenne temporelle des différences temporaires de modulation *ModDiff[n]* qui utilise les facteurs de pondération *TempWt[n]* est décrite au § 5.2. Les valeurs des constantes *negWt*, *offset* et *levWt* sont données dans le Tableau 10.

TABLEAU 10

Variables de sortie de modèle dans l'estimation  
de la différence globale de modulation

| MOV (Xxx=Win/Avg/Rms) | negWt | écart | levWt |
| --- | --- | --- | --- |

|  |  |  |  |
| --- | --- | --- | --- |
| *XxxModDiff1B* | 1 | 1 | 100 |
| *XxxModDiff2B* | 0,1 | 0,01 | 100 |
| *XxxModDiffA* | 1 | 1 | 1 |

### 4.2.1 RmsModDiffA

La variable de sortie de modèle *RmsModDiffA* est la moyenne quadratique de la différence de modulation calculée à partir du modèle auditif fondé sur le banc de filtres. Voir le calcul de la moyenne temporelle au § 5.2.2 et les constantes dans le Tableau 10.

### 4.2.2 WinModDiff1B

La variable de sortie de modèle *WinModDiff1B* est la moyenne avec fenêtrage de la différence de modulation calculée à partir du modèle auditif TFR. Voir le calcul de la moyenne temporelle au § 5.2.3 et les constantes dans le Tableau 10. Le facteur de pondération temporelle indiqué dans l'équation (65) n'est pas appliqué à cette variable.

### 4.2.3 AvgModDiff1B et AvgModDiff2B

Les variables de sortie de modèle *AvgModDiff1B* et *AvgModDiff2B* sont la moyenne linéaire de la différence de modulation calculée à partir du modèle auditif TFR. La différence entre *AvgModDiff2B* et *AvgModDiff1B* repose sur le choix des constantes. Voir le calcul de la moyenne temporelle au § 5.2.1 et les constantes dans le Tableau 10.

## 4.3 Intensité acoustique du bruit

Les variables de sortie du modèle sont des estimations de l'intensité partielle des distorsions ajoutées en présence du Signal de référence masquant. L'équation de l'intensité partielle (équation (66)) sert à déterminer l'intensité spécifique du bruit selon [Zwicken et Feldtkeller, 1967] s'il n'y a pas de masque, et une valeur qui serait en quelque sorte un rapport entre le bruit et le masque si le bruit est très faible par rapport au masque.

L'intensité acoustique partielle du bruit se calcule avec l'équation:

 (66)

où *E*0 vaut toujours 1, *EThres* est la fonction de bruit interne *EThres*[*k*] telle qu'elle est définie dans l'équation (36) et *s* se calcule avec la formule:

 (67)

Sauf précisions contraires, on utilise comme entrées les *caractéristiques d'excitation spectralement adaptées* (voir le § 3.1): *ETest**EP,Test*[*k, n*] et *ERef**EP,Ref* [*k, n*]. Le coefficient β, qui détermine la quantité de masquage, se calcule de la façon suivante:

 (68)

Les valeurs d'intensité de bruit temporaire ne sont pas prises en compte avant 50 ms après que l'intensité globale du canal de gauche ou du canal de droite ait dépassé une fois la valeur de *NThres  0,1 sone* pour le signal d'essai et le Signal de référence (voir le § 5.2.4.2).

Lors du calcul de la moyenne spectrale, les valeurs instantanées sont normalisées par le nombre de bandes de filtrage par bande critique et non pas par le nombre total de bandes de filtrages; en d'autres termes, le résultat du calcul de la moyenne spectrale est multiplié par le facteur 24.

Si l'intensité de bruit temporaire est inférieure à une valeur seuil de *NLmin* on la considère comme nulle.

TABLEAU 11

Variables de sortie de modèle pour l'estimation  
de l'intensité acoustique globale du bruit

| MOV (Xxx=Win/Avg/Rms) | α | ThresFac0 | S0 | NLmin |
| --- | --- | --- | --- | --- |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| *XxxMissingComponentsB* | 1,5 | 0,15 | 1 | 0 |
| *XxxNoiseLoudB* | 1,5 | 0,15 | 0,5 | 0 |
| *XxxMissingComponentsA* | 1,5 | 0,15 | 1 | 0 |
| *XxxNoiseLoudA* | 2,5 | 0,3 | 1 | 0,1 |
| *XxxLinDistA* | 1,5 | 0,15 | 1 | 0 |

### 4.3.1 RmsNoiseLoudA

La variable de sortie de modèle *RmsNoiseLoudA* est la moyenne quadratique de l'intensité du bruit calculée à partir du modèle auditif fondé sur le banc de filtres. Voir le calcul de la moyenne temporelle au § 5.2.2 et les constantes dans le Tableau 11.

### 4.3.2 RmsMissingComponentsA

La variable de sortie de modèle *RmsMissingComponentsA* est la moyenne quadratique de l'intensité du bruit calculée à partir du modèle auditif fondé sur le banc de filtres. Elle est calculée avec les caractéristiques d'excitation adaptées en spectre du signal d'essai et du Signal de référence afin d'obtenir l'intensité des composantes du Signal de référence qui sont perdues dans le signal d'essai. Voir le calcul de la moyenne temporelle au § 5.2.2 et les constantes dans le Tableau 11.

### 4.3.3 RmsNoiseLoudAsymA

La variable de sortie de modèle *RmsNoiseLoudAsymA* est la somme pondérée des carrés des moyennes d'intensité du bruit (voir le § 4.3.1) et de l'intensité des composantes de signal perdues (voir le § 4.3.2), les deux valeurs étant calculées à partir du modèle auditif fondé sur le banc de filtres.

 (69)

### 4.3.4 AvgLinDistA

La variable de sortie de modèle *AvgLinDistA* est la mesure de l'intensité des composantes de signal perdues lors de l'adaptation spectrale du Signal testé et du Signal de référence. La référence en est l'excitation spectralement adaptée du Signal de référence et le signal d'essai en est l'excitation non adaptée du Signal de référence. Cette MOV est calculée à partir du modèle auditif fondé sur le banc de filtres. Voir le calcul de la moyenne temporelle au § 5.2.1 et les constantes dans le Tableau 11.

### 4.3.5 RmsNoiseLoudB

La variable de sortie de modèle *RmsNoiseLoudB* est la moyenne quadratique de l'intensité du bruit calculée à partir du modèle auditif TFR. Voir le calcul de la moyenne temporelle au § 5.2.2 et les constantes dans le Tableau 11.

## 4.4 Largeur de bande

Ces valeurs de sortie de modèle sont des estimations de la largeur de bande moyenne du signal d'essai et du Signal de référence dans les raies TFR.

Pour chaque trame la largeur de bande locale *BwRef*[*n*] et *BwTest*[*n*] est calculée selon le pseudo-code ci-dessous.

### 4.4.1 Pseudo-code

|  |  |
| --- | --- |
| /\* inputs \*/ |  |
| FLevRef[], FlevelTest[] | : niveau des sorties TFR en dB |
| /\* outputs \*/ |  |
| BwRef, BwTest | : caractéristiques des sorties |
| /\* intermediate values \*/ |  |
| k | : indice des raies TFR |
| ZeroThreshold | : seuil de largeur de bande |

ZeroThreshold = FLevelTst(921);

BwRef = BwTst = 0.0;

for(k=921;k<1024;k++)

{

ZeroThreshold=max(ZeroThreshold,FLevelTst(k));

}

for (k = 920; k>=0; k--)

{

if (FLevelRef[k] >= 10.0+ZeroThreshold)

{

BwRef = k+1;

break;

}

}

for (k = BwRef-1; k>=0; k--)

{

if(FLeveltest[k] >= 5.0+ZeroThreshold)

{

BwTest=k+1;

break;

}

}

### 4.4.2 BandwidthRefB et BandwidthTestB

BandwidthRefB est la moyenne linéaire de BwRef, et BandwidthTestB est la moyenne linéaire de BwTest. Pour le calcul de la moyenne, seules les trames de BwRef  346 sont prises en compte. Les trames de faible énergie au début et à la fin des éléments sont ignorées (voir le § 5.2.4.4). Voir le calcul de la moyenne temporelle au § 5.2.1.

## 4.5 Rapport bruit/masque (NMR)

Les valeurs de modèle suivantes sont calculées à partir des valeurs de bruit et de masque.

Le *NMR* local de la trame courante *n* est:

 (70)

### 4.5.1 Total NMRB

La variable de sortie de modèle *Total NMRB* est la moyenne linéaire du rapport bruit/masque.

 (71)

Les trames de faible énergie au début et à la fin des éléments sont ignorées (voir le § 5.2.4.4).

### 4.5.2 Segmental NMRB

La variable de sortie de modèle *Segmental NMRB* est la moyenne linéaire du NMR local. Voir le calcul de la moyenne temporelle au § 5.2.1.

Les trames de faible énergie au début et à la fin des éléments sont ignorées (voir le § 5.2.4.4).

## 4.6 Relative Disturbed FramesB

La variable de sortie de modèle *Relative* *Disturbed FramesB* (abréviation: RelDistFramesB) représente le nombre de trames ayant:



par rapport au nombre total de trames de l'élément.

Les trames de faible énergie au début et à la fin des éléments sont ignorées (voir le § 5.2.4.4).

## 4.7 Probabilité de détection

Les MOV définies dans ce paragraphe sont fondées sur  (bande *k*, trame *n*), qui sont les *caractéristiques d'excitation E*[*k, n*] exprimées en dB:

 (72)

Pour chaque trame *n*:

Les calculs qui suivent sont effectués indépendamment pour chaque canal *c* (les valeurs de *c* sont gauche et droite). Les *caractéristiques logarithmiques* *d'excitation* sont  pour le Signal de référence et  pour le Signal testé.

Pour chaque bande *k*:

1. Calcul de l'excitation asymétrique moyenne.

 (73)

– Calcul de la dimension du pas de détection réel *s*. La formule qui suit est une approximation de la différence de niveau juste identifiable mesurée par [Zwicker et Fastl, 1990].

Si *L*[*k, n*]  0:

*s*[*k, n*]    5,95072  ((6,39468)/*L* [*k, n*])1,71332  9,01033  10–11  *L* [*k, n*]4  5,05622   
 10–6  *L* [*k, n*]3 – 0,00102438  *L* [*k, n*]2 0,0550197  *L*[*k, n*] –0,198719

sinon

*s* [*k, n*]  1,0  1030 (74)

– Calcul de l'erreur signée *e*:

 (75)

– Si  alors la caractéristique de pente *b* vaut 4,0 sinon elle vaut 6,0. Ceci modélise l'effet suivant: un accroissement de l'énergie de signal du Signal testé par rapport au Signal de référence est plus frappant qu'une diminution.

– Calcul du facteur de graduation *a*:

 (76)

– Calcul de la probabilité de détection. L'équation (76) détermine le facteur de graduation *a* de façon à ce que si *e*[*k, n*] est égal à *s*[*k, n*], *pc*[*k, n*] devient *0,5*:

 (77)

– Calcul du nombre total de pas au-dessus du seuil:

 (78)

– La probabilité de détection bilatérale est:

*pbin* [ *k, n*]  max( *pleft* [ *k, n*], *pright* [*k, n*]) (79)

– Le nombre de pas au-dessus du seuil pour le canal bilatéral est:

*qbin* [ *k, n*]  max(*qleft* [ *k, n*], *qright* [*k, n*]) (80)

La probabilité globale de détection du canal *c* de la trame *n* est:

 (81)

où *c* peut être *left, right* ou *bin*. Le nombre de pas au-dessus du seuil pour le canal *c* de la trame *n* est:

 (82)

### 4.7.1 Probabilité de détection maximum filtrée (MFPDB)

On calcule une version pondérée de la probabilité de détection pour chaque canal *c*:

 (83)

où *Pc[–1]* *0*. La constante *c*0 dépend de *StepSize*:

*c*0  0,9*StepSize*/1 024 (84)

*c*0 réduit la sensibilité aux distorsions infimes.

On calcule la probabilité de détection maximum filtrée (abréviation: MFPD):

 (85)

où *PMc*[*–1*] est nul. La constante *c*1 dépend de *StepSize*:

*c*1  0,99*StepSize*/1 024 (86)

*c*1 modélise l'effet suivant: les distorsions au début d'un extrait sonore sont moins graves qu'à la fin de cet extrait du fait du phénomène d'oubli. Il convient de noter que cette constante est utile pour la modélisation des essais d'écoute lors desquels les sujets ne sont pas autorisés à sélectionner des plus petites parties de l'extrait. Pour le présent modèle, étalonné avec des données provenant des essais d'écoute conformément à la Recommandation UIT-R BS.1116, *c*1 doit être égal à 1,0.

La MOV MFPD est la valeur de *PMbin*[*n*] pour la dernière trame.

### 4.7.2 Bloc distordu moyen[[5]](#footnote-5)5 (ADBB)

On compte le nombre de trames valides présentant une probabilité de détection du canal central *Pbin*[*n*] supérieure à 0,5 (*ndistorted*).

On calcule pour toutes les trames valides le nombre total de pas au-dessus du seuil du canal central *Qbin*[*n*]:



On calcule la distorsion du bloc moyen distordu, *ADB*:

– si *ndistorted* égal zéro alors *ADB*  0 (pas de distorsion audible);

– si *ndistorted*  0 et *Qsum*  0 alors *ADB*  log10 ((*Qsum*)/*ndistorted*));

– si *ndistorted* 0 et *Qsum* égal zéro alors *ADB*  –0,5.

## 4.8 Structure harmonique de l'erreur

Un Signal de référence contenant des harmoniques forts (ex. clarinette basse, clavecin) a un spectre caractérisé par un certain nombre de pics régulièrement espacés séparés par des ventres profonds. Sous certaines conditions, le signal d'erreur peut hériter de cette structure. Par exemple, un bruit mélangé à un tel signal a plus de chances de ne pas être masqué quand le signal est faible dans les ventres du spectre. Le spectre de l'erreur qui en résulte aurait alors une structure semblable à celle du spectre original mais avec un écart en fréquence pour correspondre aux emplacements des ventres. De cette structure peut résulter une distorsion dont les qualités tonales pourraient augmenter la saillance de l'erreur.

L'erreur est définie comme étant la différence des signaux de référence et des signaux traités dans chaque spectre logarithmique, chacun étant pondéré par la réponse en fréquence de l'oreille externe et de l'oreille moyenne (voir § 2.1.4, équation (7)). La structure d'excitation tirée du modèle psychoacoustique n'est pas utilisée ici parce que la transformation non linéaire de Bark, appliquée aux fréquences, estomperait la structure harmonique.

### 4.8.1 EHSB

On mesure l'amplitude de la structure harmonique en identifiant et en mesurant le pic le plus élevé dans le spectre de la fonction d'autocorrélation. Chaque corrélation est calculée comme le cosinus de l'angle compris entre deux vecteurs conformes à la formule suivante, où  est le vecteur d'erreur et  est le même vecteur retardé d'une certaine grandeur. La longueur de la corrélation est la même que le retard maximum (par exemple, 256 dans l'exemple ci-après).

 (87)

Le retard maximal pour obtenir la fonction d'autocorrélation est la plus grande puissance de deux qui est plus petite que la moitié du nombre de la composante fréquentielle de transformation TFR correspondant à 18 kHz.

Par exemple, à une fréquence d'échantillonnage de 48 kHz et dans une fenêtre TFR ayant un effectif de 2 048 échantillons, la composante TFR correspondant à 18 kHz est égale à (18/24) × 1 024  768. Le retard maximal sera donc de 384. Le nombre de retards réel sera 256, qui est la plus grande puissance de deux inférieure à 384. La première valeur de la fonction de corrélation sera obtenue par alignement de *Ft*[0] sur *F*0[0] et la dernière valeur par alignement de *Ft*[0] sur *F*0[255].

Le vecteur résultant des corrélations est intégré dans une fenêtre de Hann normalisée et, après suppression de la composante continue par soustraction de la valeur moyenne, on calcule un spectre de puissance avec transformation TFR. Le pic maximal de ce spectre, après le premier ventre correspond à la fréquence dominante dans la fonction d'autocorrélation. La valeur moyenne au cours des trames de ce pic maximal, multipliée par 1 000,0, est la variable de structure harmonique d'erreur (EHS).

# 5 Moyennes

## 5.1 Calcul de la moyenne spectrale

Sauf indications contraires dans les descriptions des variables de sortie de modèle (voir le § 4), on utilise l'algorithme suivant pour obtenir la moyenne des valeurs locales sur les bandes de fréquences.

### 5.1.1 Moyenne linéaire

On calcule la valeur de la moyenne linéaire avec:

 (88)

où *S* est le nom de la variable de sortie de modèle et *Z* le nombre de groupes de fréquences.

## 5.2 Calcul de la moyenne temporelle

Sauf indications contraires dans les descriptions des variables de sortie de modèle (voir le § 4), on utilise un ou plusieurs des algorithmes suivants pour obtenir la moyenne des valeurs temporaires dans le temps. Le facteur de pondération temporelle est désigné (s'il est appliqué) par le symbole *W* et *Z* est le nombre de bandes de fréquences.

### 5.2.1 Moyenne linéaire

On calcule la valeur de la moyenne linéaire (préfixe «*Avg*») avec:

 (89)

où *X* est le nom de la variable de sortie de modèle et *N* le nombre d'échantillons temporels pour lesquels on a calculé les valeurs temporaires de *X*.

Si on applique une pondération temporelle (voir le § 4.2), on calcule la moyenne linéaire avec:

 (90)

### 5.2.2 Moyenne quadratique

On calcule la moyenne quadratique (préfixe «*Rms»*) selon l'équation:

 (91)

où *X* est le nom de la variable de sortie de modèle et *N* le nombre d'échantillons temporels pour lesquels les valeurs temporaires de *X* ont été calculées.

Si on applique une pondération temporelle (voir le § 4.2), la moyenne quadratique se calcule de la façon suivante:

 (92)

### 5.2.3 Moyenne avec fenêtrage

La valeur de la moyenne avec fenêtrage (préfixe *«Win»*) se calcule avec l'équation:

 (93)

où *X* est le nom de la variable de sortie de modèle, *N* le nombre d'échantillons temporels pour lesquels les valeurs temporaires de *X* ont été calculées et *L* la longueur des fenêtres temporelles dans les échantillons temporels. La longueur de la fenêtre est approximativement 100 ms, ce qui signifie que *L* vaut *4* pour le modèle auditif TFR et *25* pour le modèle auditif fondé sur le banc de filtres.

### 5.2.4 Sélection des trames

#### 5.2.4.1 Moyenne retardée

Pour les variables de sortie de modèle qui utilisent ce critère, les valeurs calculées pendant la première demi‑seconde de la mesure ne sont pas prises en compte dans la moyenne temporelle. On utilise la *Moyenne retardée* pour toutes les variables de sortie de modèle suivantes:

WinModDiff1, AvgModDiff1, AvgModDiff2, RmsNoiseLoudness, RmsNoiseLoudAsym, RmsModDiff, AvgLinDist.

#### 5.2.4.2 Seuil d'intensité

Pour les variables de sortie de modèle qui utilisent ce critère, toutes les valeurs temporaires calculées jusqu'à 50 ms après que l'intensité globale d'un des canaux correspondants a atteint une fois une valeur de *NThres* sonie pour le signal d'essai et le Signal de référence ne sont pas prises en compte dans la moyenne temporelle. On n'utilise le *Seuil d'intensité* que pour les variables de sortie de modèle décrites au § 4.3.

#### 5.2.4.3 Seuil d'énergie

Si l'énergie de la moitié la plus récente d'une trame de 2 048 échantillons est inférieure à 8 000[[6]](#footnote-6)\* dans le canal mono ou dans chacun des canaux droite et gauche à la fois pour la référence et pour les données de Test, la trame est ignorée. Les trames présentent un recouvrement de 50% et seule la moitié de la trame contenant des données nouvelles est évaluée. L'application de ce critère permet d'éviter de traiter des trames n'ayant que peu d'énergie.

Ce critère n'est utilisé que pour la variable de sortie de modèle décrite au § 4.8.

#### 5.2.4.4 Limite de données

Si le fichier traité contient du bruit avant ou après les données du fichier de référence, l'erreur relative peut être très importante puisque le niveau de référence approche –. Si cette erreur est considérée comme un artefact, on peut l'ignorer en appliquant le critère de rejet de limite de données.

A la première ouverture des fichiers, on identifie l'emplacement du début et de la fin des données réelles dans le fichier de référence. Le début ou la fin des données est défini comme premier emplacement, selon que l'on analyse à partir du début ou de la fin du fichier, où la somme des valeurs absolues sur cinq échantillons successifs est supérieure à 200 dans l'un des canaux audio correspondant. Les trames qui sortent entièrement de cette fourchette sont en conséquence ignorées.

On utilise ce critère pour toutes les variables de sortie de modèle.

## 5.3 Moyenne sur les canaux audio

Sauf indications différentes, dans le cas de signaux stéréo, les variables MOV des canaux de gauche et de droite sont calculées linéairement après calcul moyennage temporel.

# 6 Estimation de la qualité audio de base perçue

On estime la *qualité audio de base perçue* en transformant plusieurs variables de sortie de modèle en un seul chiffre grâce à une structure de réseau neuronal artificiel avec une couche cachée.

## 6.1 Réseau neuronal artificiel

La fonction d'activation du réseau neuronal est une sigmoïde asymétrique:

 (94)

Le réseau utilise *I* entrées et *J* nœuds dans la couche cachée. La transformation est définie par un ensemble de facteurs de graduation d'entrée *amin*[*i*], *amax*[*i*], un ensemble de pondération d'entrée *wx*[*i*], un ensemble de pondération de sortie *wy*[*j*] et deux facteurs de graduation de sortie *bmin* et *bmax*. Les entrées sont transformées en un indice de distorsion:

 (95)

qui est directement lié à l'estimation de la *qualité audio de base perçue* en termes d'une *note différentielle objective* (ODG). La relation entre *l'indice de distorsion* et *la note différentielle objective* est donnée par la formule:

 (96)

## 6.2 Version de base

La version de base n'utilise que le modèle auditif TFR. Elle utilise les variables de sortie de modèle suivantes: BandwidthRefB*,* BandwidthTestB*,* Total NMRB*,* WinModDiff1B*,* ADBB*,* EHSB*,* AvgModDiff1B*,* AvgModDiff2B*,* RmsNoiseLoudB*,* MFPDB *et* RelDistFramesB. Ces 11 variables de sortie de modèle sont transformées en un indice de qualité unique par un réseau neuronal décrit au § 6.1 avec trois nœuds dans la couche cachée. On trouvera dans les Tableaux 12 à 16, les paramètres de la transformation.

TABLEAU 12

Variables de sortie de modèle utilisées dans la version de base

| Variable de sortie de modèle (MOV) | Objet |
| --- | --- |

|  |  |
| --- | --- |
| WinModDiff1B |  |
| AvgModDiff1B | Changements de modulation (lié à la rugosité) |
| AvgModDiff2B |  |
| RmsNoiseLoudB | Intensité de la distorsion |
| BandwidthRefB | Distorsions linéaires (réponse en fréquence, etc.) |
| BandwidthTestB |  |
| RelDistFramesB | Fréquence des distorsions audibles |
| Total NMRB | Rapport bruit/masque |
| MFPD | Probabilité de détection |
| ADBB |  |
| EHSB | Structure harmonique de l'erreur |

TABLEAU 13

Facteurs de graduation pour les entrées de la version de base

|  |  |  |  |
| --- | --- | --- | --- |
| indice (i) | MOV (x[i]) | amin[i] | amax[i] |

|  |  |  |  |
| --- | --- | --- | --- |
| 0 | BandwidthRefB | 393,916656 | 921 |
| 1 | BandwidthTestB | 361,965332 | 881,131226 |
| 2 | Total NMRB | –24,045116 | 16,212030 |
| 3 | WinModDiff1B | 1,110661 | 107,137772 |
| 4 | ADBB | –0,206623 | 2,886017 |
| 5 | EHSB | 0,074318 | 13,933351 |
| 6 | AvgModDiff1B | 1,113683 | 63,257874 |
| 7 | AvgModDiff2B | 0,950345 | 1145,018555 |
| 8 | RmsNoiseLoudB | 0,029985 | 14,819740 |
| 9 | MFPDB | 0,000101 | 1 |
| 10 | RelDistFramesB | 0 | 1 |

TABLEAU 14

Pondérations pour les nœuds d'entrée de la version de base

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| indice (i) | MOV (x[i]) | nœud 1 (wx[i,0]) | nœud 2 (wx[i,1]) | nœud 3 (wx[i,2]) |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 0 | BandwidthRefB | –0,502657 | 0,436333 | 1,219602 |
| 1 | BandwidthTestB | 4,307481 | 3,246017 | 1,123743 |
| 2 | Total NMRB | 4,984241 | –2,211189 | –0,192096 |
| 3 | WinModDiff1B | 0,051056 | –1,762424 | 4,331315 |
| 4 | ADBB | 2,321580 | 1,789971 | –0,754560 |
| 5 | EHSB | –5,303901 | –3,452257 | –10,814982 |
| 6 | AvgModDiff1B | 2,730991 | –6,111805 | 1,519223 |
| 7 | AvgModDiff2B | 0,624950 | –1,331523 | –5,955151 |
| 8 | RmsNoiseLoudB | 3,102889 | 0,871260 | –5,922878 |
| 9 | MFPDB | –1,051468 | –0,939882 | –0,142913 |
| 10 | RelDistFramesB | –1,804679 | –0,503610 | –0,620456 |
| 11 | bias | –2,518254 | 0,654841 | –2,207228 |

TABLEAU 15

Pondération pour les nœuds de sortie de la version de base

| nœud 1 (wy[0]) | nœud 2 (wy[1]) | nœud 3 (wy[2]) | écart (wy[3]) |
| --- | --- | --- | --- |

|  |  |  |  |
| --- | --- | --- | --- |
| –3,817048 | 4,107138 | 4,629582 | –0,307594 |

TABLEAU 16

Facteurs de graduation pour la sortie de la version de base

|  |  |  |
| --- | --- | --- |
|  | bmin | bmax |

|  |  |  |
| --- | --- | --- |
| ODG | –3,98 | 0,22 |

## 6.3 Version avancée

La *version avancée* utilise le modèle auditif fondé sur le banc de filtres et le modèle auditif TFR. Elle utilise les variables de sortie de modèle RmsModDiffA, RmsNoiseLoudAsymA, AvgLinDistA, Segmental NMRB et EHSB. Ces 5 variables de sortie de modèle sont transformées en un indice de qualité unique par un réseau neuronal décrit au § 6.1 avec cinq nœuds dans la couche cachée. On trouvera dans les Tableaux 17 à 21, les paramètres de la transformation.

TABLEAU 17

Variables de sortie de modèle utilisées dans la version avancée

| Variable de MOV | Objet |
| --- | --- |

|  |  |
| --- | --- |
| RmsNoiseLoudAsymA | Intensité de la distorsion |
| RmsModDiffA | Changements de modulation (lié à la rugosité) |
| AvgLinDistA | Distorsions linéaires (réponse en fréquence etc.) |
| Segmental NMRB | Rapport bruit/masque |
| EHSB | Structure harmonique de l'erreur |

TABLEAU 18

Facteurs de graduation pour les nœuds d'entrée de la version avancée

| indice (i) | MOV (x[i] ) | amin[i] | amax[i] |
| --- | --- | --- | --- |

|  |  |  |  |
| --- | --- | --- | --- |
| 0 | RmsModDiffA | 13,298751 | 2166,5 |
| 1 | RmsNoiseLoudAsymA | 0,041073 | 13,24326 |
| 2 | Segmental NMRB | –25,018791 | 13,46708 |
| 3 | EHSB | 0,061560 | 10,226771 |
| 4 | AvgLinDistA | 0,024523 | 14,224874 |

TABLEAU 19

Pondérations pour les entrées de la version avancée

| indice (i) | MOV (x[i] ) | nœud 1 (wx[i,0]) | nœud 2 (wx[i,1]) | nœud 3 (wx[i,2]) | nœud 4 (wx[i,3]) | nœud 5 (w4[i,4]) |
| --- | --- | --- | --- | --- | --- | --- |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 0 | RmsModDiffA | 21,211773 | –39,913052 | –1,382553 | –14,545348 | –0,320899 |
| 1 | RmsNoiseLoudAsymA | –8,981803 | 19,956049 | 0,935389 | –1,686586 | –3,238586 |
| 2 | Segmental NMRB | 1,633830 | –2,877505 | –7,442935 | 5,606502 | –1,783120 |
| 3 | EHSB | 6,103821 | 19,587435 | –0,240284 | 1,088213 | –0,511314 |
| 4 | AvgLinDistA | 11,556344 | 3,892028 | 9,720441 | –3,287205 | –11,031250 |
| 5 | Écart | 1,330890 | 2,686103 | 2,096598 | –1,327851 | 3,087055 |

TABLEAU 20

Pondération pour les nœuds de sortie de la version avancée

| nœud 1 (wx[i,0]) | nœud 2 (wx[i,1]) | nœud 3 (wx[i,2]) | nœud 4 (wx[i,3]) | nœud 5 (w4[i,4]) | écart (wy[5]) |
| --- | --- | --- | --- | --- | --- |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| –4,696996 | –3,289959 | 7,004782 | 6,651897 | 4,009144 | –1,360308 |

TABLEAU 21

Facteurs de graduation pour la sortie de la version avancée

|  |  |  |
| --- | --- | --- |
|  | bmin | bmax |

|  |  |  |
| --- | --- | --- |
| ODG | –3,98 | 0,22 |

# 7 Conformité des mises en œuvre

## 7.1 Généralités

Cette section propose un ensemble d'éléments de tests pour contrôler la conformité de la mise en œuvre de la méthode.

## 7.2 Sélection

Les éléments de tests ont été sélectionnés dans la base de données 3 (DB3), utilisées pour la validation des modèles. Afin de simplifier les tests, on a défini un sous‑ensemble des 84 éléments de DB3. Ce sous-ensemble contient 16 éléments. Le critère principal de cette sélection fut que les valeurs de MOV et de DI *(Indice de distorsion)* couvrent une large gamme.

## 7.3 Réglages pour les tests de conformité

Les éléments de tests peuvent être téléchargés sous forme de fichiers WAV (format Microsoft RIFF). Tous les éléments sont échantillonnés à 48 kHz, 16 bits PCM. Les signaux de référence et de tests fournis par l'UIT sont déjà adaptés les uns aux autres quant au temps et au niveau; il n'y a donc pas besoin de compensation en gain ou en retard. L'algorithme de mesure doit être réglé pour un niveau d'écoute de 92 dB SPL.

## 7.4 Intervalle de tolérance acceptable

Afin d'être conforme à la recommandation, les valeurs calculées de DI doivent reproduire les valeurs données dans les Tableaux 22 et 23, avec une tolérance inférieure à ± 0,02[[7]](#footnote-7)6 pour tous les éléments de test. Si une mise en œuvre ne donne pas des résultats situés dans cette fourchette de tolérance, elle n'est pas conforme à cette Recommandation.

## 7.5 Éléments de tests

Les tableaux qui suivent donnent le nom des éléments de référence et de tests[[8]](#footnote-8)7, les numéros de piste et les valeurs de DI qui en résultent. Le Tableau 22 concerne la version de base, le Tableau 23 la version avancée.

TABLEAU 22

Éléments de tests et valeurs DI qui en résultent pour la version de base

|  |  |  |
| --- | --- | --- |
| Élément | DI | ODG |

|  |  |  |
| --- | --- | --- |
| acodsna.wav | 1,304 | –0,676 |
| bcodtri.wav | 1,949 | –0,304 |
| ccodsax.wav | 0,048 | –1,829 |
| ecodsmg.wav | 1,731 | –0,412 |
| fcodsb1.wav | 0,677 | –1,195 |
| fcodtr1.wav | 1,419 | –0,598 |
| fcodtr2.wav | –0,045 | –1,927 |
| fcodtr3.wav | –0,715 | –2,601 |
| gcodcla.wav | 1,781 | –0,386 |
| icodsna.wav | –3,029 | –3,786 |
| kcodsme.wav | 3,093 | 0,038 |
| 1codhrp.wav | 1,041 | –0,876 |
| 1codpip.wav | 1,973 | –0,293 |
| mcodcla.wav | –0,436 | –2,331 |
| ncodsfe.wav | 3,135 | 0,045 |
| scodclv.wav | 1,689 | –0,435 |

TABLEAU 23

Éléments de tests et valeurs DI qui en résultent pour la version avancée

|  |  |  |
| --- | --- | --- |
| Élément | DI | ODG |

|  |  |  |
| --- | --- | --- |
| acodsna.wav | 1,632 | –0,467 |
| bcodtri.wav | 2,000 | –0,281 |
| ccodsax.wav | 0,567 | –1,300 |
| ecodsmg.wav | 1,594 | –0,489 |
| fcodsb1.wav | 1,039 | –0,877 |
| fcodtr1.wav | 1,555 | –0,512 |
| fcodtr2.wav | 0,162 | –1,711 |
| fcodtr3.wav | –0,783 | –2,662 |
| gcodcla.wav | 1,457 | –0,573 |
| icodsna.wav | –2,510 | –3,664 |
| kcodsme.wav | 2,765 | –0,029 |
| lcodhrp.wav | 1,538 | –0,523 |
| lcodpip.wav | 2,149 | –0,219 |
| mcodcla.wav | 0,430 | –1,435 |
| ncodsfe.wav | 3,163 | 0,050 |
| scodclv.wav | 1,972 | –0,293 |

Pièce jointe 1  
de l'Annexe 2  
  
Processus de validation

# 1 Généralités

En 1994, l'UIT-R a adopté la Question UIT-R 210/10 – Méthode objective d'évaluation de la qualité perceptuelle, et un Groupe de travail a été mis sur pied. Une des premières actions fut de lancer un appel à propositions, qui reçut six réponses proposant un modèle.

Beaucoup de travail fut nécessaire pour définir les procédures du processus de validation. On estima nécessaire de compiler une première base de données, DB1, composée de matériaux provenant des essais d'écoute déjà effectués. L'intérêt principal étant la qualité audio moyenne et haute, seuls les résultats des essais d'écoute conformes à la Recommandation UIT-R BS.1116 furent pris en considération. Les matériaux de ces essais étaient des matériaux de radiodiffusion critiques pour des codecs à faible débit binaire comme MPEG1 Couche II, MPEG1 Couche III, Dolby AC2, Mini Disc, et NICAM. On créa la base de données 1 pour donner aux créateurs des modèles une plate-forme commune, composée de matériaux couvrant une large gamme de dégradations, divers codecs et une dégradation due à des codecs en cascade. On trouvera dans la Pièce jointe 2 de l'Annexe 2 une description détaillée des essais compilés dans la base de données 1.

Il était évident qu'une méthode de mesure objective de la qualité du son perçu qui imiterait le comportement humain ne pouvait être validée que sur une base de données contenant les résultats d'essais subjectifs. Une validation appropriée exige une base de données fondée sur des matériaux inconnus. C'est pourquoi il fut nécessaire d'effectuer des nouveaux essais d'écoute. La méthode de mesure devant idéalement cibler tout type d'artefact pouvant apparaître dans les applications de radiodiffusion, il fallait intégrer non seulement les artefacts de codage, mais aussi des artefacts plus traditionnels comme la distorsion et le bruit. Les bases de données 2 et 3 furent créées respectivement en 1996 et 1997 pour satisfaire à ces exigences. Outre les codecs déjà intégrés dans DB1, on intégra Dolby AC-3 et AAC. On trouvera des détails complémentaires dans la Pièce jointe 2 de l'Annexe 2.

La validation devait prendre en compte les incertitudes, souvent présentées comme des intervalles de confiance, inhérentes aux essais d'écoute subjective. La taille de l'intervalle de confiance dépend d'un certain nombre de facteurs. Les plus importants sont: l'expérience des sujets, les procédures d'entraînement, le contexte dans lequel les éléments de test sont présentés, et le nombre de sujets.

L'adaptation et la validation de la méthode objective proposée dans cette Recommandation sont fondées sur un «auditeur expert moyen». Pour définir «l'auditeur expert moyen», on utilise les valeurs moyennes des évaluations de qualité subjectives et les intervalles de confiance à 95%.

Les essais d'écoute subjective sont très sensibles à différents facteurs qui ont une influence sur les résultats. Les SDG des bases de données 2 et 3 furent produites sur trois sites d'essais différents, et un certain nombre d'études ont cherché à savoir si les données pouvaient vraiment être associées. Ces études ne sont pas toutes parvenues à des résultats identiques mais il apparut raisonnable de fusionner les données et c'est cette base de données fusionnées qui servit de base à la validation.

Le processus de validation s'est effectué en trois phases:

– Phase 1: Phase de compétition

– Phase 2: Phase de collaboration

– Phase 3: Choix final

Ces phases sont décrites en détail dans les paragraphes qui suivent.

# 2 Phase de compétition

Six méthodes (DIX, NMR, PAQM, PERCEVAL, POM, TTA) ont été proposées pour la mesure objective de la qualité du son perçu et il fut décidé de comparer les performances de ces méthodes avec la base de données 2 et un sous-ensemble de la base de données 1. La base de données 2 fut générée début 1996. Le choix final des matériaux d'essai fut le résultat des efforts conjoints de SR (Suède) et de la BBC (Royaume-Uni). Les essais d'écoute furent effectués à NRK en Norvège, DR au Danemark et NHK au Japon. L'analyse statistique des données de ces essais fut préparée par Deutsche Telekom (Allemagne) et Teracom (Suède). Lors de la Phase 1, les données objectives étaient générées sur un site neutre (Swisscom, Suisse). Les créateurs des modèles reçurent alors la première moitié de la base de données 2 pour adaptation finale des méthodes (Phase 2). Enfin, des nouvelles notes différentielles objectives furent générées chez Swisscom.

L'analyse des performances des méthodes fut menée par Teracom (Suède) et par les créateurs eux‑mêmes. Les résultats de certaines méthodes proposées montraient une forte corrélation avec les SDG, mais un consensus fut néanmoins trouvé pour dire qu'aucune des méthodes ne satisfaisait aux exigences des utilisateurs. Une étude indépendante a montré qu'aucune des méthodes proposées n'était de façon significative meilleure que les autres. Il fut donc décidé de développer une méthode de mesure améliorée conjointement par tous les créateurs. Les performances de la nouvelle méthode devaient être comparées avec une des méthodes déjà en place dénommée modèle B3.

# 3 Phase de collaboration

La phase de collaboration était fondée sur l'idée suivante: associer les meilleurs éléments de chaque méthode dans une nouvelle méthode unique. Pour mieux satisfaire aux besoins des utilisateurs, il fut décidé de développer deux versions de la méthode: une adaptée aux réalisations en temps réel, et une autre, demandant plus de puissance de calcul, pour obtenir une plus grande précision.

La procédure de validation des deux nouvelles méthodes fut mise au point de la même façon que pour la phase de compétition. Il fallait créer une nouvelle base de données (DB3). Les éléments et conditions furent arrêtés au printemps 1997 et compilés à SR, Swisscom et BBC. On trouvera dans la Pièce jointe 2 de l'Annexe 2 une description complète de la base de données. Les essais d'écoute subjectifs eurent lieu sur trois sites d'essai, Deutsche Telekom, NHK et SR. Tous les sites ont appliqué la «méthode en double aveugle à référence cachée et stimulus triple», décrite dans la Recommandation UIT‑R BS.1116. Les résultats des essais d'écoute ont été collationnés en Suède. Une analyse statistique complète des résultats des essais d'écoute fut effectuée par Teracom et d'autres parties. Après cette analyse, certains auditeurs furent exclus de la suite de l'évaluation. On associa les résultats des sites d'essai pour créer la base de données 3.

À l'automne 1997, 52 éléments de la base de données furent donnés aux créateurs. Les nouvelles méthodes furent adaptées à ces nouvelles données. Comme plusieurs paramètres donnaient des résultats similaires le choix final fut retardé aussi longtemps que possible. Enfin, en Suisse, on utilisa les 32 éléments restants pour valider les nouvelles méthodes sur un ensemble de données «inconnu».

En outre, les résultats d'un nouvel essai d'écoute, mené par CRC (Canada) servirent à la validation des nouvelles méthodes sur des matériaux «inconnus». Les processus de choix et de vérification sont décrits dans les paragraphes qui suivent.

# 4 Vérification

On a effectué des essais complets des 18 versions spécifiées de la méthode de mesure objective. Les critères de sélection et les résultats des comparaisons entre les SDG et les résultats obtenus par les 18 versions de la méthode de mesure sont décrits dans ce paragraphe. L'objectif était de choisir et de vérifier les versions optimales qui seraient recommandées à l'UIT.

Critères de choix

Le critère le plus évident pour valider une méthode objective est la corrélation entre les résultats subjectifs et objectifs. En outre, on a introduit pour la validation deux critères supplémentaires qui prennent en compte la fiabilité des valeurs moyennes – la AES et les caractéristiques de tolérance.

La AES, introduite pour relier la précision du modèle à la précision du test d'écoute, se définit par l'expression suivante:

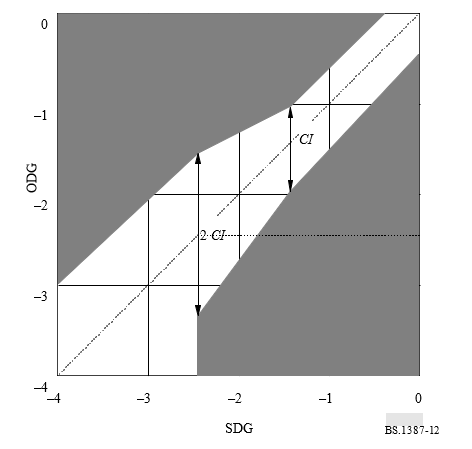
 si *CI*  0,25 alors *CI*  0,25

où *CI* est l'intervalle de confiance.

Les caractéristiques de tolérance sont conçues pour permettre diverses déviations des ODG par rapport aux SDG aux extrémités supérieure et inférieure de l'échelle des dégradations. La gamme tolérée dépend des intervalles de confiance des tests d'écoute. Cette gamme se limite à une valeur minimum de 0,25 graduations. On utilisa la distance entre les ODG hors caractéristiques de tolérance et les caractéristiques de tolérance pour évaluer la qualité de la méthode de mesure.

FIGURE 12

Caractéristiques de tolérance, intervalle de confiance *CI* ≥ *0,25*



## 4.1 Comparaison des valeurs de SDG et de ODG

Les mesures objectives ont été divisées en trois phases. Lors de la Phase 1, les 84 éléments de tests étaient inconnus de tout le monde, sauf du groupe de sélection. Lors de la Phase 2, des informations furent dévoilées sur 52 éléments. Ces informations étaient constituées des valeurs de SDG et des extraits sonores. Lors de la Phase 3 on a utilisé cette connaissance pour optimiser les performances des versions de la méthode. Il convient de noter que, par rapport à la Phase 1, quatre versions supplémentaires ont été évaluées lors de la Phase 3. Les valeurs de SDG présentées étaient calculées à partir de données générées par 75 sujets qualifiés.

Il existe diverses méthodes pour évaluer la façon dont les ODG sont corrélés aux SDG. Il n'y a malheureusement pas de valeur unique qui reflète la totalité des performances. On doit donc prendre en considération un certain nombre d'aspects. On trouvera les corrélations au § 4.2 et les AES au § 4.3. Le Modèle B3 est un des modèles testés par l'UIT-R en 1996 et il avait été décidé que les diverses nouvelles versions devaient être comparées à ce modèle antérieur.

## 4.2 Corrélation

Les chiffres de corrélation de la Phase 1 et de la Phase 3 sont représentés sur la Fig. 13 (84 éléments) et la Fig. 14 (32 éléments).

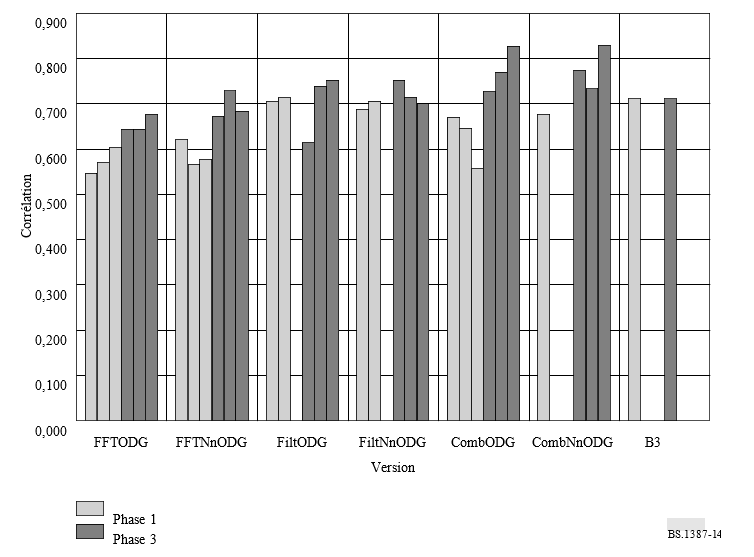
FIGURE 13

Corrélation entre les SDG et les ODG. Les 84 éléments sont intégrés



FIGURE 14

Corrélation entre les SDG et les ODG. Les 32 éléments non diffusés sont intégrés



## 4.3 Valeurs de l'erreur absolue (AES)

Un modèle qui produit en moyenne des valeurs d'ODG dans l'intervalle de confiance des SDG aura une valeur AES proche de 2. On trouvera dans la Fig. 15 et la Fig. 16 un résumé des valeurs AES.

FIGURE 15

AES des différentes versions.  
Les 84 éléments sont intégrés

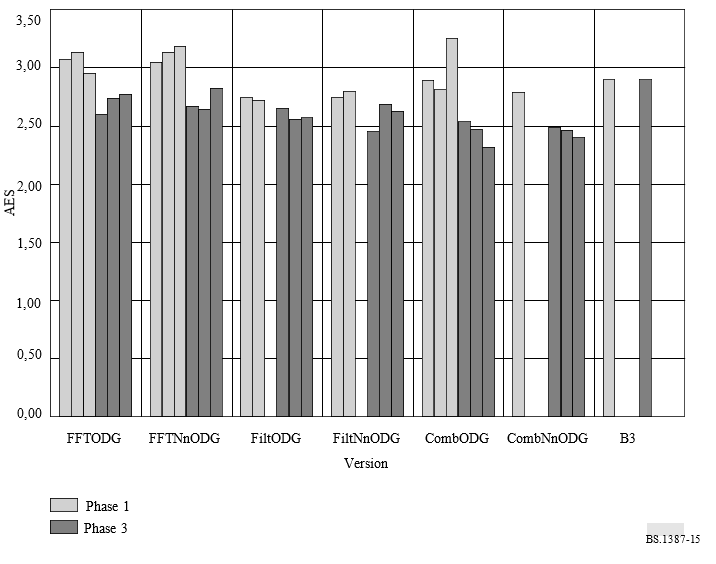
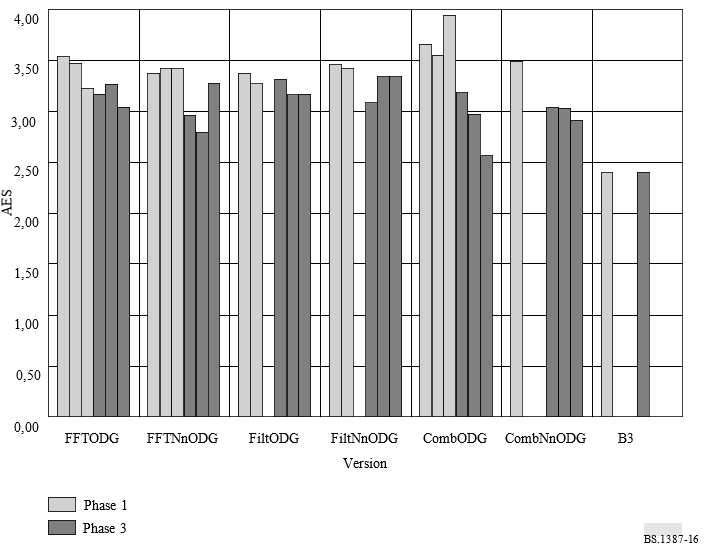


FIGURE 16

AES des différentes versions.   
Les 32 éléments non diffusés sont intégrés



## 4.4 Comparaison des ODG et de l'intervalle de confiance

La plupart des versions ont des performances similaires, comme on pouvait probablement s'y attendre. De nombreuses représentations ont été présentées lors de la réunion mais dans la présente Recommandation le répertoire est limité. On trouvera plus de détails dans le rapport complet des essais de vérification.

Les Figures 17 à 22 sont les représentations, pour le Modèle B3 et les versions FFTNnODG1 et CombNnODG3 du modèle, des moyennes de SDG, de l'intervalle de confiance et des ODG pour les 32 éléments non diffusés.







Les Figures 20 et 21 sont des représentations similaires, mais pour les 84 éléments de la Phase 3. En outre, la Fig. 22 est une illustration des performances de la version CombNnODG3.





## 4.5 Comparaison des ODG et de l'intervalle de confiance (tolérance)

L'UIT-R a défini une exigence pour l'utilisateur cible que l'on peut représenter par l'intervalle de confiance. Les exigences cibles sont plus drastiques pour les hauts niveaux de qualité audio que pour les niveaux de qualité audio inférieurs. Les Figures 23 à 25 illustrent les performances dans ce domaine du modèle B3 et des versions FFTNnODG1 et CombNnODG3 du modèle pour les 84 éléments lors de la Phase 3.







# 5 Sélection des versions optimales du modèle

Dix-huit versions différentes du modèle ont été évaluées sur le site des essais objectifs: six versions sont fondées sur la TFR, six versions utilisent un banc de filtres perceptuels adapté et six versions utilisent une combinaison de TFR et du banc de filtres. Le groupe TFR est destiné à être utilisé dans un système de mesure pouvant fonctionner en temps réel, les deux autres groupes nécessitent des calculs plus complexes et fourniront une meilleure précision. On peut diviser en deux sous‑groupes les six versions du modèle de chaque groupe: un qui utilisent les réseaux neuronaux et un qui ne les utilisent **pas**. On intègre également pour comparaison les performances du modèle de référence d'avril 1996 (B3) pour la base de données 3.

## 5.1 Critères de présélection fondés sur la corrélation

– On a évalué les performances des 18 versions du modèle et du modèle de référence B3 avec deux ensembles de données qui n'étaient pas utilisés majoritairement pour l'apprentissage des modèles (la Phase 1, la deuxième partie de la Phase 3 et CRC n'ont pas été utilisées).

– La base de données 3 a été créée spécialement pour l'évaluation des modèles perceptuels. Environ la moitié de cette base de données a été utilisée pour l'apprentissage des modèles. On a utilisé la corrélation entre les résultats subjectifs et les résultats objectifs pour les éléments restants (DB3\_2nd) pour évaluer les modèles objectifs.

– La base de données CRC a été créée par le CRC pour l'évaluation des schémas de codage sonores perceptuels. Cette base de données n'a pas été utilisée pour l'apprentissage des modèles perceptuels. La corrélation entre les résultats subjectifs et les résultats objectifs pour tous les éléments a été utilisée pour évaluer les modèles objectifs.

On trouvera dans les Tableaux 24 et 25 (corrélations et AES) les résultats de la base de données 3.

Lors de la phase de présélection, on a pris en compte toutes les entrées permettant une quantification des performances de toutes les versions du modèle avec une pondération plus importante pour DB3. Sur la base de la comparaison globale, il fut décidé d'effectuer une nouvelle comparaison des deux versions du modèle de chaque groupe qui semblaient donner les meilleurs résultats. On trouvera dans le Tableau 24 la corrélation de ces six versions (trois fois 2 versions du modèle).

TABLEAU 24

Corrélation entre les SDG et les ODG

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | FFTNnODG1 | FFTNnODG2 | FiltODG2 | FiltODG3 | CombNnODG3 | CombODG3 | B3 |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| DB3\_2nd | 0,671 | 0,728 | 0,738 | 0,751 | 0,828 | 0,826 | 0,710 |
| CRC | 0,837 | 0,779 | 0,862 | 0,839 | 0,851 | 0,777 | 0,656 |

TABLEAU 25

Valeurs de l'erreur absolue

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | FFTNnODG1 | FFTNnODG2 | FiltODG2 | FiltODG3 | CombNnODG3 | CombODG3 | B3 |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| DB3\_2nd | 2,96 | 2,79 | 3,16 | 3,16 | 2,91 | 2,56 | 2,39 |
| CRC | 1,55 | 1,85 | 1,61 | 1,67 | 1,61 | 1,90 | 2,78 |

Les deux versions TFR du modèle ont des performances différentes avec les deux bases de données. Il fallait des critères supplémentaires pour choisir une des deux versions. En prenant en compte le fait que la base de données CRC était totalement inconnue, FFTNnODG1 semble avoir de meilleures performances.

Les versions combinées du modèle donnent de meilleurs résultats avec les deux bases de données par rapport aux versions TFR. Mais, les résultats sont moins bons avec la base de données CRC par rapport aux versions banc de filtres. On peut dire que les versions banc de filtres sont un cas particulier des versions combinées où la pondération des variables de sortie de la version TFR est zéro. On a donc préféré les versions combinées du modèle.

## 5.2 Analyse du nombre d'éléments non conformes

L'intervalle de 95% autour de la moyenne sur de nombreux auditeurs exprime la précision des données subjectives provenant des tests d'écoute. On a également évalué les performances des versions du modèle en prenant en compte les éléments non conformes. On définit un élément comme non conforme si la différence entre les données subjectives et les données objectives est supérieure à deux fois l'intervalle de confiance.

On trouvera dans le Tableau 26 le nombre d'éléments non conformes pour les six versions du modèle avec DB3. Sensible signifie qu'une version du modèle indique une qualité sonore inférieure à celle indiquée par l'évaluation subjective, insensible signifie l'inverse.

TABLEAU 26

Éléments non conformes

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | FFTNnODG1 | FFTNnODG2 | FiltODG2 | FiltODG3 | CombNnODG3 | CombODG3 |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Sensible | 10 | 4 | 4 | 4 | 3 | 5 |
| Insensible | 13 | 13 | 11 | 13 | 12 | 14 |
| Total | 23 | 17 | 15 | 17 | 15 | 19 |

Si l'on considère le nombre d'éléments non conformes, la version FFTNnODG2 semble donner de meilleurs résultats que la version FFTNnODG1. La version CombNnODG3 donne les meilleurs résultats parmi les versions les plus précises.

## 5.3 Analyse de la gravité des non-conformités

TABLEAU 27

Éléments présentant des déviations de plus de 1,0 graduation  
entre les prédictions et les SDG

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Nombre d'éléments non conformes | 14 | 4 | 4 | 12 | 12 | 9 | 6 |
| Version du modèle | B3 | CombODG3 | CombNnODG3 | FFTNnODG1 | FFTNnODG2 | FiltODG2 | FiltODG3 |
| Élément | jcodsmg | jcodsmg | jcodsmg | jcodsmg | jcodsmg | jcodsmg | jcodsmg |
|  | qcodket | qcodket | pcodsmg | qcodket | qcodket | rcodpip | rcodpip |
|  | pcodsmg | pcodsmg | ccodclv | pcodsmg | pcodsmg | qcodket | qcodket |
|  | pcodcla | icodmar | acodclv | ocodsfe | pcodcla | pcodsmg | pcodsmg |
|  | ocodclv |  |  | ncodcas | ocodsfe | ocodsax | ocodsax |
|  | ncodcas |  |  | mcodcla | ncodglo | kcodpip | fcodsb2 |
|  | mcodpip |  |  | jcodveg | kcodcas | jcodveg |  |

TABLEAU 27 (*fin*)

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Nombre d'éléments non conformes | 14 | 4 | 4 | 12 | 12 | 9 | 6 |
| Version du modèle | B3 | CombODG3 | CombNnODG3 | FFTNnODG1 | FFTNnODG2 | FiltODG2 | FiltODG3 |
| Élément | jcodryc |  |  | jcodryc | jcodveg | fcodsb2 |  |
|  | icodtam |  |  | icodtam | jcodryc | fcodcl2 |  |
|  | icodsmg |  |  | icodmar | icodtam |  |  |
|  | fcodsb2 |  |  | fcodcl2 | ecodhrp |  |  |
|  | ecodhrp |  |  | ecodhrp | ccodclv |  |  |
|  | ccodclv |  |  |  |  |  |  |
|  | acodclv |  |  |  |  |  |  |

TABLEAU 28

Éléments présentant des déviations de plus de 1,5 graduation  
entre les prédictions et les SDG

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Nombre d'éléments non conformes | 8 | 2 | 2 | 3 | 2 | 2 | 2 |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Version du modèle | B3 | CombODG3 | CombNnODG3 | FFTNnODG1 | FFTNnODG2 | FiltODG2 | FiltODG3 |
| Élément |  | jcodsmg | jcodsmg | jcodsmg | jcodsmg | jcodsmg | jcodsmg |
|  | qcodket |  | pcodsmg |  |  |  |  |
|  | pcodsmg | pcodsmg |  | pcodsmg | pcodsmg |  |  |
|  |  |  |  |  |  | pcodsmg | pcodsmg |
|  |  |  |  | ncodcas |  |  |  |
|  | ncodcas |  |  |  |  |  |  |
|  | jcodryc |  |  |  |  |  |  |
|  | icodtam |  |  |  |  |  |  |
|  | icodsmg |  |  |  |  |  |  |
|  | ccodclv |  |  |  |  |  |  |
|  | acodclv |  |  |  |  |  |  |

# 6 Conclusion

Suite aux décisions prises lors du développement, deux versions du modèle, sur les dix‑huit versions proposées, ont été choisies pour satisfaire aux exigences des applications définies pour les méthodes de mesure objective. Une version simple, destinée à des réalisations en temps réel à moindre coût, et une version plus précise, qui ne fonctionne pas nécessairement en temps réel, ont été définies. Pour le processus de sélection, les critères définis ci-dessus ont été appliqués et analysés.

Pour la version en temps réel, on a choisi un modèle TFR, dénommé «FFTNnODG1», pour les raisons suivantes:

Aucune des versions TFR ne présentait un avantage significatif par rapport aux autres sur un quelconque des critères ci-dessus. En ce qui concerne les coefficients de corrélation entre les ODG et les SDG, et le nombre et la gravité des non conformités, chacune des versions évaluées a ses avantages et ses inconvénients. Néanmoins, il faut noter que «FFTNnODG1» propose la meilleure corrélation avec la base de données CRC'97 qui était totalement inconnue (0,837).

Pour la version de plus grande précision, on a préféré la version combinée, qui intègre une TFR et un banc de filtres, parce que cette approche intègre également le sous-ensemble d'un modèle purement banc de filtres et devrait donc avoir de meilleurs résultats. Six versions différentes d'un modèle combiné ont été évaluées lors du processus de sélection. La version choisie, «CombNnODG3», présente moins de non conformités et une meilleure corrélation (r  0,851 avec CRC'97) que les autres versions. La corrélation de cette version pour la totalité de la base de données 3 avait le même ordre de grandeur qu'une des versions «CombODG3», mais avait une meilleure corrélation que les autres versions.

Pièce jointe 2  
de l'Annexe 2  
  
Description des bases de données de référence

# 1 Introduction

Lors du développement de la méthode de mesure objective de la qualité du son perçu, on a utilisé un certain nombre de bases de données pour l'apprentissage et la validation.

Certaines des bases de données contenaient des données casque et haut-parleur mélangées et, certaines uniquement des données casque. Pour les bases de données ayant des ensembles distincts de données casque et de données haut-parleur, on n'a utilisé que les données casque.

Un élément est un fragment sonore utilisé pour une évaluation subjective. Une condition est une dégradation unique. Tous les éléments ont été utilisés pour toutes les conditions expérimentales sauf pour les études avec DB2 et DB3. DB3 a été utilisée en partie pour l'apprentissage et en partie pour la validation (52 des 84 éléments ont été utilisés pour l'apprentissage lors de la deuxième phase de validation).

Apprentissage

– MPEG90

– La moyenne des SDG par élément couvrait uniformément une gamme de 0,0 à −4,0.

– Référence: ISO/IEC JTC 1/SC 2/WG 11 MPEG/Audio test report, Document MPEG90/N0030, octobre 1990.

– MPEG91

– Au moins 88% de la moyenne des SDG par élément étaient supérieurs à –2,0, et la gamme s'étendait de 0,1 à −3,8.

– Référence: ISO/IEC JTC 1/SC 2/WG 11 MPEG/Audio test report, Document MPEG91/N0010, juin 1991.

– ITU92DI

– 80% de la moyenne des SDG par élément étaient supérieurs à −2,0, et la gamme s'étendait de 0,1 à −3,4.

– ITU92CO

– Au moins 96% de la moyenne des SDG par élément étaient supérieurs à −2,0, et la gamme s'étendait de 0,2 à −2,4.

– ITU93

– La plupart de la moyenne des SDG par élément étaient supérieurs à −2,0, et la gamme s'étendait de −0,1 à −2,3. Il n'y avait pas de différence significative entre les données de ces deux laboratoires.

– Références: Grusec et coll. [1997].

– MPEG95

– Au moins 63% de la moyenne des SDG par élément étaient supérieurs à −2,0, et la gamme s'étendait de −0,2 à −3,8.

– Référence: Meares et Kim [1995].

– EIA95

– Au moins 93% de la moyenne des SDG par élément étaient supérieurs à −2,0, et la gamme s'étendait de 0,1 à −3,7.

– Référence: Grusec et coll. [1997].

– DB2

– Les éléments n'ont pas tous été utilisés dans toutes les conditions.

Validation

– DB3

– Les éléments n'ont pas tous été utilisés dans toutes les conditions.

– CRC97

– La moyenne des SDG par élément couvrait uniformément une gamme de 0,1 à −3,6.

– Référence: Soulodre et autres [1998].

Les paragraphes qui suivent décrivent les éléments intégrés aux différentes bases de données et les conditions appliquées.

# 2 Éléments par base de données



# 3 Conditions expérimentales

Pour tous les débits binaires comportant l'indication kbit/s stéréo, on donne le débit binaire total, ex. 256 kbit/s stéréo signifie que 256 kbit/s sont alloués au total pour les deux canaux d'un signal stéréo. Si rien d'autre n'est indiqué, stéréo renvoie à un codage indépendant par canal.

## 3.1 MPEG90

Trois débits binaires: 64 kbit/s mono, 192 kbit/s et 256 kbit/s stéréo, les matériaux n'étaient pas tous disponibles pour cette base de données.

– Musicam.

– SB-ADPCM.

## 3.2 MPEG91

Trois débits binaires: 64 kbit/s mono, 192 kbit/s, et 256 kbit/s stéréo.

– MPEG1 Couche I.

– MPEG1 Couche II.

– MPEG1 Couche III.

– MUSICAM.

– ASPEC.

– NICAM.

## 3.3 ITU92DI

Cinq codecs de distribution: 240 kbit/s stéréo.

Chaque élément a été traité par le même codec trois fois en tandem, avec une perte en niveau de 0,1 dB avant chaque passe.

– MPEG1 Couche II.

– MPEG1 Couche III.

– Dolby AC-2.

– Aware.

– NHK.

## 3.4 ITU92CO

Six codecs de contribution: 360 kbit/s stéréo. Chaque élément a été traité par le même codec trois fois en tandem, avec une perte en niveau de 0,1 dB avant chaque passe.

– MPEG1 Couche II.

– MPEG1 Couche III.

– Dolby AC-2.

– Dolby Low-Delay.

– Aware.

## 3.5 ITU93

Configurations codec tandem MPEG1 Couche II:

– Codec Émission uniquement à 256 kbit/s stéréo.

– Codec Émission uniquement à 192 kbit/s stéréo (codage stéréo commun).

– Huit codecs de contribution à 360 kbit/s suivis d'un codec d'émission à 256 kbit/s, tous stéréo.

– Huit codecs de contribution à 360 kbit/s suivis d'un codec d'émission à 192 kbit/s, tous stéréo.

– Cinq codecs de contribution à 360 kbit/s suivis de trois codecs de distribution à 240 kbit/s et un codec d'émission à 256 kbit/s, tous stéréo.

– Cinq codecs de contribution à 360 kbit/s suivis de trois codecs de distribution à 240 kbit/s et un codec d'émission à 192 kbit/s, tous stéréo.

## 3.6 MPEG95

Réalisations codec (64 kbit/s):

– Vingt-deux variations d'encodage ont été choisies dans un ensemble plus important de méthodes d'encodage disponible dans 6 codecs pour réaliser un sous-ensemble de 4 modèles temps/fréquence basse résolution et 17 modèles temps/fréquence haute résolution.

– Les organisations participantes étaient AT&T, Fraunhofer, Sony, GCL, RAI/Alcatel, et Philips.

– Tous les éléments étaient des enregistrements mono présentés sur deux canaux.

## 3.7 EIA95

– Eureka 147/MPEG1 Couche II #1 224 kbit/s stéréo (codage stéréo commun)

– Eureka 147/MPEG1 Couche II #2 192 kbit/s stéréo (codage stéréo commun)

– AT&T/Lucent 160 kbit/s stéréo

– AT&T/Lucent/Amati #1 128 kbit/s stéréo

– AT&T/Lucent/Amati #2 160 kbit/s stéréo

– VOA/JPL 160 kbit/s stéréo

– USADR-FM #1 128-256 kbit/s stéréo (débit binaire variable)

– USADR-FM #2 128-256 kbit/s stéréo (débit binaire variable)

– USADR-AM 96 kbit/s stéréo

## 3.8 DB2

– MPEG1 Couche II 256 kbit/s stéréo, 1, 3, 5, 7, et 9 étages

– Dolby AC2 256 kbit/s stéréo, 1, 3, 5, 7, et 9 étages

– MPEG1 Couche II 192 kbit/s stéréo (codage stéréo commun)

– MPEG1 Couche II 64 kbit/s mono

– MPEG2 Couche II 64 kbit/s mono

– MPEG1 Couche II 384 kbit/s stéréo

– MPEG1 Couche III 128, 160, 192 kbit/s, tous stéréo

– APT-X 256 et 384 kbit/s les deux stéréo

– Quantification distorsion

– Enregistrement analogique 1, 2, 3 étages

– Écrêtage

## 3.9 DB3

– NICAM

– MiniDisc et MiniDisc + Couche II 192 kbit/s, stéréo (codage stéréo commun)

– Dolby AC2 256 kbit/s stéréo, 1, 3, 5, 7, et 9 étages

– MPEG1 Couche II sélection de la base de données Swisscom, > 192 kbit/s stéréo

– MPEG1 Couche III 128 et 160 kbit/s les deux stéréo (codage stéréo commun)

– MPEG AAC 128 kbit/s stéréo (codage stéréo commun)

– MPEG Couche III 128 + Couche II, 384 + Couche II, 224 kbit/s, tous stéréo

– Dolby AC3 256 kbit/s stéréo

– Dolby AC3 256 + MPEG Couche II, 224 kbit/s, les deux stéréo

– Quantification distorsion

– THD

– Bruit

## 3.10 CRC97

– AT&T PAC 64, 96, 128, et 160 kbit/s, tous stéréo

– Dolby AC3 128, 160, et 192 kbit/s, tous stéréo

– MPEG1 Couche II software 128, 160, et 192 kbit/s, tous stéréo

– MPEG1 Couche II hardware (ITIS) 96, 128, 160, 192 kbit/s, tous stéréo

– MPEG4 AAC 96 et 128 kbit/s, les deux stéréo

– MPEG1 Couche III 128 kbit/s stéréo

# 4 Éléments par condition pour DB2 et DB3

## 4.1 DB2

**Condition N°** **Éléments**

Site test I, NHK Japon

Couche II, 256 kbit/s 1 étage CO13 CLA,RYC,SB1,STR

3 étages CO11 CLA,RYC,SB1,STR

5 étages CO19 CLA,RYC,SB1,STR

7 étages CO18 CLA,RYC,SB1,STR

9 étages CO15 CLA,RYC,SB1,STR

NBC (Dolby AC2) 1 étage CO1A CAS,RYC,STR,WIN

3 étages CO12 CAS,RYC,STR,WIN

5 étages CO17 CAS,RYC,STR,WIN

7 étages CO16 CAS,RYC,STR,WIN

9 étages CO14 CAS,RYC,STR,WIN

**Condition N°** **Éléments**

Site test II, DR Danemark

Couche II, 256 kbit/s 1 étage CO2B CLA,RYC,SB1,STR

Couche II, 192 kbit/s js CO25 CLA,RYC,SB1,STR

Couche II, 64 kbit/s mono CO27 MLA,MPE,MTR,MYC

NBC (Dolby AC2) 5 étages CO29 CAS,RYC,STR,WIN

MPEG2/L2 LSF CO22 MLA,MPE,MTR,MYC

Analogique 1 CO23 PER

Analogique 2 CO2A PER

Analogique 3 CO28 PER

Erreurs 1 CO24 GLO,HRN,TRI

Erreurs 2 CO21 GLO,HRN,TRI

Écrêtage CO26 BAS,CL2,TUB

Site test III, NRK Norvège

Couche II, 384 kbit/s CO34 CLA,RYC,SB1,STR

Couche II, 256 kbit/s 1 étage CO31 CLA,RYC,SB1,STR

NBC (Dolby AC2) 5 étages CO3B CAS,RYC,STR,WIN

Couche III (ASPEC3), 192 kbit/s CO32 CLA,STR,TAM,VEG

Couche III (ASPEC3), 128 kbit/s CO39 CLA,STR,TAM,VEG

Couche III (ASPEC3), 160 kbit/s CO3A CLA,STR,TAM,VEG

APT-X, 256 kbit/s CO33 HAR,SB2,STR,TPT

APT-X, 384 kbit/s CO36 HAR,SB2,STR,TPT

Quantification dist. 1 CO35 DRU

Quantification dist. 2 CO37 DRU

Quantification dist. 3 CO38 DRU

Éléments de tests

STR Musique populaire suédoise, enregistrement SR, utilisé auparavant

SB1 Cornemuses, enregistrement SR

SB2 Cornemuses, enregistrement SR

CLA Clarinette, SQUAM 16/2

TAM Tambourin, enregistrement SR, utilisé auparavant

WIN Stravinski, ensemble de vents, utilisé auparavant

TPT Trompette, SQUAM 21/2

HAR Ensemble Harlequin, enregistrement BBC G 49/17

VEG Suzanne Vega, master ancien, utilisé auparavant

CAS Castagnettes, SQUAM 27

SPE Voix allemande, SQUAM 54

RYC Ry Cooder, CD: JAZZ tr 11 (0,25 – 0,47)

PER Percussion, Marimba Basse Japonaise, CD: Sony/CBS 32DC 5027

HRN Cor, SQUAM 23/2

GLO Glockenspiel, SQUAM 35/1, utilisé auparavant

TRI Triangle, SQUAM 32/2

DRU Tambours, SQUAM 28

CL2 Clarinette, SQUAM 16/2

BAS Clarinette Basse, SQUAM 17

TUB Tuba, SQUAM 24

MPE Mixage mono de SPE

MTR Mixage mono de STR

MLA Mixage mono de CLA

MYC Mixage mono de RYC



**Test/Élément**

## 4.2 DB3



Glossaire

Valeur de l'erreur absolue (AES)

La valeur de l'erreur absolue est dérivée d'une formule développée spécialement pour l'évaluation de la qualité des résultats obtenus à partir d'une méthode de mesure perceptuelle objective. Il prend en compte les intervalles de confiance des valeurs moyennes des tests d'écoute subjective.

Qualité audio de base

La qualité audio de base est un attribut subjectif global qui intègre toutes les différences détectées entre le Signal de référence et sa version traitée.

Marge de codage

La marge de codage est un paramètre de la qualité qui mesure la distance entre les artefacts de codage inaudibles et le seuil où ces artefacts deviennent audibles.

Variables de sortie de modèle (MOV)

Les variables de sortie de modèle sont des valeurs de sortie intermédiaires de la méthode de mesure perceptuelle. Ces variables sont fondées sur les découvertes psychoacoustiques de base et peuvent donc servir à définir plus avant les artefacts de codage.

Note différentielle objective (ODG)

La note différentielle objective est le paramètre de sortie principal de la méthode de mesure perceptuelle. Elle correspond à la SDG et est le paramètre de mesure qui donne la qualité sonore de base globale. L'ODG est située entre 0 et −4.

Mesure Off-line

Procédure de mesure qui n'interfère pas avec la transmission d'un programme en cours.

Mesure en ligne

Procédure de mesure qui repose sur la transmission de programmes en cours, ou partie de cette transmission.

Note différentielle subjective (SDG)

Dans un test d'écoute conforme à la Recommandation UIT-R BS.1116 la qualité audio de base de la référence cachée et de la version traitée de la référence sont notées sur l'échelle de dégradation à cinq notes. La note différentielle est la note donnée au Signal testé moins la note donnée au Signal de référence. Idéalement, la SDG doit être située entre 0 et −4. Si la référence n'a pas été identifiée correctement, la SDG est positive.

Abréviations

ADB bloc distordu moyen (*average distorted block*)

AES valeur de l'erreur absolue (*absolute error score*)

ASD différence auditive spectrale (*auditory spectral difference*)

Avg moyenne (linéaire) (*average (linear*)

BAQ qualité audio de base (*basic audio quality*)

Bw largeur de bande (*bandwidth*)

CI intervalle de confiance (*confidence interval*)

CM marge de codage (*coding margin*)

DBn base de données (*database*) *n* (1, 2 ou 3)

DC courant direct (*direct current*)

DIX indice de perturbation (*disturbance index*)

DUT système testé (*device under test*)

EHS structure harmonique d'erreur (*error harmonic structure*)

ERB largeur de bande rectangulaire équivalente (*equivalent rectangular bandwidth*)

fac facteur

FIR réponse impulsionnelle finie (*finite impulse response*)

IIR réponse impulsionnelle infinie (*infinite impulse response*)

ISO Organisation internationale de normalisation (*International Standards Organization*)

JNLD différence de niveau juste audible (*just noticeable level difference*)

MFPD probabilité de détection maximum filtrée (*maximum filtered probability of detection*)

MOV variable de sortie de modèle (*model output variable*)

MPEG moving picture expert group (*moving picture expert group*)

NL intensité acoustique du bruit (*noise loudness*)

NMR rapport bruit/masque (*noise-to-mask ratio*)

OASE évaluation objective du signal sonore (*objective audio signal evaluation*)

OCM marge de codage objective (*objective coding margin*)

ODG note différentielle objective (*objective difference grade*)

PAQM mesure perceptuelle de la qualité du son (*perceptual audio quality measure*)

PEAQ mesure objective de la qualité du son perçu (*objective measurements of perceived audio quality*)

PERCEVAL évaluation perceptuelle *(perceptual evaluation*)

POM mesure perceptuelle objective (*perceptual objective measure*)

REF signal de référence (*reference signal*)

res Résolution

r.m.s. moyenne quadratique (*root mean squared*)

ROEX exponentielle arrondie (*rounded exponential*)

ROV débit des valeurs de sortie (*rate of output values*

SCM marge de codage subjective (*subjective coding margin*)

SDG note différentielle subjective (*subjective difference grade*)

*S*/*N* rapport signal/bruit (*signal-to-noise ratio*)

SPL niveau de pression sonore (*sound pressure level*)

SUT signal testé (*signal under test*)

TFD transformation de Fourier discrète (*discrete Fourier transform*)

TFR transformation de Fourier rapide (*fast Fourier transform*)

THD distorsion harmonique totale (*total harmonic distortion*)

Win moyenne avec fenêtrage (*windowed average*)

Références

AURAS, W. [septembre, 1984] Berechnungsverfahren für den Wohlklang beliebiger Schallsignale, ein Beitrag zur gehörbezogenen Schallanalyse. Dissertation an der Fakultät für Elektrotechnik der Technischen Universität München, République fédérale d'Allemagne.

BEERENDS, J. G. et STEMERDINK, J. A. [décembre, 1992] A perceptual audio quality measure based on a psychoacoustic sound representation. *J. Audio Eng. Soc.*, Vol. 40, p. 963‑978.

BEERENDS, J. G. et STEMERDINK, J. A. [février, 1994] Modeling a cognitive aspect in the measurement of the quality of music codecs. Contribution to the 96th AES Convention, preprint 3800. Amsterdam, Pays-Bas.

BEERENDS, J. G. et STEMERDINK, J. A. [mars, 1994] A perceptual speech quality measure based on a psychoacoustic sound representation. *J. Audio Eng. Soc.*, Vol. 42, p. 115-123.

BEERENDS, J. G., van den BRINK, W. A. C. et RODGER, B. [mai, 1996] The role of informational masking and perceptual streaming in the measurement of music codec quality. Contribution to the 100th AES Convention, preprint 4176. Copenhague, Danemark.

BRANDENBURG, K. [1987] Evaluation of quality for audio encoding at low bit rates. Contribution to the 82nd AES Convention, preprint 2433. Londres, Royaume-Uni.

BREGMAN, A. S. [1990] *Auditory Scene Analysis: The Perceptual Organisation of Sound*, MIT Press, Cambridge MA, États-Unis d'Amérique.

COHEN, E. A. et FIELDER, L. D. [mai, 1992] Determining noise criteria for recording environments. *J. Audio Eng. Soc.*, Vol. 40, p. 384-402.

COLOMES, C., LEVER, M., RAULT, J. B. et DEHERY, Y. F. [avril, 1995] A perceptual model applied to audio bit-rate reduction. *J. Audio Eng. Soc.*, Vol. 43, p. 233-240.

FEITEN, B. [mars, 1997] Measuring the Coding Margin of Perceptual Codecs with the Difference Signal. 102nd AES‑Convention preprint 4417. Munich, République fédérale d'Allemagne.

GRUSEC, T., THIBAULT, L. et SOULODRE, G. [septembre, 1997] EIA/NRSC DAR systems subjective tests. Part 1: Audio codec quality. *IEEE Trans. on Broadcasting*, Vol. 43, 3.

KARJALAINEN, J. [mars, 1985] A new auditory model for the evaluation of sound quality of audio system. Proceedings of the ICASSP, p. 608-611. Tampa, Floride, États‑Unis d'Amérique.

LEEK, M. R. et WATSON, C. S. [1984] Learning to detect auditory pattern components. *J. Acoust. Soc. Am.*, Vol. 76, p. 1037‑1044.

MEARES, D. J. et KIM, S. W. [juillet, 1995] «NBC time/frequency module subjective tests: overall results», ISO/IEC JTC 1/SC 29/WG 11 N0973 MPEG95/208.

MOORE, B. C. [1986] *Frequency Selectivity in Hearing*. Academic Press. Londres, Royaume‑Uni.

MOORE, B. C. [1989] *An introduction to the psychology of hearing*. Academic Press. Londres, Royaume‑Uni.

PAILLARD, B., MABILLEAU, P., MORISETTE, S. et SOUMAGNE, J. [1992] Perceval: Perceptual evaluation of the quality of audio signals. *J. Audio Eng. Soc.*, Vol. 40, p. 21-31.

SCHROEDER, M. R., ATAL, B. S. et HALL, J. L. [décembre, 1979] Optimizing digital speech coders by exploiting masking properties of the human ear. *J. Acoust. Soc. Am*., Vol. 66, p. 1647‑1652.

SOULODRE, G., GRUSEC, T., LAVOIE, M. et THIBAULT, L. [mars, 1998] Subjective evaluation of state-of-the-art 2-channel audio codecs. *J. Audio Eng. Society*.

SPORER, T. [octobre, 1997] Objective audio signal evaluation – applied psychoacoustics for modeling the perceived quality of digital audio. 103rd AES-Convention, preprint 4512. New York, États‑Unis d'Amérique.

TERHARDT, E. [1979] Calculating Virtual Pitch, Hearing Research. Vol. 1, p. 155-182.

THIEDE, T. et KABOT, E. [1996] A New Perceptual Quality Measure for Bit Rate Reduced Audio. Contribution to the 100th AES Convention, preprint 4280. Copenhague, Danemark.

TREURNIET, W. C. [1996] Simulation of individual listeners with an auditory model. Proceedings of the Audio Engineering Society, Reprint Number 4154. Copenhague, Danemark.

von BISMARCK, G. [1974] Sharpness as an attribute of the timbre of steady sounds. *Acústica*, 30, p. 159‑172.

ZWICKER, E. et FASTL, H. [1990] *Psycho-acoustics, Facts and Models*. Berlin; Heidelberg: Springer Verlag, République fédérale d'Allemagne.

ZWICKER, E. et FELDTKELLER, R. [1967] *Das Ohr als Nachrichtenempfänger*. Stuttgart: Hirzel Verlag, République fédérale d'Allemagne.

Bibliographie

GRUSEC, T., THIBAULT, L. et SOULODRE, G. [1995] Subjective evaluation of high quality audio coding systems: methods and results in the two-channel case. Preprint 4065 (F‑5), Proceedings of the AES. New York, États-Unis d'Amérique.

1. 1 Voir la Recommandation UIT-R BS.1116. [↑](#footnote-ref-1)
2. 2 Les inventeurs de la technologie décrite dans la présente Recommandation ont déposé des brevets conformément à l'Annexe 1 de la Résolution UIT‑R 1. La technologie décrite dans la présente Recommandation est protégée par des brevets internationaux, et comme toutes les Recommandations de l'UIT, soumise aux droits d'auteur. L'accord préalable des propriétaires, sous la forme d'une licence, est obligatoire pour pouvoir exploiter cette technologie. Pour toute information complémentaire sur les droits de licence de cette technologie, consulter la base de données brevets de l'UIT-R, ou le Secrétariat du BR. [↑](#footnote-ref-2)
3. 3 Le retard supplémentaire d'un échantillon n'est pas nécessaire pour une réalisation. Il convient de noter que la réalisation de référence utilisée pour les essais de conformité inclut ce retard supplémentaire. [↑](#footnote-ref-3)
4. 4 En fait, les enveloppes des filtres dans les bandes de hautes fréquences ne sont pas nécessairement conformes au théorème d'échantillonnage. Même si l'aliasing n'apparaît que dans certaines conditions très particulières (à savoir composantes de hautes fréquences modulées avec des fréquences supérieures à 1,5 kHz) et si les problèmes liés à ces effets ne sont jamais apparus dans les bases de données connues, il faut préciser qu'il peut y avoir des problèmes d'aliasing en particulier avec des signaux d'essai synthétisés. [↑](#footnote-ref-4)
5. 5 Le terme de «bloc» est équivalent à celui de «trame» dans ce contexte. [↑](#footnote-ref-5)
6. \* Ce nombre se rapporte à des données d'entrée en format d'entiers signés de 16 bits dans une étendue de ‑32 768 à 32 767, comme cela est le cas sur un disque compact. [↑](#footnote-ref-6)
7. 6 Pour atteindre cette précision, utiliser une arithmétique IEEE en virgule flottante. [↑](#footnote-ref-7)
8. 7 Les noms des éléments de référence correspondants sont calculés par remplacement de la sous‑chaîne «cod» dans les noms des éléments de test par la sous‑chaîne «ref». Par exemple, l'élément de référence pour le fichier «bcodtri.wav» devient «breftri.wav». [↑](#footnote-ref-8)