



UNION INTERNATIONALE DES TÉLÉCOMMUNICATIONS

**UIT-T**

**P.830**

SECTEUR DE LA NORMALISATION  
DES TÉLÉCOMMUNICATIONS  
DE L'UIT

(02/96)

**QUALITÉ DE LA TRANSMISSION TÉLÉPHONIQUE  
MÉTHODES D'ÉVALUATION OBJECTIVE  
ET SUBJECTIVE DE LA QUALITÉ**

---

**ÉVALUATION SUBJECTIVE DE LA QUALITÉ  
DES CODECS NUMÉRIQUES À BANDE  
TÉLÉPHONIQUE ET À LARGE BANDE**

**Recommandation UIT-T P.830**

(Antérieurement «Recommandation du CCITT»)

---

## AVANT-PROPOS

L'UIT-T (Secteur de la normalisation des télécommunications) est un organe permanent de l'Union internationale des télécommunications (UIT). Il est chargé de l'étude des questions techniques, d'exploitation et de tarification, et émet à ce sujet des Recommandations en vue de la normalisation des télécommunications à l'échelle mondiale.

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

L'approbation des Recommandations par les Membres de l'UIT-T s'effectue selon la procédure définie dans la Résolution n° 1 de la CMNT (Helsinki, 1<sup>er</sup>-12 mars 1993).

La Recommandation révisée UIT-T P.830, que l'on doit à la Commission d'études 12 (1993-1996) de l'UIT-T, a été approuvée le 6 février 1996 selon la procédure définie dans la Résolution n° 1 de la CMNT.

---

### NOTE

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

© UIT 1996

Droits de reproduction réservés. Aucune partie de cette publication ne peut être reproduite ni utilisée sous quelque forme que ce soit et par aucun procédé, électronique ou mécanique, y compris la photocopie et les microfilms, sans l'accord écrit de l'UIT.

## TABLE DES MATIÈRES

|    |   | <i>Page</i> |
|----|---|-------------|
| 1  | Domaine d'application.....  | 1           |
| 2  | Références.....   | 1           |
| 3  | Définitions.....  | 2           |
| 4  | Abréviations.....   | 2           |
| 5  | Conventions.....  | 3           |
| 6  | Résumé des procédures d'évaluation subjective.....  | 3           |
| 7  | Enregistrements de la source.....   | 3           |
|    | 7.1 Système d'enregistrement.....   | 3           |
|    | 7.2 Echantillons vocaux.....  | 4           |
| 8  | Choix des paramètres d'expérience.....  | 5           |
|    | 8.1 Conditions requises pour les codecs.....  | 5           |
|    | 8.2 Conditions de référence.....  | 11          |
| 9  | Conception des essais.....  | 14          |
| 10 | Procédure des essais d'écoute.....  | 14          |
|    | 10.1 Système de réception.....  | 14          |
|    | 10.2 Echelles d'appréciation subjective.....  | 14          |
|    | 10.3 Bruit électrique.....  | 15          |
| 11 | Analyse des résultats.....  | 15          |
|    | Annexe A – Comparaison des différentes définitions du rapport SNR.....                                  | 17          |
|    | Annexe B – Ensemble de paramètres permettant de déterminer la qualité de fonctionnement des codecs..... | 17          |
|    | B.1 Systèmes à bande étroite (300-3400 Hz).....   | 18          |
|    | B.2 Systèmes à large bande (100-7000 Hz).....   | 18          |
|    | Annexe C – Codec MIC correspondant à une seule unité qdu.....   | 19          |
|    | C.1 Interface analogique commune.....   | 19          |
|    | C.2 Interface numérique.....  | 19          |
|    | C.3 Interface analogique.....   | 19          |
|    | Annexe D – Caractéristiques d'émission et de réception de l'appareil IRS modifié.....                   | 20          |
|    | Bibliographie.....  | 21          |

## RÉSUMÉ

La présente Recommandation révisée décrit des méthodes et des procédures permettant d'effectuer des évaluations subjectives de la qualité de transmission de codecs vocaux numériques.

Les révisions reprises dans cette version de la Recommandation consistent à lui ajouter de nouvelles informations reflétant les pratiques actuelles en matière d'évaluation subjective de la qualité des codecs numériques, en développant l'article relatif à la création d'enregistrements de la source et en ajoutant deux annexes. Une des annexes décrit une réalisation de codec MIC (en loi A/ $\mu$ ) qui produit une seule unité de distorsion de quantification (qdu) affectant les signaux d'entrée. L'autre annexe nouvelle décrit les caractéristiques d'émission et de réception de l'appareil IRS modifié. L'usage de ces caractéristiques est recommandé en tant que réponses en émission et en réception lorsque le codec examiné est destiné à être utilisé dans des circuits entièrement numériques.

## INTRODUCTION

La téléphonie numérique est désormais courante dans le réseau téléphonique public commuté (RTPC). Depuis les années 60, la transmission numérique dans le RTPC est synonyme de modulation par impulsions et codage (MIC) en loi A ou en loi  $\mu$  à 64 kbit/s (voir la Recommandation G.711). La combinaison d'une meilleure connaissance des techniques de traitement des signaux avec les progrès technologiques, notamment des techniques d'intégration LSI/VLSI, a fait que l'on s'intéresse davantage à des méthodes de codage plus efficaces que la MIC. Des exemples de cette évolution sont l'introduction de la MICDA à 32 kbit/s (Recommandation G.726), la prédiction linéaire à faible délai avec excitation par code (LDCELP) à 16 kbit/s (Recommandation G.728) ainsi qu'un codec à large bande de 7 kHz (Recommandation G.722). On accorde actuellement un intérêt considérable au codage de la parole à des débits binaires inférieurs à 16 kbit/s.

La méthode utilisée le plus souvent pour évaluer la qualité des codecs numériques consiste à les soumettre à des essais subjectifs. Lorsque le conduit de transmission est numérique et/ou non linéaire, de simples mesures objectives, comme celles qui sont spécifiées dans la Recommandation G.712, sont insuffisantes pour assurer une qualité de transmission adéquate. Le but d'une méthode d'essai subjective est de mesurer la dégradation à laquelle la partie non linéaire du conduit de transmission contribue, afin que la qualité de l'ensemble du système soit satisfaisante. A cet effet, les mesures doivent être:

- a) fiables; et
- b) effectuées de telle sorte qu'elles tiennent compte des principales interactions entre la partie linéaire et les autres parties du système de transmission.

Cela implique aussi bien la possibilité d'imputer sans ambiguïté une contribution numérique à chaque processus numérique que la capacité d'utiliser cette contribution, conjointement avec d'autres dégradations, pour estimer la qualité des connexions téléphoniques.

## MOTS CLÉS

Essais subjectifs, évaluation subjective.

# ÉVALUATION SUBJECTIVE DE LA QUALITÉ DES CODECS NUMÉRIQUES À BANDE TÉLÉPHONIQUE ET À LARGE BANDE

(Helsinki, 1993; révisée en 1996)

## 1 Domaine d'application

La présente Recommandation est fondée sur l'Annexe B/P.80 et sur l'expérience acquise à la suite de plusieurs études internationales [5], [6], [9], [10], [12], [13], [15] et [17]. Par ailleurs, la présente Recommandation définit une méthode d'essai spécifique pour évaluer les processus numériques. Cette méthode permet de prendre en considération, lors de l'évaluation de la qualité du réseau international en cours d'évolution, les effets subjectifs de la distorsion introduite par le processus de codage des signaux vocaux. La présente Recommandation contient des avis sur la façon d'évaluer la qualité des codecs numériques. Elle doit être lue conjointement avec la Recommandation P.80. La bibliographie donne des renseignements complémentaires sur les sources disponibles.

Diverses méthodes sont possibles pour caractériser la qualité des processus numériques. Actuellement, ces méthodes comprennent des essais d'écoute seulement, qui impliquent:

- 1) des notes d'opinion (par catégorie);
- 2) des comparaisons par paires ou multiples;
- 3) des tests d'articulation.

Pour la plupart des applications, l'UIT-T recommande d'utiliser la méthode d'évaluation par catégories absolues (ACR) (*absolute category rating*) fondée sur l'échelle de qualité d'écoute. Il y a toutefois des cas où d'autres échelles et méthodes d'évaluation sont plus appropriées et mieux adaptées (par exemple, on a constaté que l'échelle d'effort d'écoute était utile lorsque l'on recherche une bonne corrélation avec les mesures de qualité vocale); ces méthodes sont également utilisées dans la présente Recommandation. Seuls seront signalés ci-après les cas où, exceptionnellement, la méthode d'évaluation ACR fondée sur l'échelle de qualité d'écoute n'est pas utilisée.

NOTE – La méthode d'évaluation par catégories de dégradation (DCR) (*degradation category rating*) [8] est décrite en détail dans l'Annexe D/P.80. Cette méthode est supposée convenir lorsque la qualité absolue du matériel source est faible (par exemple paroles mutilées par un bruit de fond) ou lorsque les dégradations numériques sont faibles. Elle peut donc être particulièrement utile pour l'optimisation des systèmes. Il convient en outre de noter qu'une méthode de seuil pour comparaison directe est décrite en détail dans l'Annexe E/P.80 et est également applicable pour l'optimisation des systèmes.

Il peut aussi être nécessaire de compléter les essais d'écoute seulement par des essais de conversation planifiés et conduits conformément à la Recommandation P.80. Tant qu'un complément d'étude n'aura pas été effectué quant aux circonstances dans lesquelles des essais de conversation sont nécessaires et quant au choix des conditions à réaliser au cours de telles expériences, aucun plan détaillé ne pourra toutefois être inclus dans la présente Recommandation.

## 2 Références

Les Recommandations contiennent des dispositions qui, par suite de la référence qui y est faite, constituent des dispositions valables pour la présente Recommandation. Au moment de la publication, les éditions indiquées étaient en vigueur. Toute Recommandation ou autre référence est sujette à révision; tous les utilisateurs de la présente Recommandation sont donc invités à rechercher la possibilité d'appliquer les éditions les plus récentes des Recommandations indiquées ci-après. Une liste des Recommandations UIT-T en vigueur est publiée régulièrement.

- Publication 581-7 de la CEI (1986), *Equipements et systèmes électroacoustiques haute fidélité; Valeurs limites des caractéristiques – Septième partie: Haut-parleurs.*
- Recommandation UIT-T G.113 (1996), *Dégradations de la transmission.*
- Recommandation UIT-T G.191 (1993), *Outils logiciels pour la normalisation du codage des signaux vocaux et audiofréquence.*

<sup>1)</sup> Antérieurement P.83.

- Recommandation UIT-T G.192 (1996), *Interface parallèle numérique commune pour la normalisation des signaux vocaux.*
- Recommandation G.711 du CCITT (1988), *Modulation par impulsions et codage (MIC) des fréquences vocales.*
- Recommandation G.712 du CCITT (1992), *Caractéristiques de qualité de transmission des canaux MIC.*
- Recommandation G.722 du CCITT (1988), *Codage audiofréquence à 7 kHz à un débit inférieur ou égal à 64 kbit/s.*
- Recommandation G.726 du CCITT (1990), *Modulation par impulsions et codage différentiel adaptatif (MICDA) à 40, 32, 24, 16 kbit/s.*
- Recommandation G.728 du CCITT (1992), *Codage de la parole à 16 kbit/s en utilisant la prédiction linéaire à faible délai avec excitation par code.*
- Recommandation UIT-T O.41 (1994), *Psophomètre utilisé sur des circuits de type téléphonique.*
- Recommandation O.131 du CCITT (1988), *Appareil pour la mesure de la distorsion de quantification utilisant un bruit pseudo-aléatoire comme signal d'essai.*
- Recommandation O.132 du CCITT (1988), *Appareil de mesure de la distorsion de quantification utilisant un signal d'essai sinusoïdal.*
- Recommandation UIT-T O.133 (1993), *Appareils destinés à mesurer la qualité de fonctionnement de codeurs et décodeurs de modulation par impulsions et codage.*
- Recommandation UIT-T P.10 (1993), *Vocabulaire des termes relatifs à la qualité de la transmission téléphonique et aux appareils téléphoniques.*
- Recommandation P.48 du CCITT (1988), *Spécification d'un système de référence intermédiaire.*
- Recommandation UIT-T P.50 (1993), *Voix artificielle.*
- Recommandation UIT-T P.56 (1993), *Mesure objective du niveau vocal actif.*
- Recommandation UIT-T P.80 (1993), *Méthodes d'évaluation subjective de la qualité de transmission.*
- Recommandation UIT-T P.810 (1996), *Appareil de référence à bruit modulé (MNRU).*
- Recommandation UIT-T P.84 (1993), *Méthode d'essai d'écoute subjective pour évaluer les équipements de multiplication de circuit numérique et les systèmes téléphoniques avec mise en paquets.*
- Recommandation Q.35 du CCITT (1988), *Caractéristiques techniques des tonalités du service téléphonique.*

### 3 Définitions

Pour les besoins de la présente Recommandation, les définitions suivantes s'appliquent.

**3.1 dBov:** décibels par rapport au point de surcharge d'un système numérique.

**3.2 Q:** rapport, en décibels, entre la puissance des signaux vocaux et la puissance du bruit modulé dans l'appareil de référence à bruit modulé, décrit dans la Recommandation P.810.

**3.3  $Q_N$ :** rapport Q pour un appareil de référence à bruit modulé en bande étroite.

**3.4  $Q_W$ :** rapport Q pour un appareil de référence à bruit modulé en large bande.

### 4 Abréviations

Pour les besoins de la présente Recommandation, les abréviations suivantes s'appliquent.

|     |  |
|-----|--|
| ACR | évaluation par catégories absolues ( <i>absolute category rating</i> )           |
| BER | taux d'erreur sur les bits ( <i>bit error rate</i> )                             |
| DCR | évaluation par catégories de dégradation ( <i>degradation category rating</i> )  |
| eif | facteur de dégradation due à l'équipement ( <i>equipment impairment factor</i> ) |

|           |   |
|-----------|---|
| IRS       | système de référence intermédiaire ( <i>intermediate reference system</i> ) (Recommandation P.48)               |
| ITU-T STL | bibliothèque d'outils logiciels de l'UIT-T ( <i>ITU-T software tools library</i> )                              |
| LDCELP    | prédiction linéaire à faible délai avec excitation par code ( <i>low delay code excited linear prediction</i> ) |
| MIC       | modulation par impulsions et codage   |
| MICDA     | MIC différentielle adaptable  |
| MNRU      | appareil de référence à bruit modulé ( <i>modulated noise reference unit</i> )                                  |
| MOS       | note moyenne d'opinion ( <i>mean opinion score</i> )  |
| qdu       | unité de distorsion de quantification ( <i>quantization distortion unit</i> )                                   |
| RMS       | valeur quadratique moyenne ( <i>root mean square</i> )  |
| RTPC      | réseau téléphonique public commuté  |
| SNR       | rapport signal/bruit ( <i>signal-to-noise ratio</i> )   |

## 5 Conventions

L'évaluation subjective des codecs numériques vocaux peut en principe être conduite au moyen de méthodes d'essais subjectifs ne prévoyant que l'écoute ou avec conversation. Sur un plan pratique, les essais à écoute seulement sont la seule méthode d'essais subjectifs qui soit applicable au cours de la mise au point de codecs vocaux si une mise en œuvre en temps réel du codec n'est pas possible. La présente Recommandation traite des méthodes d'écoute seulement pour l'évaluation subjective des codecs vocaux.

## 6 Résumé des procédures d'évaluation subjective

Lors de l'exécution d'une évaluation subjective d'un codec vocal, il faut tenir compte de l'application prévue du codec soumis à l'essai. Le système utilisé pour enregistrer les matériels vocaux, le choix des conditions expérimentales des essais et le système utilisé pour acheminer les matériels vocaux traités jusqu'aux auditeurs doivent être choisis de façon à reprendre les principales caractéristiques de la chaîne de connexion téléphonique de bout en bout. Il convient en particulier que les caractéristiques des systèmes d'enregistrement et de réception reflètent les caractéristiques de l'émetteur et du récepteur qui seront utilisés dans l'application prévue du codec.

Les évaluations subjectives des codecs numériques sont exécutées selon un certain nombre d'étapes:

- 1) préparation des matériels vocaux de source, y compris l'enregistrement des locuteurs;
- 2) sélection des paramètres expérimentaux afin de mettre à contribution les caractéristiques recherchées du codec;
- 3) conception de l'expérience;
- 4) sélection d'une procédure d'essai et conduite de l'expérience;
- 5) analyse des résultats.

Chacune de ces étapes sera décrite ci-dessous.

## 7 Enregistrements de la source

### 7.1 Système d'enregistrement

Comme indiqué ci-dessus, il est important d'utiliser un système représentatif pour le côté émission du système soumis à l'essai. Pour les essais de codecs numériques à bande étroite (300-3400 Hz) destinés à des applications en réseau, l'UIT-T recommande l'emploi d'une caractéristique d'émission du système de référence intermédiaire (IRS) modifié. On construit un tel dispositif en retirant de l'émetteur IRS le filtre SRAEN (système de référence pour la détermination des affaiblissements équivalents pour la netteté). La caractéristique d'émission du système IRS modifié est décrite dans l'Annexe D. La bibliothèque d'outils logiciels de l'UIT-T (ITU-T STL) propose une mise en œuvre logicielle de la caractéristique d'émission du système IRS modifié.

NOTE – La bibliothèque d'outils logiciels de l'UIT-T contient un certain nombre d'outils qui sont utiles pour la mise au point de normes de codage de signaux de parole et de données audio. Ces outils sont également utiles pour préparer les matériels de base pour les évaluations subjectives des codecs vocaux. La Recommandation G.191 décrit les outils qui sont disponibles dans la bibliothèque ITU-T STL.

Dans certaines circonstances (par exemple lorsque l'on prévoit que les basses fréquences seront fortement limitées), il peut être nécessaire ou souhaitable d'évaluer la qualité du codec en utilisant le côté émission du système de référence intermédiaire (IRS) conforme à la Recommandation P.48 et étalonné selon la Recommandation P.64. On a cependant constaté que la qualité des codecs vocaux à faible débit binaire pouvait dépendre sensiblement de la caractéristique de fréquence appliquée au signal vocal d'entrée. Le système IRS est en particulier représentatif des chaînes de connexion de la téléphonie analogique (années 1970). C'est pourquoi il convient d'utiliser, dans la plupart des circonstances, la caractéristique d'émission du système IRS modifié.

Si l'on se propose d'utiliser un système à large bande (100-7000 Hz) pour audioconférence, il y a lieu que le côté émission soit conforme à la Publication 581-7 de la CEI.

## 7.2 Echantillons vocaux

Il convient que le matériel vocal consiste en phrases simples, courtes, précises et que ces phrases soient choisies de façon à être faciles à comprendre (par exemple extraites de la littérature non technique ou de périodiques actuels). Il y a également lieu que ces phrases soient groupées par deux ou par trois, de manière qu'il n'y ait aucune liaison évidente entre la signification des phrases d'un même groupe. Il convient d'éviter les phrases très courtes ou très longues, le but étant que la durée de prononciation de chaque phrase soit de 2 à 3 s.

NOTE – Sur le plan des erreurs sur les bits, cette méthode permet, par exemple, d'évaluer des taux d'erreur sur les bits (BER) de 1:10000, 1:1000 et 1:100 (voir 8.1.4). Les BER inférieurs à 1:10000 nécessitent un plus grand nombre de phrases par échantillon.

### 7.2.1 Production de matériel de source

Il convient de faire asseoir (un par un) les locuteurs dans une salle présentant une durée de réverbération inférieure à 500 ms et un niveau de bruit de salle inférieur à 30 dBA sans crêtes dominantes dans le spectre. Il y a lieu de faire les enregistrements au moyen d'un système de haute qualité.

Les locuteurs prononceront les phrases avec aisance mais sans effets particuliers. Ils conserveront sans effort un volume constant, tout en évitant les bruits provenant par exemple du froissement de papiers. Les phrases peuvent être présentées au locuteur une par une sur un écran d'affichage piloté par ordinateur, ou sur une liste placée à une distance permettant une lecture aisée.

Il y a lieu d'observer, pendant l'enregistrement, le niveau de conversation active. Au cours du processus d'enregistrement, il convient de veiller à ce que le niveau de conversation active soit, pendant l'enregistrement, compris entre 20 dB et 30 dB au-dessous du point de surcharge du système d'enregistrement, avec une mesure particulière pour chaque phrase. Il conviendra de réenregistrer toute phrase pour laquelle cette condition n'est pas observée.

NOTE – Au cours de l'enregistrement, il faut veiller à ne pas ajouter de bruits parasites (comme un ronflement de réseau d'alimentation ou des pointes de commutation).

### 7.2.2 Egalisation du niveau des signaux vocaux

La mesure des niveaux des signaux dans la partie numérique du réseau international est normalement exprimée en y dBm0 par les ingénieurs en télécommunication, c'est-à-dire par rapport à 1 mW sur 600 Ω. Les matériels vocaux enregistrés numériquement par ordinateur sont cependant stockés sous forme d'entiers, le plus souvent. Il est plus pratique de spécifier les niveaux de ces matériels par rapport à la puissance du signal maximal qui puisse être mise en mémoire dans un ordinateur, par exemple z dBov (où, par convention, l'indice ov désigne le niveau numérique du signal de surcharge). Il est nécessaire de comprendre la relation de ces termes entre eux, de façon à pouvoir exprimer les niveaux en dBov. Une telle relation est de la forme suivante:

$$z \text{ dBov} = y \text{ dBm0} + C \quad (7-1)$$

Pour un système codé en loi A, une onde sinusoïdale qui excite jusqu'au niveau maximal a une puissance donnée par  $T_{\max} = 3,14 \text{ dBm0}$  (Recommandation G.711). De même, pour un système codé en loi  $\mu$ ,  $T_{\max} = 3,17 \text{ dBm0}$ .

La puissance de référence pour le niveau dBov est la plus grande valeur d'entier pour le système utilisé. Une onde sinusoïdale d'amplitude maximale (par exemple 32 767 pour un système à 16 bits) aura un niveau de puissance de  $-3,01 \text{ dBov}$ . On aura donc  $C = -6,15 \text{ dB}$  pour les systèmes en loi A et  $C = -6,18 \text{ dB}$  pour les systèmes en loi  $\mu$ .

Une valeur nominale courante pour le niveau de conversation active moyen (mesuré conformément à la Recommandation P.56) est  $-20$  dBm0, correspondant à environ  $-26$  dBov. Il convient donc de mesurer le niveau de conversation active pour chaque phrase, ainsi que le niveau enregistré après réglage sur  $-27 \pm 1$  dBov pour le système numérique, c'est-à-dire en format numérique. Pour les langues ou les voix dont le niveau crête moyen dépasse 23 dB, il convient de réduire en conséquence le niveau du matériel vocal. Pour déterminer le niveau moyen des signaux vocaux en cours de conversation, il y a lieu d'utiliser le logiciel mis en œuvre selon la Recommandation P.56 et disponible dans la logithèque ITU-T STL.

Finalement, les groupes de phrases sont combinés (avec une période de silence appropriée entre les phrases) afin de produire les données de source finales qui seront utilisées dans l'essai subjectif.

## 8 Choix des paramètres d'expérience

Par commodité, on divise les conditions expérimentales selon qu'elles mettent en jeu le codec soumis à l'essai ou qu'elles servent de références. Les conditions avec codec donnent des renseignements essentiels sur la qualité de fonctionnement du codec, alors que les conditions de référence permettent d'effectuer des comparaisons fines sur le même codec essayé dans différents laboratoires ou dans le même laboratoire à différents moments. Les Figures 1 à 3 et les Figures 8 et 9 sont des exemples montrant des effets typiques sur la note moyenne d'opinion (MOS) concernant quelques paramètres expérimentaux.

### 8.1 Conditions requises pour les codecs

Les conditions d'essai décrites dans ce paragraphe servent à indiquer dans quelle mesure la qualité du codec est affectée par des variations du niveau d'entrée et du niveau d'écoute, par différents locuteurs, par la présence d'un bruit ambiant, etc.

#### 8.1.1 Niveaux d'entrée de parole

On admet que les codecs vocaux numériques sont sensibles au niveau des signaux vocaux d'entrée et que leur qualité de fonctionnement dépend de ce niveau. Lors de l'évaluation de la qualité des codecs, on règle généralement les niveaux d'entrée par rapport au point de surcharge des codecs. Pour évaluer l'influence de variations de niveaux d'entrée sur la qualité des codecs, il est recommandé d'utiliser des niveaux d'entrée de 14, 26 et 38 dB au-dessous du point de surcharge des codecs (ce qui équivaut à  $-8$ ,  $-20$  et  $-32$  dBm0 environ). Ces niveaux correspondent approximativement à la valeur moyenne mesurée au point de commutation international  $\pm$  deux écarts types. D'autres applications peuvent exiger des niveaux d'entrée différents. La Figure 1 illustre l'effet des variations du niveau d'entrée vocal sur la note moyenne d'opinion (MOS).

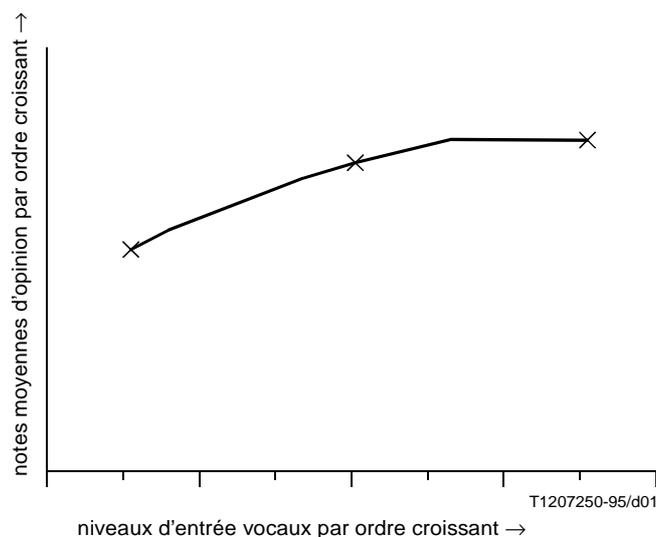


FIGURE 1/P.830

Notes moyennes d'opinion pour les niveaux d'entrée vocaux

### 8.1.2 Niveaux d'écoute

Lorsqu'il s'agit d'évaluer l'effet du niveau de réception, il est recommandé d'utiliser au moins trois niveaux d'écoute. Ces niveaux doivent normalement se situer à 10 dB de part et d'autre du niveau d'écoute préféré (choisi en l'occurrence à 79 dB SPL au plan de référence oreille). Il convient de noter que le niveau d'écoute préféré est généralement inférieur au «niveau d'écoute optimal», c'est-à-dire au niveau d'écoute correspondant à la meilleure note MOS, comme défini dans la Recommandation P.10.

La Figure 2 illustre l'effet, sur la note MOS, des modifications du niveau d'écoute (ainsi que l'interaction entre le niveau d'écoute selon que les locuteurs sont de sexe masculin ou féminin).

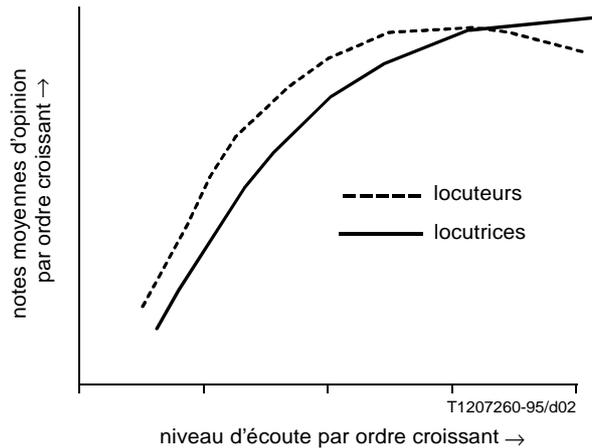


FIGURE 2/P.830

#### Notes moyennes d'opinion pour les niveaux d'écoute

### 8.1.3 Locuteurs

#### 8.1.3.1 Locuteurs différents

Il est recommandé d'utiliser au minimum deux hommes et deux femmes. La Figure 2 illustre l'effet de différents locuteurs sur la note MOS. Si cependant l'on veut tester la variation en fonction des locuteurs en tant que variable indépendante, il est recommandé de faire appel à un plus grand nombre de locuteurs, comme suit:

- 8 hommes;
- 8 femmes;
- 8 enfants.

#### 8.1.3.2 Locuteurs multiples

Il est important de connaître le degré d'insensibilité du codec à la présence simultanée de multiples signaux d'entrée vocaux et de pouvoir expliquer l'apparition d'éventuels phénomènes défavorables tels que les interruptions de transmission, les signaux parasites, etc.

Si l'on ne veut prendre en considération que l'utilisation du combiné, une combinaison de deux locuteurs différents avec des niveaux vocaux d'entrée séparés de 20 dB est généralement suffisante. Si toutefois le codec doit être utilisé en mode conférence, c'est-à-dire «mains-libres», avec plusieurs locuteurs pouvant parler en même temps, il faut s'assurer que la présence de locuteurs multiples n'a pas une influence défavorable sur la qualité du codec lorsque la différence de niveau vocal peut s'annuler.

Pour évaluer les effets de locuteurs multiples, il est recommandé d'utiliser soit la méthode d'évaluation par catégories de dégradation (DCR) fondée sur l'échelle à 5 points du 10.2.3 (voir également l'Annexe D/P.80) soit la méthode de détectabilité de la réponse discontinue alternative fondée sur l'échelle à 3 points du 10.2.4 (voir également l'Annexe C/P.80).

#### 8.1.4 Erreurs

La sélection des conditions d'erreur doit toujours être faite dans le contexte des applications prévues pour le codec. Si celui-ci doit être utilisé dans le RTPC (systèmes de lignes), il y a lieu que les erreurs sur les bits suivent une loi de répartition aléatoire. Si, par ailleurs, l'application prévue est un environnement radioélectrique tel que les radiocommunications mobiles, les erreurs se produiront soit isolément soit en salves et leur effet sera une corruption de trames entières des signaux vocaux codés, plutôt que d'éléments binaires isolés.

Il est recommandé d'utiliser, pour les systèmes de lignes, des erreurs à répartition aléatoire avec des taux d'erreur sur les bits (BER) compris entre 0 et 1:1000. Naturellement, cela dépendra du nombre de phrases utilisées dans un échantillon (voir 7.2). Dans certaines circonstances, il peut être nécessaire de tester avec des BER pouvant atteindre 1:100. On pourra utiliser, à cette fin, le dispositif d'insertion d'erreurs (EID) (*error insertion device*) proposé dans la bibliothèque ITU-T STL.

Pour d'autres applications, telles que les radiocommunications mobiles, les erreurs peuvent être du type à salves et il convient donc d'utiliser des erreurs de ce type. Là encore, le dispositif EID de la bibliothèque ITU-T STL peut être utilisé pour ces essais.

On utilise généralement la méthode d'évaluation ACR fondée sur l'échelle de qualité d'écoute du 10.2.1. Si cependant on s'attend à des conditions médiocres, la méthode d'évaluation ACR fondée sur l'échelle d'effort d'écoute du 10.2.2 peut être plus appropriée.

#### 8.1.5 Débits binaires

Le codec doit être testé à tous les débits binaires auxquels il est capable de fonctionner (ces débits étant par exemple de 48 kbit/s, 56 kbit/s et 64 kbit/s pour la Recommandation G.722 et de 16 kbit/s, 24 kbit/s, 32 kbit/s et 40 kbit/s pour la Recommandation G.726). Le fonctionnement à certains débits binaires peut cependant dépendre des conditions d'exploitation et la charge du système peut entrer en jeu (voir la Recommandation P.84). La Figure 3 illustre l'effet, sur la note MOS, des variations du débit binaire.

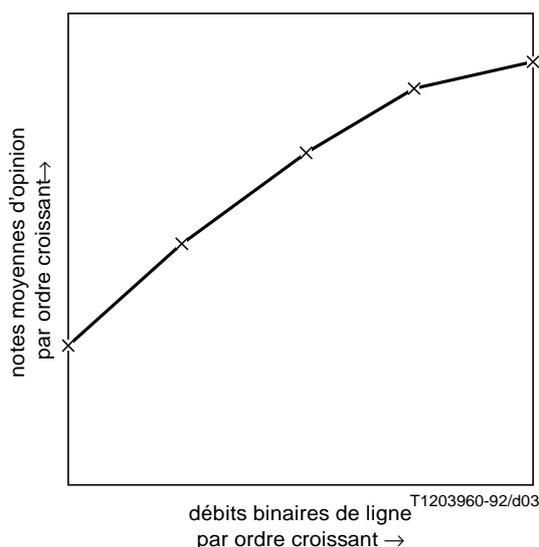


FIGURE 3/P.830

Notes moyennes d'opinion à différents débits binaires

### 8.1.6 Transcodages

Etant donné que les codecs normalisés par l'UIT-T sont susceptibles d'être utilisés dans le RTPC (et dans l'architecture à 64 kbit/s associée), il convient d'évaluer la qualité des codecs vocaux en combinaison avec un processus de compression-extension en loi A ou en loi  $\mu$ . Un seul codec inséré dans une connexion allant d'un émetteur analogique à un récepteur analogique aura donc la configuration indiquée sur la Figure 4.

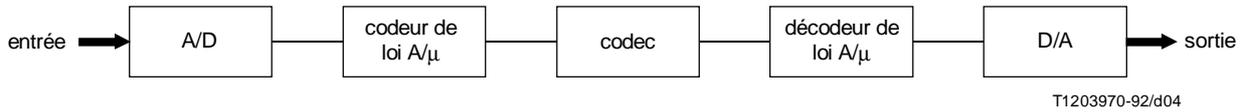


FIGURE 4/P.830  
Système de transcodage

#### 8.1.6.1 Mise en cascade asynchrone

Lorsque le codec en essai doit être utilisé dans le RTPC mixte analogique et numérique, des circonstances peuvent apparaître où les signaux vocaux sont codés et décodés au moins deux fois par le codec, c'est-à-dire que divers segments numériques seront dispersés parmi des segments analogiques. Dans de telles situations, la distorsion introduite par le codec se cumule à chaque conversion de numérique en analogique puis d'analogique en numérique, avec recodage par le codec. De plus, lorsque le codec est déployé dans le réseau, chaque transcodage comporte une distorsion de quantification supplémentaire, produite par le processus linéaire de compression-extension en loi A ou en loi  $\mu$ , ainsi qu'une distorsion d'affaiblissement supplémentaire résultant des tolérances dans les filtres anti-distorsion de repliement et de reconstitution, plus le bruit accumulé dans la voie inactive. Cette configuration est illustrée sur la Figure 5.

L'UIT-T recommande de tester au moins 1, 2 et 3 codecs en cascade. La méthode recommandée pour réaliser une mise en cascade asynchrone (afin de simuler le processus de conversion D/A et A/D) consiste à utiliser le filtre de conversion dans les deux sens (suréchantillonnage et sous-échantillonnage) proposé dans la bibliothèque ITU-T STL.

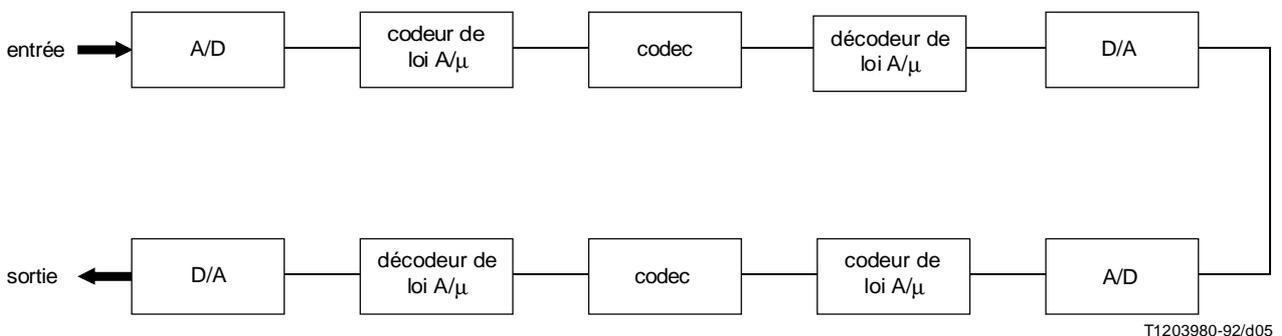


FIGURE 5/P.830  
Mise en cascade asynchrone

### 8.1.6.2 Mise en cascade synchrone

La configuration de circuit représentée sur la Figure 6 illustre des situations similaires à celles du 8.1.6.1. Dans ces situations, les signaux vocaux sont codés et décodés à multiples reprises mais sans conversions D/A ni en loi A ou  $\mu$  (c'est-à-dire dans un réseau tout numérique). Cette situation présente donc l'avantage de ne nécessiter qu'un seul processus de compression-extension de loi A ou  $\mu$  et une distorsion de quantification réduite par rapport à celle du 8.1.6.1. La distorsion due au codec d'essai continue cependant à se cumuler et doit donc être évaluée.

L'UIT-T recommande de tester au moins 1, 2 et 3 codages et décodages successifs par les codecs en cascade.

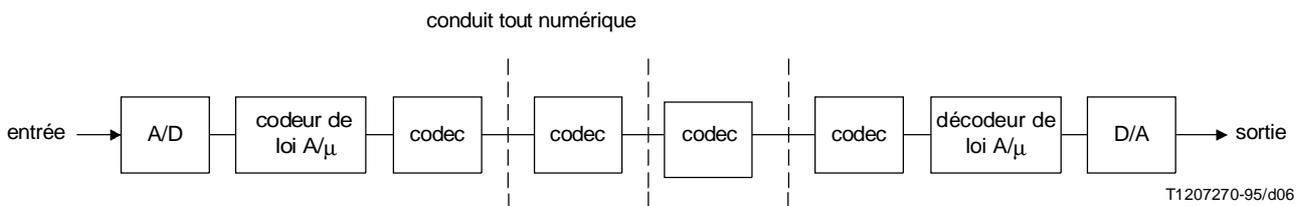


FIGURE 6/P.830

Exemple de mise en cascade synchrone

### 8.1.6.3 Interopérabilité avec d'autres normes de codage de parole

Il est important de déterminer les effets de systèmes en cascade qui utilisent différents débits binaires ou différentes méthodes de codage. Une configuration de ce type est illustrée sur la Figure 7 à titre d'exemple. On trouvera en [14] d'autres exemples de configurations d'interopérabilité. Il est essentiel que les plus probables combinaisons de codecs soient testées.

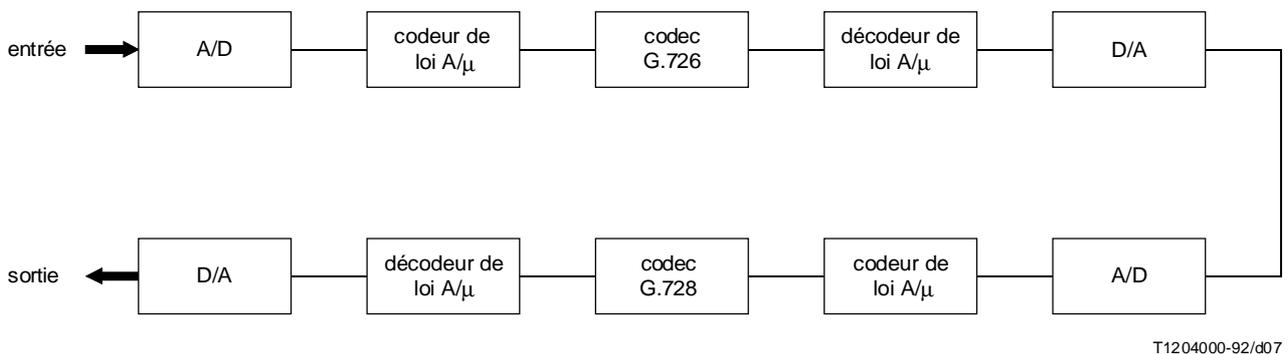


FIGURE 7/P.830

Interopérabilité des normes de codage de parole

L'exemple suivant illustre les combinaisons possibles:

|   |                          |   |                          |                     |
|---|--------------------------|---|--------------------------|---------------------|
| 8-16 kbit/s →   | 64 kbit/s →              | 32 kbit/s →                                     | 64 kbit/s →              | 16-32 kbit/s →      |
| radiocommunications<br>mobiles, annonces<br>enregistrées et applications<br>aéronautiques | circuits<br>interurbains | équipements<br>DCME à débit<br>binaire variable | circuits<br>interurbains | téléphonie sans fil |

Pour les systèmes à bande élargie (G.722 par exemple), il est nécessaire de tester ces systèmes en cascade avec des systèmes à bande normale (étroite), par exemple un système G.722 en cascade avec un système G.726 et vice versa.

### 8.1.7 Défauts d'adaptation

Certains codecs vocaux (conformes à la Recommandation G.722 par exemple) sont capables de fonctionner à divers débits binaires. Dans de tels systèmes, il est possible que l'émetteur et le récepteur travaillent dans des modes différents (c'est-à-dire à des débits binaires différents). Lorsque cette situation est possible, il est nécessaire de calculer la qualité de fonctionnement des combinaisons appropriées en termes de débit binaire, aux deux extrémités de la chaîne de connexion. Par exemple, dans le cas d'un codec répondant aux spécifications de la Recommandation G.722, il y a lieu d'effectuer l'essai dans les conditions suivantes, où codeur et décodeur fonctionnent à des débits différents:

| <i>Débit binaire de<br/>l'émetteur</i> |       | <i>Débit binaire du<br/>récepteur</i> |
|--|-------|---------------------------------------|
| 56 kbit/s                              | ————→ | 64 kbit/s                             |
| 48 kbit/s                              | ————→ | 56 kbit/s                             |
| 48 kbit/s                              | ————→ | 64 kbit/s                             |

### 8.1.8 Bruit d'origine externe (émission)

Comme dans le cas des considérations relatives aux locuteurs multiples (8.1.3.2), l'interaction du bruit d'origine externe avec la voix du locuteur est un facteur important pour les raisons suivantes:

- les codecs à faible débit binaire peuvent utiliser des techniques de codage autres que le codage de forme d'onde;
- l'application d'un codec pourrait consister à l'utiliser dans un environnement bruyant, par exemple un véhicule en mouvement (s'il est utilisé pour les radiocommunications mobiles) ou un bureau bruyant.

Il convient d'effectuer un nombre suffisant d'essais avec le bruit approprié (voir A.1.1.2.2/P.80) afin d'examiner cet effet. Les rapports signal/bruit suivants (voir 8.2.3 pour les définitions) sont recommandés:

- 30 dB pour le bruit de salle;
- 10 dB et 20 dB pour le bruit à l'intérieur d'un véhicule.

Comme la qualité perçue des paroles corrompues par le bruit d'origine externe est censée être un peu inférieure à celle de paroles sur faible bruit de fond, certains laboratoires ont constaté que la méthode d'évaluation ACR, fondée sur l'échelle d'effort d'écoute (voir 10.2.2), était particulièrement utile car elle est fondée sur l'aptitude à comprendre la signification de phrases. Ces laboratoires ont constaté que l'échelle d'effort d'écoute donnait des résultats corrélés plus étroitement avec les notes moyennes d'opinion que les notes obtenues par l'échelle de qualité d'écoute.

Les techniques modernes de codage de la parole permettent parfois de réduire ou d'éliminer le bruit de fond. Il est donc possible qu'un matériel vocal corrompu par du bruit de fond puis traité par un codec soit de meilleure qualité que les signaux d'entrée originaux (c'est-à-dire les paroles corrompues par le bruit). Dans ces situations, l'échelle de comparaison (voir 10.2.5) a été jugée très utile.

Les bruits d'origine externe qui ont été enregistrés avec des microphones de studio doivent être filtrés afin de compenser la différence entre les sources diffuses (en champ lointain) et les sources ponctuelles (en champ proche). Une réalisation logicielle d'un tel filtre est proposée dans la bibliothèque ITU-T STL. Les rapports SNR définis ci-dessus doivent normalement être mesurés après un tel filtrage.

## NOTES

1 Le bruit peut être combiné électriquement avec les enregistrements de la source, de telle sorte que le niveau du bruit et de la parole soit connu de manière précise. Il n'est pas recommandé que les enregistrements de la source soient effectués dans un environnement bruyant.

2 Il ne faut pas supposer automatiquement que la combinaison de signaux vocaux enregistrés séparément et d'un bruit externe élevé produira les mêmes effets qu'un sujet parlant dans un environnement bruyant. Cela tient au fait que les locuteurs adaptent leurs caractéristiques vocales (niveau et spectre) ainsi que leur comportement en matière de parole dans un environnement bruyant. Etant donné que certains codecs peuvent effectuer un traitement différent selon les parties du spectre audiofréquence considérées, cette modification des caractéristiques vocales peut influencer les résultats.

3 Certains laboratoires utilisent une technique d'enregistrement des échantillons vocaux qui tire parti du fait que les locuteurs adaptent leurs niveaux d'élocution en environnement bruyant; mais cette technique évite les problèmes associés aux enregistrements en environnement bruyant. Les locuteurs portent des casques qui reproduisent le bruit de fond à un niveau connu. Les paroles du locuteur sont ensuite enregistrées de façon normale, l'orateur s'exprimant comme s'il se trouvait en environnement bruyant. Le bruit est ensuite ajouté électroniquement aux paroles enregistrées, au niveau souhaité par rapport à celles-ci.

### 8.1.9 Signaux d'information de réseau

Dans tout réseau national, un grand nombre de signaux ou de tonalités d'information sont adressés à l'utilisateur, d'autres constituant des instructions pour les équipements. Ces signaux peuvent provenir de l'intérieur du RTPC ou de réseaux privés qui lui sont rattachés. Il est important que la dégradation de ces signaux, lorsqu'ils passent par un dispositif quelconque de traitement de la parole, ne les rende méconnaissables ni pour le client ni pour l'équipement conçu pour y donner suite à la réception. Il est possible que cette dernière situation tolère moins bien une dégradation que la première lorsque certains signaux sont présents, comme les tonalités DTMF, mais elle offre l'avantage que de simples essais objectifs devraient être suffisants pour détecter une limite admissible de cette dégradation.

Il est recommandé de tester subjectivement les tonalités de signalisation conformes à la Recommandation Q.35 émises par le réseau et au minimum les tonalités suivantes:

- tonalité d'invitation à numéroter;
- tonalité de retour d'appel de l'abonné demandé;
- tonalité d'occupation de l'abonné demandé;
- tonalité d'occupation de l'équipement;
- tonalité de numéro impossible à obtenir.

Il est recommandé d'utiliser la méthode d'évaluation par catégories de dégradation (DCR) fondée sur l'échelle à 5 points du 10.2.3 (voir également l'Annexe D/P.80) pour déterminer si les signaux d'information sont appropriés (reconnaissables).

### 8.1.10 Musique

Certains équipements utilisés dans le RTPC, notamment les commutateurs privés, ont la possibilité de transmettre de la musique au client en cas de «mise en attente». L'UIT-T recommande de n'utiliser que des essais simples pour s'assurer que la musique est d'une qualité raisonnable.

## 8.2 Conditions de référence

Les conditions de référence constituent un moyen pratique pour effectuer des comparaisons utiles entre résultats d'essai subjectif issus de laboratoires différents ou du même laboratoire à des moments différents. Ces conditions comportent une *meilleure condition possible*, ainsi que des conditions où des dégradations connues et contrôlées ont été ajoutées aux données vocales.

### 8.2.1 Condition directe

Il s'agit de la meilleure condition que l'on puisse obtenir lors de l'expérience; elle correspond théoriquement à un rapport  $Q_N$  ou  $Q_W$  infini (voir 8.2.2). Cette condition n'implique en particulier aucun codage des paroles d'entrée, qui ne subissent que le même filtrage d'entrée, le même niveau d'entrée et le même filtrage de sortie que les paroles qui ont été traitées par le codec en essai.

### 8.2.2 Appareil de référence à bruit modulé (MNRU)

Il convient d'utiliser comme système de référence un bruit aléatoire avec une amplitude proportionnelle à l'amplitude du signal instantanée en termes de rapport  $Q_N$  ou  $Q_W$ , conformément à l'appareil de référence à bruit modulé (MNRU) (*modulated noise reference unit*) spécifié dans la Recommandation P.810; la qualité subjective des processus numériques devra être exprimée en fonction de ce système de référence pour les raisons suivantes:

- a) pour les codecs harmoniques éventuellement certains codecs autres qu'harmoniques, le signal traité par l'appareil MNRU est, du point de vue perceptif, sensiblement identique aux signaux traités par ces codecs, ce qui facilite, en principe, l'évaluation par les sujets participant aux essais;
- b) l'expérience a montré que l'appareil MNRU est un étalon de transfert utile qui permet d'effectuer des comparaisons significatives entre différents laboratoires et entre expériences du même laboratoire mais conduites à des moments différents;
- c) une expérience et des informations considérables ont été accumulées avec l'appareil MNRU.

Il y a lieu d'inclure dans l'essai subjectif une gamme de valeurs pour le rapport  $Q_N$  ou  $Q_W$ . Cette gamme de bruit pour les systèmes à bande étroite doit aller de  $Q_N = 5$  à 35 dB (de préférence 5 à 7 valeurs différentes) et pour les systèmes à large bande de  $Q_W = 10$  à 45 dB (soit 5 à 7 valeurs différentes).

NOTE – La distorsion introduite par l'appareil MNRU est similaire, subjectivement à la distorsion de quantification. De telles distorsions ne sont pas forcément similaires, lors d'une évaluation subjective, à la distorsion produite par les techniques modernes de codage de la parole. Les dégradations de référence qui produisent des distorsions plus proches des dégradations dues aux techniques modernes de codage sont à l'étude.

La Figure 8 illustre l'effet des variations de  $Q_N$  ou  $Q_W$  sur les notes MOS.

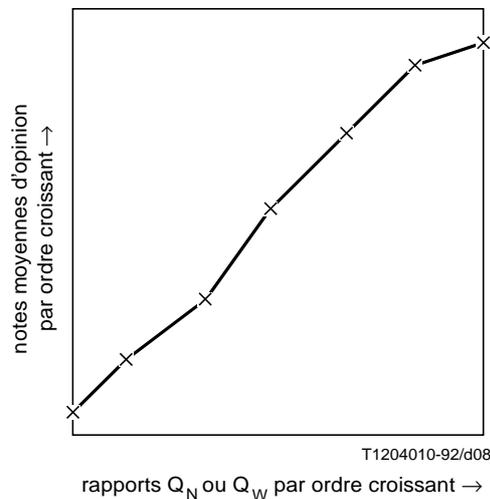


FIGURE 8/P.830  
Notes moyennes d'opinion pour le rapport  $Q_N$  ou  $Q_W$

### 8.2.3 Rapport signal/bruit (SNR)

Les Administrations et les compagnies d'exploitation ont jugé utile d'évaluer les effets des dégradations en termes de rapport signal/bruit.

L'UIT-T recommande d'utiliser les définitions suivantes du rapport signal/bruit, mesuré sur les connexions, en ce qui concerne le bruit en régime permanent:

- *mesures téléphoniques (bande étroite) – valeurs psophométriques* (voir la Recommandation O.41)  
SNR(p) = niveau de parole active (voir la Recommandation P.56)/mesure de bruit psophométriquement pondérée;
- *mesures à large bande – valeurs pondérées A* (voir la Recommandation P.54)  
SNR(A) = niveau de parole active (voir la Recommandation P.56)/mesure de bruit pondérée A.

Si des définitions autres que celles qui sont indiquées ci-dessus sont utilisées, il convient d'adopter le système de notation suivant:

- *bande étroite 300-3400 Hz – valeurs non pondérées*  
SNR(N) = niveau de parole active (voir la Recommandation P.56)/mesure de bruit non pondérée;
- *large bande 100-7000 Hz – valeurs non pondérées*  
SNR(W) = niveau de parole active (voir la Recommandation P.56)/mesure de bruit non pondérée;
- *pondération pour message C* (voir la Recommandation O.41)  
SNR(C) = niveau de parole active (voir la Recommandation P.56)/mesure de bruit pondérée pour message C;
- *largeur de bande 100-5000 Hz*  
SNR(0,1-5 kHz) = niveau de parole active (voir la Recommandation P.56)/mesure de bruit non pondérée.

NOTE – Si d'autres largeurs de bande sont utilisées, il convient de remplacer «0,1-5 kHz» par la largeur de bande de mesure.

On trouvera en Annexe A une comparaison des rapports SNR établie à l'aide des définitions ci-dessus pour différents types de spectres de bruit et différentes caractéristiques d'efficacité en fonction de la fréquence à l'émission.

Si on utilise un bruit gaussien lors de l'essai, il est suggéré que, pour les systèmes à bande étroite et à large bande, la gamme de bruit soit comprise entre 15 dB et 45 dB.

La Figure 9 illustre l'effet typique des variations du rapport SNR sur les notes MOS.

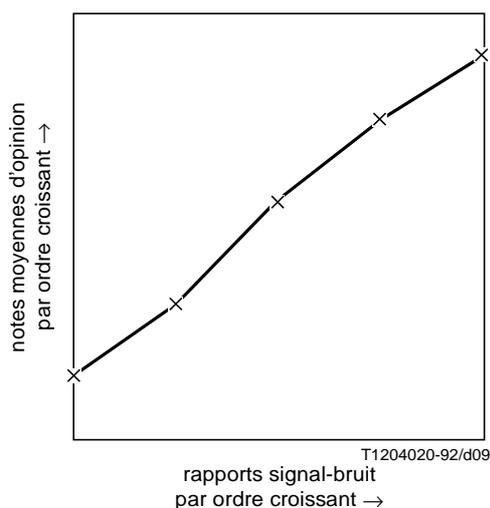


FIGURE 9/P.830

Notes moyennes d'opinion pour le rapport SNR

## 8.2.4 Codecs de référence

Les codecs de référence servent à deux usages:

- 1) on peut les utiliser pour déterminer les règles de planification soit sous la forme d'unités de distorsion de quantification (qdu) soit sous la forme de facteurs de dégradation d'équipement (eif) (voir à ce sujet la Recommandation G.113); et
- 2) on peut les utiliser comme norme pour évaluer la qualité globale en termes de paramètres (voir 8.1).

Si les codecs de référence servent à déterminer les règles de planification, l'UIT-T recommande d'utiliser la mise en cascade asynchrone (voir 8.1.6.1) de 1, 2, 4, 8 et 16 codecs de loi A ou  $\mu$  conformes aux Recommandations G.711 et G.712. La spécification relative à un codec de référence G.711 se trouve dans l'Annexe C.

S'ils servent à déterminer la qualité relative, par exemple par rapport aux Recommandations G.726 ou G.722, les considérations formulées aux 8.1.6.1 et 8.1.6.2 s'appliquent.

## 9 Conception des essais

Les considérations formulées en détail dans B.3/P.80 s'appliquent.

L'utilisation de chaque combinaison des paramètres décrits à l'article 7 conduirait à effectuer une seule expérience qui serait logistiquement trop importante. Il est recommandé de mener un ensemble minimal d'expériences qui, même s'il ne couvre pas chaque combinaison, fournira des données suffisantes pour prendre des décisions en connaissance de cause. L'Annexe B indique des ensembles d'expériences jugés appropriés par l'UIT-T à la suite des études de codecs à bande étroite et à large bande qui ont conduit à l'élaboration des Recommandations de la série G.700.

Il convient d'être extrêmement prudent lorsqu'on compare, même lors d'un même essai, des systèmes présentant des dégradations très différentes (par exemple, codecs numériques, systèmes de multiplexage par répartition en fréquence, vocodeurs, etc.).

## 10 Procédure des essais d'écoute

Cet article est identique à celui de B.4/P.80, avec les exceptions suivantes.

### 10.1 Système de réception

Lors des essais de codecs numériques à bande étroite (300-3400 Hz) pour applications en réseau, l'UIT-T recommande d'utiliser une caractéristique de réception par le système IRS modifié. Comme dans le cas de la caractéristique d'émission du système IRS modifié, ce dispositif est construit par suppression du filtre de système SRAEN dans le récepteur du système IRS. La caractéristique de réception du système IRS modifié est décrite dans l'Annexe D. La bibliothèque ITU-T STL contient une réalisation logicielle de la caractéristique de réception du système IRS. Dans certaines circonstances (par exemple lorsque l'on prévoit que les basses fréquences seront fortement limitées), il sera nécessaire ou souhaitable d'évaluer la performance du codec au moyen du côté réception du système de référence intermédiaire (IRS) conforme à la Recommandation P.48 et étalonné selon la Recommandation P.64.

Si l'on se propose d'utiliser un système à large bande (100-7000 Hz) pour audioconférence, le côté réception doit être conforme à la Publication 581.7 de la CEI.

NOTE – Certains laboratoires utilisent des casques pour la présentation de matériels vocaux à des auditeurs lors d'évaluations subjectives: il faut alors veiller à ce que cette présentation assure un filtrage approprié, afin de simuler les effets de récepteurs réels, y compris les effets, le cas échéant, d'une fuite acoustique. On trouvera en [11] de plus amples renseignements sur l'utilisation de casques au lieu de combinés téléphoniques.

### 10.2 Echelles d'appréciation subjective

Il est recommandé d'utiliser les échelles d'appréciation subjective suivantes pour évaluer les processus numériques.

#### 10.2.1 Echelle de qualité d'écoute

Qualité de la parole:

- 5 Excellent.
- 4 Bonne.

- 3 Assez bonne.
- 2 Médiocre.
- 1 Mauvaise.

### 10.2.2 Echelle d'effort d'écoute

Effort nécessaire pour comprendre la signification des phrases:

- 5 Détente absolue; aucun effort.
- 4 Attention nécessaire; pas d'effort appréciable.
- 3 Effort modéré.
- 2 Effort considérable.
- 1 Signification incompréhensible en dépit de tous les efforts possibles.

### 10.2.3 Echelle de catégorie de dégradation

- 5 Dégradation inaudible.
- 4 Dégradation audible mais pas gênante.
- 3 Dégradation un peu gênante.
- 2 Dégradation gênante.
- 1 Dégradation très gênante.

### 10.2.4 Echelle d'appréciation subjective de détectabilité

- 3 Gênant
- 2 Décelable
- 1 Indécelable

### 10.2.5 Echelle de comparaison

Par rapport au premier échantillon, le second est:

- 3 Bien meilleur
- 2 Meilleur
- 1 Légèrement meilleur
- 0 A peu près équivalent
- 1 Un peu moins bon
- 2 Moins bon
- 3 Beaucoup moins bon

## 10.3 Bruit électrique

Il convient d'ajouter un bruit gaussien équivalant à  $-68$  dBmp à l'entrée du système de réception pour réduire les effets de contraste de bruit au moment où les premières phrases sont prononcées.

## 11 Analyse des résultats

Les considérations formulées en détail dans B.4.7/P.80 s'appliquent.

L'un des objectifs de l'analyse est de déterminer une fonction  $Q_2 = F(L)$  où  $Q_2$  est la valeur équivalente  $Q$  pour le codec et  $L$  le débit binaire de ligne. Une méthode simple utilise, pour déterminer cette fonction, les valeurs de note MOS (note moyenne d'opinion) indiquées sur les Figures 3 et 8 et permet d'établir un graphique de cette fonction, comme indiqué sur la Figure 10. Cette méthode est décrite sur la Figure 11, où on choisit une valeur de débit binaire de ligne, par exemple  $L_2$ , et où on détermine sa note MOS correspondante. On transfère ensuite cette note MOS dans le graphique de droite pour trouver la valeur de  $Q$ , en l'occurrence  $Q_2$ , correspondant à cette note MOS. Les valeurs de  $Q$  pour toutes les autres valeurs  $L$  sont obtenues d'une manière similaire et l'ensemble de gains  $(L_i, Q_i)$  qui en résulte est tracé comme indiqué sur la Figure 10.

NOTE – Les opérations ci-dessus sont plus fiables lorsqu'elles sont exécutées sur des courbes ajustées que lorsqu'elles le sont sur des segments de droite joignant les points observés. Les méthodes permettant de calculer les ajustements appropriés pour les notes MOS en fonction de Q sont à l'étude. A titre de procédure provisoire en attendant les résultats de cette étude, l'attention est appelée sur une méthode d'ajustement sigmoïde qui tient compte de «l'effet de saturation» aux valeurs élevées de Q. Cette méthode a été utilisée avec un certain succès lors des travaux de l'ETSI sur l'évaluation de candidats pour le système GSM à mi-débit [2].

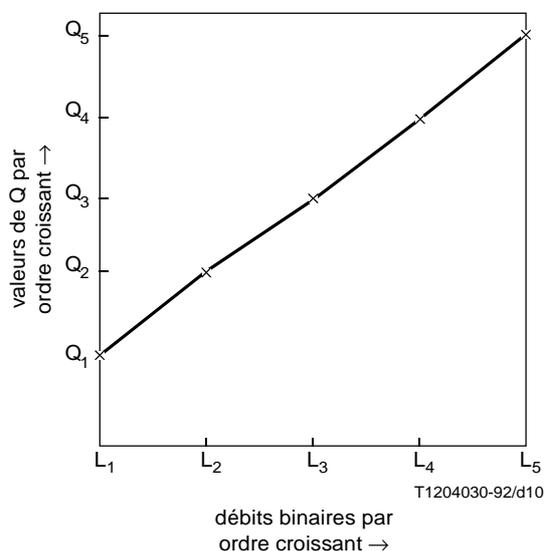


FIGURE 10/P.830  
Valeur de Q en fonction du débit binaire

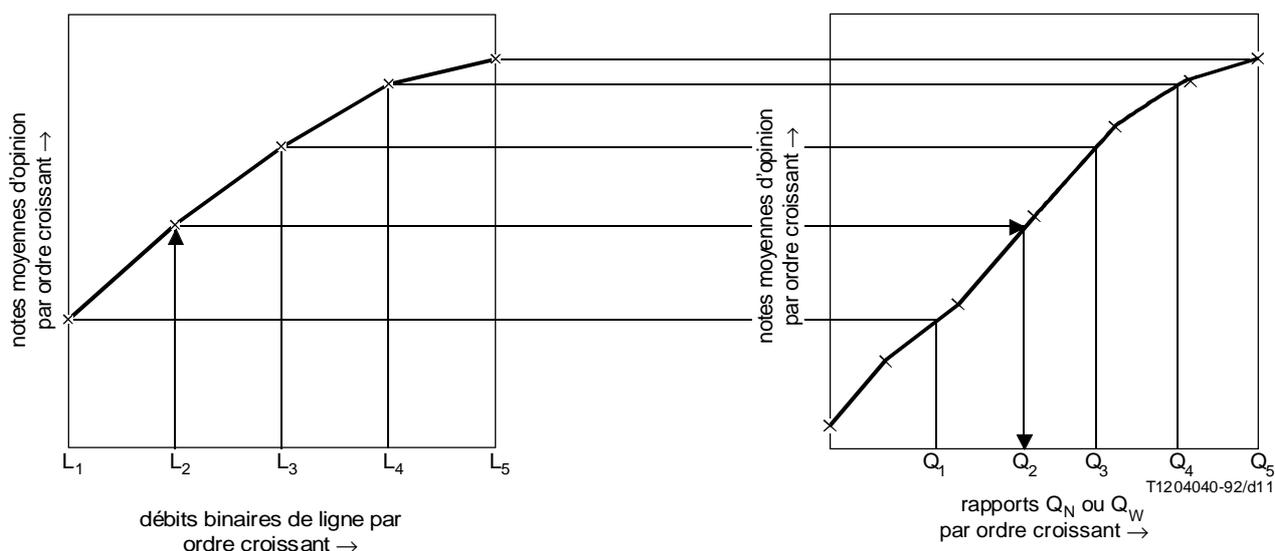


FIGURE 11/P.830  
Méthode graphique permettant d'obtenir la Figure 10  
à partir des Figures 3 et 8

## Annexe A

### Comparaison des différentes définitions du rapport SNR

(Cette annexe fait partie intégrante de la présente Recommandation)

Le Tableau A.1 indique quelques calculs utilisant les pondérations décrites dans la Recommandation O.41 et s'appliquant à des spectres connus avec deux réponses en fréquence différentes. Ces spectres ont été choisis de manière à représenter ceux qui sont le plus généralement utilisés lors d'essais subjectifs.

- Bruit blanc (gaussien) compatible avec le bruit de circuit – psophométrique et pour message C, conçu pour mesurer les effets de ce type de bruit.
- Bruit de Hoth (de salle) et bruit à l'intérieur d'un véhicule utilisés comme bruits d'origine externe lors d'essais subjectifs (voir A.1.1.2.2/P.80) et représentant le ou les bruit(s), capté(s) par le microphone et mesuré(s) aux bornes de la ligne téléphonique.

TABLEAU A.1/P.830

|                                     |                | bruit blanc | bruit de Hoth | véhicule en mouvement | véhicule à l'arrêt |
|-------------------------------------|----------------|-------------|---------------|-----------------------|--------------------|
| réponse uniforme<br>(bande étroite) | non pondérée   | 0           | 0             | 0                     | 0                  |
|                                     | message C      | -1,9        | -3,6          | -6,4                  | -4,2               |
|                                     | psophométrique | -2,6        | -2,6          | -4,1                  | -2,8               |
| réponse IRS<br>(bande étroite)      | non pondérée   | 0           | 0             | 0                     | 0                  |
|                                     | message C      | -1,9        | -2,0          | -3,7                  | -2,3               |
|                                     | psophométrique | -3,6        | -2,3          | -2,4                  | -2,3               |
| réponse uniforme<br>(large bande)   | non pondérée   | 0           | 0             | 0                     | 0                  |
|                                     | message C      | -5,2        | -7,3          | -16,5                 | -12,7              |
|                                     | psophométrique | -5,9        | -6,3          | -14,0                 | -11,3              |
| réponse IRS<br>(large bande)        | non pondérée   | 0           | 0             | 0                     | 0                  |
|                                     | message C      | -1,9        | -2,1          | -4,4                  | -2,5               |
|                                     | psophométrique | -3,6        | -2,4          | -3,1                  | -2,5               |

#### NOTES

- 1 Toutes les valeurs sont des différences en dB par rapport à la valeur non pondérée.
- 2 Le signe négatif signifie «moins bruyant».
- 3 La réponse IRS utilisée ici se rapporte à la Recommandation P.48.
- 4 Pour calculer la valeur non pondérée à bande étroite, on a utilisé les bandes d'1/3 d'octave, centrées sur les fréquences préférées comme indiqué dans l'ISO R.266, c'est-à-dire de 315 Hz à 3150 Hz inclusivement.
- 5 Pour calculer la valeur non pondérée à large bande, on a utilisé les bandes d'1/3 d'octave, centrées sur les fréquences préférées comme indiqué dans l'ISO R.266, c'est-à-dire de 100 Hz à 6300 Hz inclusivement.
- 6 Pour calculer les pondérations message C et psophométriques, on a utilisé les bandes d'1/3 d'octave, centrées sur les fréquences préférées comme indiqué dans l'ISO R.266, c'est-à-dire de 100 Hz à 5000 Hz inclusivement.
- 7 Il convient d'être prudent lorsqu'on suppose que la pondération pour message C équivaut à une pondération psophométrique, ce qui n'est vrai que pour le bruit de type gaussien (voir la Recommandation O.41).

## Annexe B

### Ensemble de paramètres permettant de déterminer la qualité de fonctionnement des codecs

(Cette annexe fait partie intégrante de la présente Recommandation)

La présente annexe donne des exemples d'expériences qu'il convient de conduire pour évaluer la qualité subjective de codecs numériques. Comme indiqué à l'article 6, il importe de compléter ces expériences par des conditions et/ou essais qui soumettent le codec aux conditions spéciales de son application prévue.

## B.1 Systèmes à bande étroite (300-3400 Hz)

- a) *Expérience 1 – Effet des erreurs, du niveau d'entrée et du niveau d'écoute*
- |  |  |
|--|--|
| Niveaux d'entrée vocaux:                       | 3 (14, 26 et 38 dB au-dessous de la surcharge) |
| Niveaux d'écoute:                              | 3 (P + 10, P, P – 10)                          |
| Taux d'erreur:                                 | 3 (0, 1:10000, 1:1000)                         |
| Transcodage:                                   | 1 transcodage                                  |
| Niveaux de bruit d'origine externe (émission): | 1 (< 30 dBA)                                   |
- b) *Expérience 2 – Effet des transcodages, du niveau d'entrée et du niveau d'écoute*
- |  |  |
|--|--|
| Niveaux d'entrée vocaux:                       | 3 (14, 26 et 38 dB au-dessous de la surcharge) |
| Niveaux d'écoute:                              | 3 (P + 10, P, P – 10)                          |
| Transcodage:                                   | «x»  |
| Taux d'erreur:                                 | 1 (1:1000)                                     |
| Niveaux de bruit d'origine externe (émission): | 1 (< 30 dBA)                                   |
- c) *Expérience 3 – Effet du bruit d'origine externe, du bruit de salle, du niveau d'entrée et du niveau d'écoute*
- |  |  |
|--|--|
| Niveaux d'entrée vocaux:                       | 3 (14, 26 et 38 dB au-dessous de la surcharge) |
| Niveaux d'écoute:                              | 3 (P + 10, P, P – 10)                          |
| Transcodage:                                   | 1 transcodage                                  |
| Taux d'erreur:                                 | 1 (1:1000)                                     |
| Niveaux de bruit d'origine externe (émission): | 2 (< 30 dBA et «y»)                            |
| Niveaux de bruit de salle:                     | «z»  |
- où: «P» est le niveau d'écoute préféré;  
«x» est le nombre de combinaisons de transcodages à tester;  
«y» est le niveau de bruit (émission) à tester;  
«z» est le nombre de conditions de bruit de salle (généralement 2).

Les expériences doivent toutes inclure également les conditions 5 à 7 de l'appareil MNRU à bande étroite.

Ces trois expériences ne sont pas nécessairement exhaustives et il y a lieu de les compléter par d'autres expériences pour mieux caractériser le codec.

## B.2 Systèmes à large bande (100-7000 Hz)

- a) *Expérience 1 – Effet du débit binaire, du BER, du niveau d'entrée et du niveau d'écoute*
- |  |  |
|--|--|
| Niveaux d'entrée vocaux:                       | 2 (20 et 38 dB au-dessous de la surcharge)       |
| Niveaux d'écoute:                              | 3 (P + 10, P, P – 10)                            |
| Transcodage:                                   | 1 transcodage                                    |
| Niveaux de bruit d'origine externe (émission): | 1 (< 30 dBA)                                     |
| Débits binaires:                               | «r» (par exemple 48, 56 et 64 kbit/s pour G.722) |
| BER:   | 3 (0, 1:10000 et 1:1000)                         |
| Bruit de salle:                                | 1 (< 30 dBA)                                     |
- b) *Expérience 2 – Effet des transcodages, du niveau d'entrée et du niveau d'écoute*
- |  |  |
|--|--|
| Niveaux d'entrée vocaux:                       | 2 (20 et 38 dB au-dessous de la surcharge)       |
| Niveaux d'écoute:                              | 3 (P + 10, P, P – 10)                            |
| Transcodages:                                  | «x», y compris synchrones et asynchrones         |
| Niveaux de bruit d'origine externe (émission): | 1 (< 30 dBA)                                     |
| Débits binaires:                               | «r» (par exemple 48, 56 et 64 kbit/s pour G.722) |

|   |  |
|---|--|
| BER:  | 3 (0, 1:10000 et 1:1000)   |
| Bruit de salle:   | 1 (< 30 dBA)   |
| c) <i>Expérience 3 – Effet des défauts d'adaptation, du niveau d'entrée et du niveau d'écoute</i> |  |
| Niveaux d'entrée vocaux:  | 2 (20 et 38 dB au-dessous de la surcharge)   |
| Niveaux d'écoute:   | 3 (P + 10, P, P – 10)  |
| Transcodage:  | 1 transcodage  |
| Niveaux de bruit d'origine externe (émission):  | 1 (< 30 dBA)   |
| Débits binaires:  | «m» (par exemple 56 → 64, 48 → 56, 48 → 64 kbit/s)   |
| BER:  | 2 (0 et 1:1000)  |
| Bruit de salle:   | 1 (< 30 dBA)   |
| où: «P»   | est le niveau d'écoute préféré;  |
| «x»   | est le nombre de combinaisons de transcodages à tester;  |
| «r»   | est le nombre de débits binaires auxquels le codec fonctionne;                                     |
| «m»   | est le nombre de combinaisons entre débits binaires différents auxquels le codec doit fonctionner. |

Les expériences doivent toutes inclure également les conditions de l'appareil MNRU à large bande.

Il est rappelé que ces trois expériences ne sont pas destinées à être exhaustives et qu'il y a lieu de les compléter par d'autres expériences pour mieux caractériser le codec.

## Annexe C

### Codec MIC correspondant à une seule unité qdu

(Cette annexe fait partie intégrante de la présente Recommandation)

#### C.1 Interface analogique commune

L'interface analogique commune (Recommandation G.192) est essentielle pour assurer des conditions d'entrée et de sortie identiques lors du test de sélection d'algorithmes pour le traitement des signaux vocaux et de leur codage. Il est également envisagé d'estimer à exactement une unité qdu près la qualité du système cible inséré dans le réseau.

#### C.2 Interface numérique

L'interface numérique connecte les équipements de traitement ou de codage de la parole, soumis aux essais de sélection, à l'interface analogique commune. Celle-ci est conforme à la configuration indiquée dans la nouvelle Recommandation concernant les outils matériels. Cette interface a une fréquence d'échantillonnage de 16 kHz et un échelon de quantification de 16 éléments binaires au plus.

#### C.3 Interface analogique

L'interface analogique se compose de convertisseurs A/D et D/A, d'un filtre qui produit la distorsion d'affaiblissement de l'équipement terminal MIC, du codage et du décodage log-MIC G.711, ainsi que de l'appareil MNRU de la Recommandation P.810 afin d'obtenir la distorsion d'une unité qdu exactement.

##### C.3.1 Convertisseurs A/D et D/A

Le convertisseur A/D transforme un signal analogique en données à codage MIC uniforme avec une précision minimale de 13 bits et une précision maximale de 16 bits à la fréquence d'échantillonnage de 16 kHz. L'entrée dans le convertisseur A/D doit toujours comprendre un filtrage anti-repliement et la sortie du convertisseur D/A doit être injectée dans un filtre de restitution. Ce filtrage présente une réponse en fréquence uniforme à l'intérieur de la bande passante et un affaiblissement suffisant dans la bande affaiblie. La distorsion prévue pour cet étage n'est pas supérieure à la valeur théorique de la distorsion de quantification en codage MIC uniforme. Il convient que le convertisseur D/A soit équilibré par rapport au convertisseur A/D.

### C.3.2 Distorsion d'affaiblissement (en fonction de la fréquence)

Il convient que la distorsion d'affaiblissement (en fonction de la fréquence) des côtés émission ou réception de l'interface soit conforme au gabarit de la Figure 3/P.66. Afin toutefois de limiter toute distorsion d'affaiblissement dans la bande passante qui serait due à une mise en cascade d'équipements, il y a lieu que l'amplitude des ondulations dans la bande passante soit inférieure à 0,1 dB.

### C.3.3 Appareil MNRU

La distorsion totale, correspondant à une seule unité qdu, est produite par un codec MIC ayant un rapport signal sur distorsion (S/D) de 35 dB lorsque le signal d'entrée est compris entre 0 dBm0 et -30 dBm0, mesuré par la méthode indiquée dans la Recommandation O.132. Le rapport S/D pour l'interface analogique commune est choisi à 2 dB au-dessus du gabarit de la Recommandation G.712; on estime qu'il est équivalent à un codec représentatif du marché (voir la Recommandation G.712).

Etant donné qu'un tel codec MIC type est un dispositif théorique, il est difficile de concevoir un codec réel présentant le rapport S/D exact. Si l'on suppose que les convertisseurs A/D et D/A sont idéaux et que l'on réalise le codage log-MIC par consultation de table, ce codec a une qualité de fonctionnement nettement meilleure à celle d'un équipement MIC représentatif du marché. Pour réaliser une émulation fidèle d'un codec MIC type, il convient d'insérer un appareil MNRU (conforme à la Recommandation P.810) avec  $Q = 36$  dB, entre le convertisseur A/D et le codeur log-MIC, ainsi qu'entre le décodeur log-MIC et le convertisseur D/A, comme représenté sur la Figure C.1.

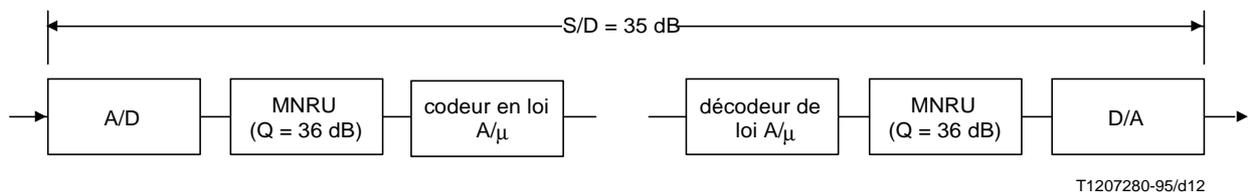


FIGURE C.1/P.830

#### Configuration pour une unité qdu

Il convient que le système soit étalonné conformément à la procédure suivante, qui suppose qu'une tolérance de  $\pm 0,1$  dB est acceptable pour le rapport S/D global:

- 1) régler les deux appareils MNRU sur  $Q = 36$  dB;
- 2) mesurer le rapport S/D du système complet par la méthode de la Recommandation O.132, le résultat étant dénoté par le symbole  $Q_0$ ;
- 3) si  $34,9 < Q_0 < 35,1$  l'étalonnage est effectué; si ce n'est pas le cas:
- 4) ajouter ( $Q_0 - 35$  dB à la valeur réglée de chaque appareil MNRU et revenir à l'étape 2).

## Annexe D

### Caractéristiques d'émission et de réception de l'appareil IRS modifié

(Cette annexe fait partie intégrante de la présente Recommandation)

L'origine des caractéristiques de fréquence en émission et en réception dans le système de référence intermédiaire (IRS) provient d'une longue série de mesures relevées au début des années 1970 sur un certain nombre de postes téléphoniques analogiques de l'époque [3] et [4]. Sur la base de ces mesures, on a calculé des caractéristiques moyennes de réponse en fréquence à l'émission et à la réception. Pour atteindre l'objectif d'équilibrage des sonies, pour lequel le système IRS est conçu, il a toutefois été nécessaire d'insérer également un filtre passe-bande de 300 à 3400 Hz, appelé *filtre SRAEN*. Les valeurs de sensibilité à l'émission et à la réception actuellement indiquées dans la Recommandation P.48 sont donc composées des réponses moyennes à l'émission et à la réception pour un certain nombre de postes téléphoniques, ainsi que de la réponse du filtre SRAEN.

Bien que ces réponses aient pu être représentatives des chaînes de connexion par ressources analogiques de transmission à grande distance, ces caractéristiques ne sont pas représentatives des connexions établies sur des ressources numériques modernes. La gigue à basse fréquence donne en particulier naissance à une dégradation de qualité inutile. Dans le cadre de l'évaluation des codecs à faible débit binaire, surtout lorsque le codec est situé dans le combiné téléphonique, on peut obtenir une meilleure caractéristique en modifiant la réponse du système IRS par mise en court-circuit du filtre SRAEN, comme indiqué dans le Tableau D.1. A titre d'information, la réponse du filtre SRAEN est également incluse dans le Tableau D.1. La tolérance sur les points nominaux indiqués dans le Tableau D.1 est de  $\pm 2,5$  dB aux fréquences comprises entre 200 Hz et 3400 Hz pour les caractéristiques d'émission et de réception du système IRS modifié. Au-dessous de 200 Hz, il convient que la décroissance de la caractéristique d'IRS modifié soit d'au moins 15 dB par octave. Au-dessus de 3400 Hz, il convient d'appliquer le filtrage passe-bas approprié afin de satisfaire aux prescriptions de repliement ou de reconstruction. Des réalisations logicielles de ces filtres sont disponibles dans la bibliothèque ITU-T STL.

TABLEAU D.1/P.830

| fréquence (Hz) | caractéristique d'émission de l'IRS modifié (dBV/Pa) | caractéristique de réception de l'IRS modifié (dBPa/V) | affaiblissement d'insertion du filtre SRAEN (dB) |
|----------------|--|--|--|
| 100            | -31,7  | -13,4  | 14,1   |
| 125            | -24,7  | -7,4   | 11,4   |
| 160            | -17,2  | -2,4   | 8,4  |
| 200            | -13,3  | 3,2  | 5,9  |
| 250            | -10,3  | 6,7  | 4,0  |
| 300            | -8,5   | 9,2  | 2,8  |
| 315            | -8,3   | 9,7  | 2,5  |
| 400            | -7,0   | 11,3   | 1,4  |
| 500            | -6,3   | 11,9   | 10,6   |
| 600            | -6,0   | 12,1   | 0,3  |
| 630            | -5,9   | 12,1   | 0,2  |
| 800            | -4,9   | 12,3   | 0,0  |
| 1000           | -3,7   | 12,6   | 0,0  |
| 1250           | -2,3   | 12,5   | 0,0  |
| 1600           | -0,5   | 13,1   | 0,1  |
| 2000           | 0,1  | 12,9   | -0,2   |
| 2500           | 1,3  | 12,6   | -0,5   |
| 3000           | 2,0  | 13,0   | 0,5  |
| 3150           | 2,1  | 12,9   | 0,3  |
| 3500           | -0,3   | 10,9   | 7,0  |
| 4000           | -3,5   | 2,1  | 33,7   |
| 5000           | -9,0   | -11,7  | 43,2   |

## Bibliographie

- [1] BOYD (I.), SOUTHCOTT (C.B.): A speech codec for the Skyphone service, *British Telecom Technology Journal*, Vol. 6, n° 2, avril 1988.
- [2] BT Laboratories: Enhanced Equivalent-Q Rating Algorithm, *ETSI/TM/TM5/TCH-HS Document TD 93/126*, décembre 1993.

- [3] CCITT – Contribution COM XII-79: *Spécification d'un système de référence intermédiaire*, période d'études 1973-1976.
- [4] CCITT – Contribution COM XII-104: *Récapitulation et analyse des résultats de mesures subjectives et objectives des équivalents pour la sonie, effectuées sur onze systèmes téléphoniques par le Laboratoire du CCITT*, période d'études 1973-1976.
- [5] CCITT – Rapport de la réunion du Groupe de travail XVIII/2 (Traitement des signaux vocaux), *COM XVIII-R 28, Annexe 1*, pp. 13 à 39, décembre 1983.
- [6] COLEMAN (A.), GLEISS (N.), SOTSCHECK (J.), USAI (P.), SCHEUERMANN (H.): Subjective performance evaluation of the RPE-LTP codec for the Pan-European cellular digital mobile radio system, *IEEE Globecom 89*, Dallas, Texas, 27-30 novembre 1989.
- [7] COLEMAN (A.), GLEISS (N.), USAI (P.): A subjective Testing Methodology for Evaluating Medium Rate Codecs for Digital Mobile Applications, *Speech Communications*, Vol. 7, pp. 151-166, juin 1988.
- [8] COMBESCURE (P.) *et al.*: Quality evaluation of speech coded at 32 kbit/s by means of degradation category ratings, *Proc. ICASSP 82 (International Conference on Acoustics, Speech and Signal Processing)*, Vol. 2, Paris, mai 1982.
- [9] CROWE (D.P.): Selection of Voice Codec for the Aeronautical Satellite Service, *European Conference on Speech Communication and Technology*, Vol. 2, S37, pp. 320-323, septembre 1989.
- [10] DAUMER (W.R.), CAVANAUGH (J.R.): A subjective comparison of selected digital codecs for speech, *Bell Systems Technical Journal*, Vol. 57, n° 9, novembre 1978.
- [11] DIMOLITSAS (S.), CORCORAN (F.), RAVISHANKAR (C.): Correlation between headphone and telephone handset listener opinion scores for single-stimulus voice coder performance assessments, *IEEE Signal Processing Letters*, Vol. 2, n° 3, mars 1995.
- [12] Groupe mixte d'experts européens: Méthode d'essai subjectif pour l'évaluation des codecs à faible débit binaire pour les radiocommunications mobiles, *CCITT, COM XII-68*, mai 1986.
- [13] GOODMAN (D.J.), NASH (R.D.): Subjective quality of the same speech transmission conditions in seven different countries, *Proc. ICASSP 82 (International Conference on Acoustics, Speech and Signal Processing)*, Vol. 2, Paris, mai 1982.
- [14] UIT-T – Contribution COM 15-20: Transmission quality of interconnected PSTN-digital cellular networks, *COMSAT*, période d'études 1993-1996.
- [15] MODENA (G.), COLEMAN (A.), USAI (P.), COVERDALE (P.): Subjective performance evaluation of the 7 kHz audio coder, *IEEE Global Telecommunications Conference 1986 (Globecom '86)*, Houston, Texas, 1-4 décembre 1986.
- [16] RICHARDS (D.L.), BARNES (G.J.): Pay-off between quantizing distortion and injected circuit noise, *Proc. ICASSP 82 (International Conference on Acoustics, Speech and Signal Processing)*, Vol. 2, Paris, mai 1982.
- [17] Administration suédoise des télécommunications: Rapport sur les essais subjectifs de codecs susceptibles d'être utilisés pour les radiocommunications mobiles, *CCITT, COM XII-147*, février 1987.
- [18] WHEDDON (C.), LINGGARD (R.): *Speech and Language Processing*, Chapman and Hall, 1990.