



UNION INTERNATIONALE DES TÉLÉCOMMUNICATIONS

UIT-T

SECTEUR DE LA NORMALISATION
DES TÉLÉCOMMUNICATIONS
DE L'UIT

P.861

(02/98)

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

Méthodes d'évaluation objective et subjective de la qualité

**Mesure objective de la qualité des codecs
vocaux fonctionnant en bande téléphonique
(300-3400 Hz)**

Recommandation UIT-T P.861

(Antérieurement Recommandation du CCITT)

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

Vocabulaire et effets des paramètres de transmission sur l'opinion des usagers	Série	P.10
Lignes et postes d'abonnés	Série	P.30 P.300
Normes de transmission	Série	P.40
Appareils de mesures objectives	Série	P.50 P.500
Mesures électroacoustiques objectives	Série	P.60
Mesures de la sonie vocale	Série	P.70
Méthodes d'évaluation objective et subjective de la qualité	Série	P.80 P.800
Qualité audiovisuelle dans les services multimédias	Série	P.900

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

RECOMMANDATION UIT-T P.861

MESURE OBJECTIVE DE LA QUALITE DES CODECS VOCAUX FONCTIONNANT EN BANDE TELEPHONIQUE (300-3400 Hz)

Résumé

La présente Recommandation décrit une méthode objective pour estimer la qualité subjective des codecs vocaux utilisés dans la bande téléphonique (300-3400 Hz).

La présente Recommandation spécifie la production de signaux vocaux de source pour la mesure objective de leur qualité, ainsi que les conditions de référence pour lesquelles on a démontré que cette méthode de mesure objective de qualité fournissait des résultats valides. Elle spécifie également le calcul de la qualité objective sur la base de la mesure objective de qualité appelée *mesure de la qualité de parole perçue* (PSQM, *perceptual speech quality measure*), l'estimation de la qualité subjective à partir des résultats des mesures objectives et l'analyse de ces résultats.

La présente Recommandation peut être appliquée lors de l'évaluation des effets, sur la qualité subjective des codecs vocaux, de différents niveaux d'entrée de parole, de différents locuteurs, de différents débits et de différents transcodages.

Source

La Recommandation UIT-T P.861, révisée par la Commission d'études 12 de l'UIT-T (1997-2000), a été approuvée le 27 février 1998 selon la procédure définie dans la Résolution n° 1 de la CMNT.

Mots clés

Evaluation subjective de la qualité des paroles, mesure objective de la qualité des paroles.

AVANT-PROPOS

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

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.

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

NOTE

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

DROITS DE PROPRIÉTÉ INTELLECTUELLE

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

A la date d'approbation de la présente Recommandation, l'UIT avait été avisée de l'existence d'une propriété intellectuelle protégée par des brevets à acquérir pour mettre en œuvre la présente Recommandation. Toutefois, comme il ne s'agit peut-être pas de renseignements les plus récents, il est vivement recommandé aux responsables de la mise en œuvre de consulter la base de données des brevets du TSB.

© UIT 1998

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

		Page
1	Domaine d'application	1
2	Références normatives	2
3	Abréviations	3
4	Définitions	3
5	Conventions	3
6	Résumé de la procédure de mesure objective	4
7	Préparation du matériel vocal source	4
7.1	Voix réelles	5
7.2	Voix artificielles	5
8	Sélection des paramètres expérimentaux	5
9	Calcul de la qualité objective	6
9.1	Initialisations globales	13
	9.1.1 Alignement temporel	13
	9.1.2 Échelonnement global	14
	9.1.3 Etalonnage global	14
9.2	Mappage temps-fréquence	15
	9.2.1 Fenêtrage.....	15
	9.2.2 Densité spectrale de puissance (SPD, <i>spectral power density</i>) échantillonnée	16
9.3	Prédistorsion et filtrage des fréquences	16
	9.3.1 Densité de puissance fondamentale échantillonnée.....	16
	9.3.2 Échelonnement local.....	16
	9.3.3 Filtrage de la bande téléphonique	17
	9.3.4 Bruit spectral de Hoth.....	17
9.4	Prédistorsion d'intensité	18
9.5	Modélisation cognitive	18
	9.5.1 Échelonnement en sonie	18
	9.5.2 Densité échantillonnée du bruit perturbateur.....	19
	9.5.3 Traitement des asymétries	19
	9.5.4 Traitement du bruit perturbateur, y compris les intervalles de silence.....	19
10	Transformation de l'échelle de qualité objective en échelle de qualité subjective.....	20
10.1	Notes moyennes d'opinion	21
10.2	Notes Q équivalentes	21
11	Analyse des résultats.....	22

	Page
Appendice I – Contenu de la disquette accompagnant la Recommandation P.861	23
I.1 Introduction.....	23
I.2 Répertoire\test.....	23
Appendice II – Mesure objective de la qualité des codecs vocaux fonctionnant en bande téléphonique (300-3400 Hz) au moyen de modules de normalisation de mesure (MNB, <i>measuring normalizing blocks</i>)	25
II.1 Introduction.....	25
II.2 Calcul de la mesure objective	26
II.2.1 Spécifications des entrées et des sorties	28
II.2.2 Temps de propagation.....	29
II.2.3 Préparation des signaux	29
II.2.4 Transformation en domaine fréquentiel.....	29
II.2.5 Sélection de trame	30
II.2.6 Approximation de la sonie perçue	30
II.2.7 Module de normalisation de mesure fréquentielle (FMNB).....	31
II.2.8 Calcul des modules de normalisation de mesure temporelle.....	31
II.2.9 Combinaison linéaire de mesures pour la structure MNB.....	33

Introduction

L'évaluation de la qualité subjective des codecs vocaux est une des techniques fondamentales qui permettent de concevoir des réseaux de télécommunication numériques. La Recommandation P.830 définit des méthodes subjectives d'essai pour codecs vocaux. Comme l'évaluation de la qualité subjective est coûteuse en termes de temps et d'argent, il est souhaitable de mettre au point une méthode d'évaluation objective de la qualité pour estimer la qualité subjective des codecs vocaux avec moins d'essais subjectifs.

Pour démontrer les performances des codecs vocaux, la mesure objective de qualité des signaux vocaux la plus communément utilisée est celle du rapport signal sur bruit (S/B). On fait toutefois observer que le rapport S/B ne prédit pas convenablement la qualité subjective pour composants de réseau modernes. Cela vaut particulièrement pour les récents codecs à bas débit. Diverses mesures objectives de qualité, plus élaborées, ont donc été mises au point, comme les suivantes: la mesure de la distance (CD, *cepstrum distance measure*) spectrale par codage LPC [1]; la méthode de l'indice d'information (II) [2]; la fonction de cohérence (CHF, *coherence function*) [3]; la méthode de reconnaissance de formes par système expert (EPR, *expert pattern recognition*) [4] et la mesure de la qualité de la parole perçue (PSQM) [5]. La performance de ces systèmes, exprimée en termes de capacité à donner des estimations fidèles de la qualité subjective, font l'objet de recherches à l'UIT-T depuis les années 1980.

A la suite de comparaisons précises entre mesures objectives de qualité par ces méthodes, on a tiré la conclusion que la mesure PSQM présentait la meilleure corrélation avec la qualité subjective des paroles codées. La présente Recommandation décrit donc l'évaluation de qualité objective au moyen de la méthode PSQM [12].

Afin d'aider les lecteurs de la présente Recommandation à mettre au point leur propre réalisation de la mesure PSQM, une disquette a été jointe à la présente Recommandation. On trouvera une description du contenu de cette disquette dans le fichier README de celle-ci ainsi que dans l'Appendice I.

Recommandation P.861

MESURE OBJECTIVE DE LA QUALITE DES CODECS VOCAUX FONCTIONNANT EN BANDE TELEPHONIQUE (300-3400 Hz)

(Révisée en 1998)

1 Domaine d'application

L'évaluation subjective de la qualité des codecs vocaux peut être effectuée au moyen de tests d'écoute seulement (dans un seul sens) ou au moyen de test de conversation (dans les deux sens). La mesure objective de qualité qui est décrite dans la présente Recommandation estime la qualité subjective lors d'essais d'écoute seulement₂.

Pour démontrer la performance subjective d'un codec, il faut examiner les effets de divers facteurs de qualité (voir la Recommandation P.830). L'exactitude de la mesure objective de qualité décrite dans la présente Recommandation n'a pas été vérifiée par examen de tous les facteurs spécifiés dans la Recommandation P.830. Le Tableau 1 est destiné à servir de guide pour faciliter la détermination, par le lecteur, des facteurs d'essai, des techniques de codage et des applications auxquelles cette Recommandation s'applique.

Tableau 1/P.861 – Relation entre facteurs expérimentaux, techniques de codage et applications selon la présente Recommandation

Facteurs expérimentaux	Note
niveaux d'entrée des signaux de parole dans un codec	1
niveaux d'écoute lors d'expériences subjectives	2
variations selon le locuteur	1
locuteurs multiples simultanés	2
erreur de transmission dans la voie	2
débits lorsque le codec peut fonctionner selon plusieurs modes	1
transcodages	1
discordance de débit entre un codeur et un décodeur si un codec possède plusieurs modes de débit	2
bruit ambiant du côté émission	2
signaux d'information de couche Réseau à l'entrée d'un codec	2
signaux de musique à l'entrée d'un codec	2
temps de propagation	3
prédistorsion à court terme du signal audio	2
prédistorsion à long terme du signal audio	4
écrêtage temporel du signal vocal	2
écrêtage en amplitude du signal vocal	2

Tableau 1/P.861 – Relation entre facteurs expérimentaux, techniques de codage et applications selon la présente Recommandation (*fin*)

techniques de codage	Note
codecs temporels	1
codecs CELP et hybrides ≥ 4 kbit/s	1
codecs CELP et hybrides < 4 kbit/s	2
VOCODEURS	2
autres codeurs	2
applications	
optimisation du codeur	1
évaluation du codeur	1
sélection du codeur	2
planification du réseau	5
essais actifs dans le réseau	6
dispositifs de mesure sans intrusion en service	3
<p>NOTE 1 – La mesure objective a montré une précision acceptable en présence de cette variable.</p> <p>NOTE 2 – On ne dispose pas d'assez de renseignements au sujet de la précision de la mesure objective en ce qui concerne cette variable.</p> <p>NOTE 3 – La mesure objective est réputée fournir des prédictions inexactes lorsqu'elle est utilisée conjointement avec cette variable ou n'est pas censée être utilisée avec cette variable.</p> <p>NOTE 4 – La mesure objective est réputée fournir des prédictions inexactes en présence d'un dérapage assez important (plus de 10% de la longueur de trame). La possibilité d'appliquer la mesure en présence d'un dérapage moins important fera l'objet d'un complément d'étude.</p> <p>NOTE 5 – Au prix de certaines précautions, on peut utiliser la mesure objective pour certains objets de planification de réseau. Le lecteur doit prendre acte du fait qu'il existe d'importants facteurs de planification de réseau auxquels la présente Recommandation n'est pas applicable (voir la section "Facteurs expérimentaux" du présent tableau).</p> <p>NOTE 6 – Au prix de certaines précautions, on peut utiliser la méthode objective pour certains essais actifs sur le réseau. Le lecteur doit prendre acte du fait qu'il peut exister, dans une chaîne de connexion établie dans un réseau actif, des facteurs ou des techniques auxquels la présente Recommandation n'est pas applicable (voir les sections "Facteurs expérimentaux" et "Techniques de codage" du présent tableau).</p>	

Lorsque l'on compare un codec à un autre codec ou à une condition de référence fondée sur des résultats d'expériences subjectives, on fait souvent appel à des essais statistiques qui tiennent compte de la distribution des notes subjectives. Etant donné que, dans la présente Recommandation, la mesure objective n'estime que la moyenne des notes subjectives (par exemple notes MOS, notes DMOS), de tels essais statistiques ne peuvent pas être appliqués aux résultats de mesures objectives. La prédiction du pourcentage de notes au mieux médiocres (%PoW, *per cent poor or worse*) et du pourcentage de notes au moins bonnes (%GoB, *per cent good or better*) est actuellement à l'étude.

2 Références normatives

La présente Recommandation se réfère à certaines dispositions des Recommandations UIT-T et textes suivants qui de ce fait en sont partie intégrante. Les versions indiquées étaient en vigueur au moment de la publication de la présente Recommandation. Toute Recommandation ou tout texte

étant sujet à révision, les utilisateurs de la présente Recommandation sont invités à se reporter, si possible, aux versions les plus récentes des références normatives suivantes. La liste des Recommandations de l'UIT-T en vigueur est régulièrement publiée.

- Recommandation G.711 du CCITT (1988), *Modulation par impulsions et codage (MIC) des fréquences vocales*.
- 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 G.729 (1996), *Codage de la parole à 8 kbit/s en utilisant la prédiction linéaire avec excitation par séquences codées à structure algébrique conjuguée (CS-ACELP)*.
- Recommandation UIT-T P.50 (1993), *Voix artificielle*.
- Recommandation UIT-T P.800 (1996), *Méthodes d'évaluation subjective de la qualité de transmission*.
- Recommandation UIT-T P.810 (1996), *Appareil de référence à bruit modulé*.
- Recommandation UIT-T P.830 (1996), *Évaluation subjective de la qualité des codecs numériques à bande téléphonique et à large bande*.
- Supplément n° 13 des Recommandations de la série P du CCITT (1988), *Spectres de bruit*.

3 Abréviations

La présente Recommandation utilise les abréviations suivantes:

- ACR évaluation par catégories absolues (*absolute category rating*)
- CELP prédiction linéaire avec excitation par code (*code excited linear prediction*)
- DCR évaluation par catégories de dégradation (*degradation category rating*)
- DMOS note moyenne d'opinion de dégradation (*degradation mean opinion score*)
- MOS note moyenne d'opinion (*mean opinion score*)
- PSQM mesure de la qualité de parole perçue (*perceptual speech quality measure*)

4 Définitions

La présente Recommandation définit le terme suivant:

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

5 Conventions

L'évaluation subjective des codecs vocaux peut être conduite au moyen d'essais d'écoute seulement ou au moyen de méthodes d'essais subjectifs par conversation. Pour des raisons pratiques, les essais d'écoute seulement sont la seule méthode possible d'essais subjectifs au cours de la mise au point de codecs vocaux, lorsqu'une mise en œuvre en temps réel de ces codecs n'est pas réalisable. La présente Recommandation expose une technique de mesure objective qui permet d'estimer la qualité subjective obtenue lors d'essais d'écoute seulement.

6 Résumé de la procédure de mesure objective

La Figure 1 illustre la procédure de mesure objective.

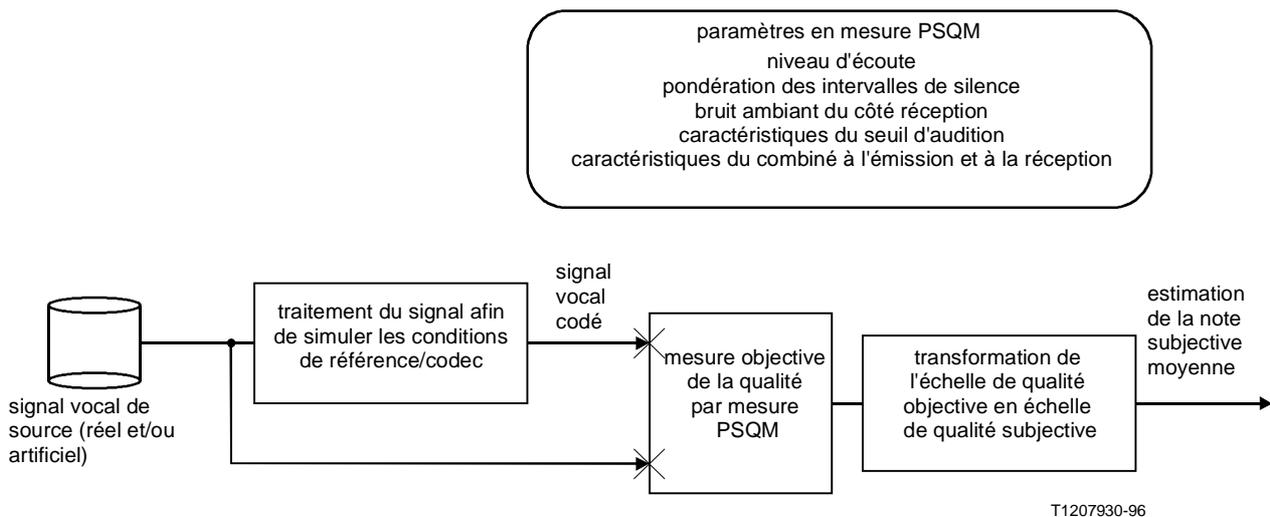


Figure 1/P.861 – Procédure de mesure objective de la qualité

La mesure objective de la qualité des codecs vocaux exige un certain nombre d'étapes:

- 1) préparation des matériaux de source, c'est-à-dire enregistrement des locuteurs et/ou production des voix artificielles conformément à la Recommandation P.50;
- 2) sélection des paramètres expérimentaux qui mettront à contribution les principales caractéristiques du codec et qui pourront être contrôlés par mesure objective;
- 3) production de signaux vocaux codés/de référence;
- 4) calcul de la qualité objective de la parole sur la base de la mesure de la qualité de parole perçue (PSQM), au moyen de signaux vocaux de source et codés;
- 5) transformation, si nécessaire, de l'échelle de qualité objective en échelle de qualité subjective;
- 6) analyse des résultats.

Chacune de ces étapes est décrite ci-dessous.

7 Préparation du matériel vocal source

Les signaux de source pour mesure objective peuvent être des voix réelles ou les voix artificielles spécifiées dans la Recommandation P.50, selon les buts de l'expérience.

Etant donné que les voix artificielles définies dans la Recommandation P.50 reproduisent les caractéristiques moyennes de la parole humaine dans diverses langues, ces voix sont utiles pour estimer objectivement la qualité subjective moyenne d'un codec utilisé avec ces langues. Lorsqu'il s'agit d'évaluer la relation d'un codec avec le locuteur ou la performance d'un codec pour des langues particulières, il est recommandé d'utiliser des voix réelles. Dans un cas comme dans l'autre, aucun bruit ambiant ne doit être ajouté.

7.1 Voix réelles

Lorsque des voix réelles sont utilisées en mesure objective, ces voix doivent être produites, enregistrées et égalisées en niveau conformément au paragraphe 7/P.830.

Il est recommandé qu'un minimum de deux locuteurs de sexe masculin et deux locutrices de sexe féminin soient mis à contribution pour chaque condition d'essai. Si la relation avec le locuteur doit être évaluée en tant que facteur autonome, il est recommandé de faire appel à un plus grand nombre de locuteurs comme suit:

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

7.2 Voix artificielles

Lorsqu'on utilise des voix artificielles conformes à la Recommandation P.50 pour une mesure objective, il est recommandé d'utiliser des voix artificielles aussi bien masculines que féminines. On doit faire passer ces signaux dans un filtre ayant des caractéristiques appropriées en fréquence afin de simuler la courbe de fréquence en émission d'un combiné téléphonique. Ces signaux doivent ensuite être égalisés en niveau comme les voix réelles (voir la Recommandation P.830).

L'UIT-T recommande l'utilisation de la caractéristique de fréquence d'émission par le système de référence intermédiaire (IRS, *intermediate reference system*) modifié, tel que défini dans l'Annexe D/P.830.

8 Sélection des paramètres expérimentaux

Pour démontrer la performance d'un codec, il y a lieu d'examiner les effets de divers facteurs de qualité sur cette performance. La Recommandation P.830 donne des directives sur l'évaluation subjective des facteurs de qualité ci-après:

- 1) niveaux d'entrée des signaux de parole dans un codec;
- 2) niveaux d'écoute lors d'expériences subjectives;
- 3) locuteurs (y compris multiples locuteurs simultanés);
- 4) erreurs dans la voie de transmission entre un codeur et un décodeur;
- 5) débits lorsque le codec peut fonctionner selon plusieurs modes;
- 6) transcodages;
- 7) désadaptation des débits entre un codeur et un décodeur si un codec possède plusieurs modes de débit;
- 8) bruit ambiant du côté émission;
- 9) signaux d'information de couche Réseau à l'entrée d'un codec;
- 10) signaux musicaux à l'entrée d'un codec.

Etant donné que la mesure objective de qualité décrite dans la présente Recommandation part du principe:

- 1) que les signaux source de parole sont "propres" (c'est-à-dire sans bruit ambiant ajouté du côté émission);

- 2) qu'il n'y a aucune dégradation de voie telle que des erreurs sur les bits de transmission, des effacements de trame (comme cela peut être le cas dans des applications de radiocommunications mobiles), ou des pertes de cellules (comme cela peut être le cas dans les réseaux en mode ATM),

les facteurs de qualité auxquels la présente Recommandation s'applique sont les niveaux d'entrée des signaux de parole, les locuteurs (sauf les multiples locuteurs simultanés), les débits et les transcodages.

NOTE 1 – La mesure objective de facteurs de qualité autres que ceux qui sont expressément indiqués dans la présente Recommandation comme étant applicables est encore à l'étude. Ces facteurs ne devront donc être mesurés qu'après vérification de l'exactitude de la mesure objective en liaison avec les essais subjectifs selon la Recommandation P.830.

NOTE 2 – Bien que certains éléments donnent à penser que la mesure objective peut prédire avec précision la qualité en conditions de dégradation des voies [10] [11], l'applicabilité de cette mesure à ces conditions est encore à l'étude.

En plus des conditions du codec, la Recommandation P.830 recommande d'utiliser des conditions de référence lors des essais subjectifs. Ces conditions sont nécessaires pour faciliter la comparaison des résultats d'essais subjectifs issus de différents laboratoires ou du même laboratoire à différentes dates. De même, lorsqu'on exprime les résultats d'essais subjectifs en termes de notes Q équivalentes, il convient de rapporter les essais aux conditions de référence au moyen de l'appareil de référence à bruit modulé (MNRU, *modulated noise reference unit*) qui est spécifié dans la Recommandation P.810.

NOTE 3 – L'inclusion dans les mesures objectives de qualité d'autres codecs normalisés tels qu'à codages MIC G.711 à 64 kbit/s, MICDA G.726 à 32 kbit/s, LD-CELP G.728 à 16 kbit/s et CS-ACELP G.729 à 8 kbit/s, ainsi que de l'appareil MNRU, peut contribuer à démontrer la performance relative du codec à l'essai et des codecs normalisés.

La Recommandation P.830 contient des explications détaillées sur ces paramètres expérimentaux.

9 Calcul de la qualité objective

Le présent paragraphe décrit une méthode de mesure de la qualité des signaux de parole codés dans la bande téléphonique (300-3400 Hz) au moyen de la mesure de qualité de la parole perçue (PSQM). La mesure PSQM consiste à simuler la perception acoustique de sujets placés en situation de la vie réelle [6]. La mesure PSQM simule des expériences au cours desquelles des sujets évaluent la qualité de codecs vocaux. Pour cela, elle compare un signal codé à un signal source (Figure 2). Bien que ce principe fondamental de comparaison soit particulièrement adapté aux essais d'évaluation par catégories de dégradation (DCR, *degradation category rating*), on peut simuler des expériences d'évaluation par catégories absolues (ACR, *absolute category rating*) comme indiqué dans les essais de validation [12]. Dans la mesure où la mesure PSQM est une représentation fidèle des processus de perception et de jugement humains, des différences inaudibles entre entrée et sortie feront l'objet de notes identiques en mesure PSQM. En particulier, si l'entrée et la sortie sont identiques, la mesure PSQM prédira une qualité parfaite sans tenir compte de la qualité du signal d'entrée.

Dans le cadre de la mesure PSQM, les signaux physiques constituant les paroles de source et les paroles codées sont mappés sur des représentations psychophysiques aussi proches que possible des représentations internes des signaux de parole (tels qu'ils sont perçus par le cerveau humain). Ces représentations internes font appel aux équivalents psychophysiques de la fréquence (niveaux de bande critique) et de l'intensité (équivalents pour la sonie comprimés). Le masquage est modélisé de manière simple: il n'est pris en compte que lorsque deux composantes de temps-fréquence coïncident dans les deux domaines, temps et fréquence.

Dans la méthode de mesure PSQM, la qualité des paroles codées est jugée sur la base de différences entre représentations internes. Ces différences permettent de calculer le bruit perturbateur en fonction du temps et de la fréquence. Dans la méthode de mesure PSQM, le bruit perturbateur moyen est directement associé à la qualité de la parole codée.

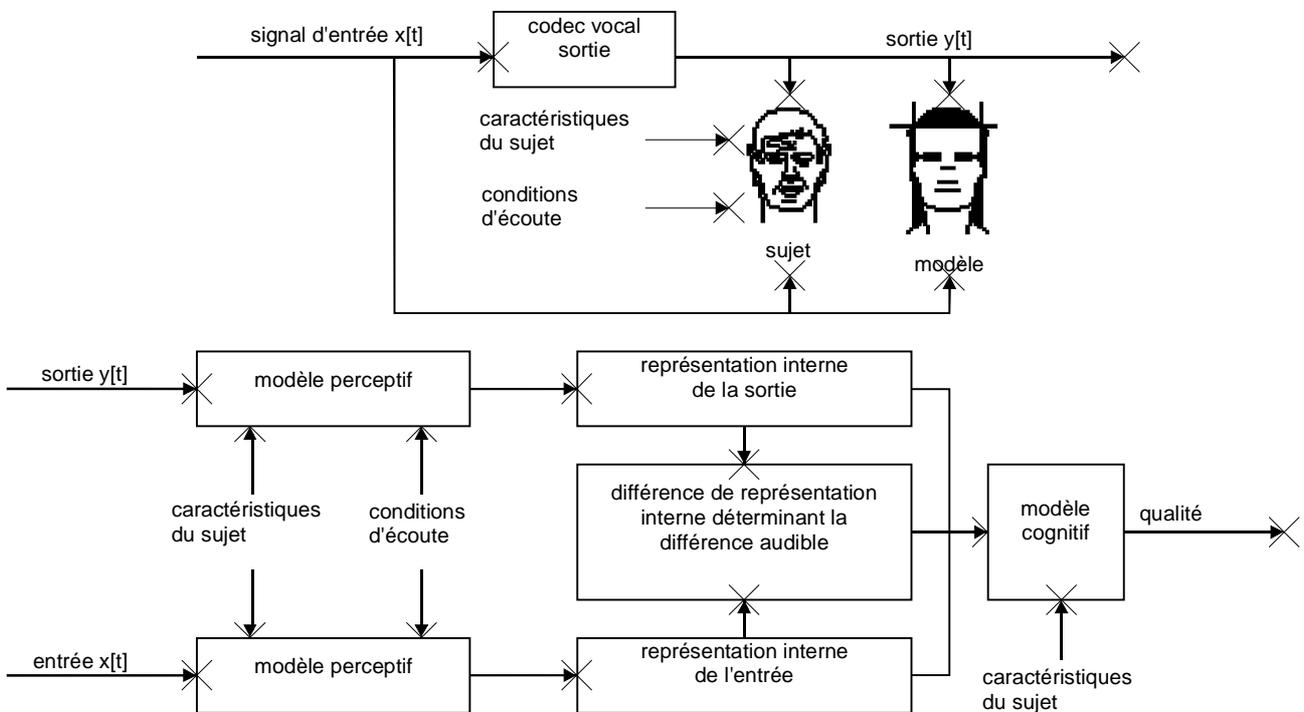
La transformation du domaine physique (externe) au domaine psychophysique (interne) s'effectue en trois phases comme suit:

- mappage temps-fréquence;
- prédistorsion en fréquence;
- prédistorsion en intensité (compression).

En plus de la modélisation perceptive, la mesure PSQM fait appel à la modélisation cognitive [7] afin d'obtenir d'importants coefficients de corrélation entre mesures subjectives et objectives.

La Figure 3 montre schématiquement l'algorithme PSQM.

Tous les paramètres et toutes les variables du présent paragraphe sont résumés dans les Tableaux 2, 3 et 4.



T1207940-96

Figure 2/P.861 – Aperçu général des principes fondamentaux suivis lors de l'élaboration de la méthode PSQM – Un modèle informatique du sujet, composé d'un modèle perceptif et d'un modèle cognitif, est utilisé pour comparer la sortie du codec vocal à son entrée

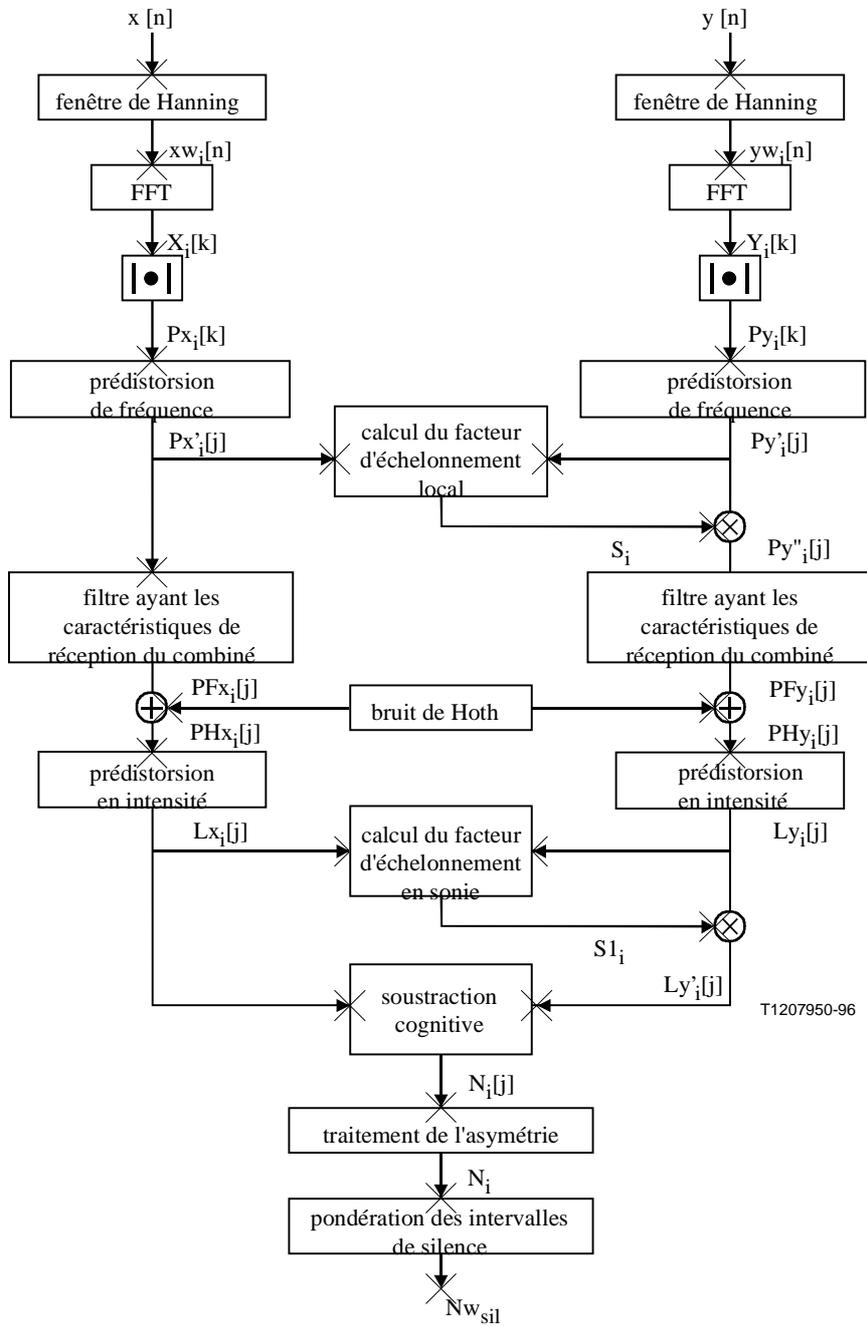


Figure 3/P.861 – Schéma fonctionnel de l'algorithme PSQM

Tableau 2/P.861 – Liste des paramètres en mesure PSQM

Nom	Description	Valeur
Nb	nombre de bandes dans le domaine des bandes critiques (de Bark)	(voir le Tableau 4)
Nf	nombre d'échantillons dans une trame temporelle	512 pour une fréquence d'échantillonnage de 16 kHz 256 pour une fréquence d'échantillonnage de 8 kHz
F[j]	réponse en fréquence du combiné à la réception	système IRS selon la Recommandation P.830 (le Tableau 4 contient la fonction de transfert de puissance pour le système IRS)
H[j]	caractéristiques de bruit spectral de Hoth	(le Tableau 4 contient la puissance additive correspondant à la caractéristique de bruit spectral de Hoth)
P ₀ [j])	seuil absolu d'audition	(le Tableau 4 contient la représentation équivalente en puissance de P ₀ [j])
Δf[j]	largeur de bande de la bande j en Hertz	(voir le Tableau 4)
Δz	largeur de chaque sous-bande dans le domaine des bandes critiques	0,312
γ	exposant de la fonction de compression	0,001
W _{sil}	facteur de pondération sur trames de silence	0,2 (provisoire)
W _{sp}	facteur de pondération sur trames de conversation active	$W_{sp} = (1 - W_{sil})/W_{sil} = 4,0$ (provisoire)

Tableau 3/P.861 – Variables utilisées dans la méthode PSQM

Nom	Description
m	index dans le domaine temporel
n	index dans le domaine temporel pour une trame (n: 1, 2, 3, , Nf)
i	index pour les trames
j	index dans le domaine fréquentiel prédistordu (domaine des bandes critiques) (j: 1, 2, 3, , Nb)
k	index dans le domaine fréquentiel (Hz) (k: 1, 2, 3, , Nf/2)
x[m]	version alignée temporellement et étalonnée globalement du signal vocal de source échantillonné
y[m]	version alignée temporellement, échelonnée globalement et étalonnée globalement du signal vocal codé et échantillonné
S _{global}	facteur d'échelonnement en échelonnement global
S _p	facteur d'étalonnage de la puissance fondamentale
S _l	facteur d'étalonnage de la sonie fondamentale
x _i [n]	x[m] dans la trame I

Tableau 3/P.861 – Variables utilisées dans la méthode PSQM (suite)

Nom	Description
$y_i[n]$	$y[m]$ dans la trame I
$xw_i[n]$	version fenêtrée de $x_i[n]$
$yw_i[n]$	version fenêtrée de $y_i[n]$
$X_i[k]$	transformée FFT de $xw_i[n]$
$Y_i[k]$	transformée FFT de $yw_i[n]$
$Px_i[k]$	densité SPD de $xw_i[n]$
$Py_i[k]$	densité SPD de $yw_i[n]$
$I_i[j]$	indice de transformée de Fourier rapide (FFT) de la première valeur de k pour $Px_i[k]$ et $Py_i[k]$ dans la bande j
$I_i[j]$	indice de transformée de Fourier rapide (FFT) de la dernière valeur de k pour $Px_i[k]$ et $Py_i[k]$ dans la bande j
$Px_i^{\prime}[j]$	densité de puissance fondamentale échantillonnée de $xw_i[n]$
$Py_i^{\prime}[j]$	densité de puissance fondamentale échantillonnée de $yw_i[n]$
Px_i^{\cdot}	puissance du signal vocal de source dans la trame i
Py_i^{\cdot}	puissance du signal vocal codé dans la trame i
$Py_i^{\prime\prime}[j]$	version échelonnée localement de $Py_i^{\prime}[j]$
$PFx_i[j]$	version filtrée en bande téléphonique de $Px_i^{\prime}[j]$
$PFy_i[j]$	version filtrée en bande téléphonique de $Px_i^{\prime\prime}[j]$
$Phx_i[j]$	$PFx_i[j]$ plus bruit spectral de Hoth en tant que bruit ambiant (en réception)
$Phy_i[j]$	$PFy_i[j]$ plus bruit spectral de Hoth en tant que bruit ambiant (en réception)
S_i	facteur d'échelonnement lors de l'échelonnement local dans la trame i
S_{av}	moyenne (arithmétique) de S_i
$Lx_i[j]$	densité d'équivalents pour la sonie comprimés après échantillonnage du signal vocal de source dans la trame I et dans la bande j
$Ly_i[j]$	densité d'équivalents pour la sonie comprimés après échantillonnage du signal vocal codé dans la trame I et dans la bande j
Lx_i	valeur instantanée des équivalents pour la sonie comprimés du signal vocal de source dans la trame I
Ly_i	valeur instantanée des équivalents pour la sonie comprimés du signal vocal codé dans la trame I
Sl_i	facteur d'échelonnement en termes de sonie dans la trame i
$Ly_i^{\prime}[j]$	version échelonnée en sonie de $Ly_i[j]$
$N_i[j]$	densité de bruit perturbateur échantillonnée dans la trame i et dans la bande j
$C_i[j]$	facteur d'effet asymétrique dans la trame i et dans la bande j
N_i	bruit perturbateur dans la trame I

Tableau 3/P.861 – Variables utilisées dans la méthode PSQM (fin)

Nom	Description
N_{wsil}	bruit perturbateur moyen avec pondération des trames de silence
M_{sp}	nombre de trames de conversation active
M_{sil}	nombre de trames de silence
N_{spav}	valeur moyenne de N_i dans les trames de conversation active
N_{silav}	valeur moyenne de N_i dans les trames de silence

Tableau 4/P.861 – Attribution de fréquences dans les bandes critiques et caractéristiques des filtres correspondants (sur la base d'une fréquence d'échantillonnage de 16 kHz)

Numéro de la bande, j	Fréquence supérieure [Hz]	Premier segment de transformée FFT dans la bande j, I_f	Dernier segment de transformée FFT dans la bande j, I_l	Caractéristique de réception, F	Seuil d'audition P_0	Bruit spectral de Hoth, H
0	15,6	0	0	valeurs éliminées lors du traitement		
1	46,9	1	1	2,45E-06	3,89E+07	1,72E+04
2	78,1	2	2	9,24E-06	1,12E+06	1,72E+04
3	109,4	3	3	3,56E-05	1,26E+05	1,72E+04
4	140,6	4	4	2,59E-04	1,86E+04	1,22E+04
5	171,9	5	5	1,18E-03	6,17E+03	8,49E+03
6	203,1	6	6	7,48E-03	2,29E+03	6,31E+03
7	234,4	7	7	3,19E-02	9,33E+02	4,91E+03
8	265,6	8	8	7,31E-02	4,37E+02	3,95E+03
9	296,9	9	9	1,37E-01	2,29E+02	3,26E+03
10	328,1	10	10	2,09E-01	1,29E+02	2,74E+03
11	359,4	11	11	2,93E-01	7,76E+01	2,35E+03
12	390,6	12	12	4,25E-01	4,27E+01	2,04E+03
13	421,9	13	13	5,23E-01	3,02E+01	1,79E+03
14	453,1	14	14	5,98E-01	2,19E+01	1,59E+03
15	484,8	15	15	6,51E-01	1,66E+01	1,44E+03
16	519,2	16	16	6,94E-01	1,32E+01	1,39E+03
17	553,6	17	17	7,31E-01	1,07E+01	1,25E+03
18	590,8	18	18	7,66E-01	8,91E+00	1,22E+03
19	631,2	19	20	7,98E-01	7,59E+00	1,19E+03
20	672,9	21	21	8,37E-01	6,31E+00	1,10E+03
21	716,6	22	22	8,63E-01	5,62E+00	1,04E+03

**Tableau 4/P.861 – Attribution de fréquences dans les bandes critiques
et caractéristiques des filtres correspondants (sur la base
d'une fréquence d'échantillonnage de 16 kHz) (suite)**

Numéro de la bande, j	Fréquence supérieure [Hz]	Premier segment de transformée FFT dans la bande j, I_f	Dernier segment de transformée FFT dans la bande j, I_l	Caractéristique de réception, F	Seuil d'audition P₀	Bruit spectral de Hoth, H
22	760,4	23	24	8,88E-01	5,13E+00	9,45E+02
23	804,6	25	25	9,12E-01	4,68E+00	8,69E+02
24	851,4	26	27	9,35E-01	4,37E+00	8,41E+02
25	898,3	28	28	9,56E-01	4,17E+00	7,68E+02
26	947,0	29	30	9,71E-01	4,07E+00	7,33E+02
27	997,0	31	31	9,80E-01	3,98E+00	6,90E+02
28	1051,0	32	33	9,87E-01	3,98E+00	6,87E+02
29	1108,0	34	35	9,90E-01	3,98E+00	6,57E+02
30	1168,0	36	37	9,91E-01	3,98E+00	6,49E+02
31	1231,0	38	39	9,93E-01	3,98E+00	6,17E+02
32	1297,0	40	41	9,95E-01	4,07E+00	5,95E+02
33	1366,0	42	43	1,00E+00	4,27E+00	5,68E+02
34	1437,0	44	45	1,01E+00	4,47E+00	5,37E+02
35	1509,0	46	48	1,02E+00	4,68E+00	5,04E+02
36	1582,0	49	50	1,04E+00	5,01E+00	4,80E+02
37	1658,0	51	53	1,06E+00	5,37E+00	4,51E+02
38	1736,0	54	55	1,07E+00	5,62E+00	4,37E+02
39	1817,0	56	58	1,09E+00	5,89E+00	4,20E+02
40	1902,0	59	60	1,10E+00	6,31E+00	4,05E+02
41	1991,0	61	63	1,11E+00	6,61E+00	3,97E+02
42	2084,0	64	66	1,12E+00	6,92E+00	3,86E+02
43	2184,0	67	69	1,12E+00	7,24E+00	3,82E+02
44	2289,0	70	73	1,12E+00	7,59E+00	3,74E+02
45	2401,0	74	76	1,11E+00	7,76E+00	3,67E+02
46	2520,0	77	80	1,10E+00	7,94E+00	3,63E+02
47	2647,0	81	84	1,08E+00	7,94E+00	3,56E+02
48	2781,0	85	88	1,01E+00	7,94E+00	3,46E+02
49	2922,0	89	93	8,62E-01	7,94E+00	3,37E+02

**Tableau 4/P.861 – Attribution de fréquences dans les bandes critiques
et caractéristiques des filtres correspondants (sur la base
d'une fréquence d'échantillonnage de 16 kHz) (fin)**

Numéro de la bande, j	Fréquence supérieure [Hz]	Premier segment de transformée FFT dans la bande j, I_f	Dernier segment de transformée FFT dans la bande j, I_l	Caractéristique de réception, F	Seuil d'audition P_0	Bruit spectral de Hoth, H
50	3069,0	94	98	6,86E-01	8,13E+00	3,25E+02
51	3225,0	99	103	5,16E-01	8,13E+00	3,16E+02
52	3392,0	104	108	3,12E-01	8,32E+00	2,92E+02
53	3572,0	109	114	1,55E-01	8,32E+00	2,69E+02
54	3765,0	115	120	3,02E-02	8,32E+00	2,47E+02
55	3971,0	121	127	2,03E-03	8,32E+00	2,25E+02
56	4193,0	128	134	1,52E-04	8,32E+00	2,06E+02
NOTE 1 – Le seuil absolu, P_0 , utilise l'étalement 0 dBA = 1,0.						
NOTE 2 – La première fréquence supérieure (15,6 Hz) équivaut à la valeur 0,156 d'une bande critique. La largeur de bande Δz est égale à la valeur 0,312 d'une bande critique.						

9.1 Initialisations globales

Avant de commencer le calcul du bruit perturbateur, qui est obtenu à la sortie de l'algorithme PSQM, les initialisations globales ci-après doivent être effectuées pour chaque paire (avant et après codage) de signaux de parole:

- alignement temporel;
- échelonnement global pour la compensation du gain systématique;
- étalement global pour le réglage de la sonie des signaux vocaux.

Etant donné que les codecs vocaux en bande téléphonique adoptent généralement une fréquence d'échantillonnage de 8 kHz à l'entrée, la présente Recommandation part du principe que les signaux vocaux de source comme codés ont une fréquence d'échantillonnage de 8 kHz ou de 16 kHz (c'est-à-dire après suréchantillonnage d'un facteur 2).

9.1.1 Alignement temporel

La première initialisation globale qui doit être effectuée est l'alignement temporel du signal source $x[m]$ et du signal codé $y[m]$. Si ces signaux ne sont pas correctement alignés, la méthode PSQM ne peut pas être appliquée.

Lorsque l'on ne connaît pas la valeur théorique du retard du signal codé par rapport au signal source, on peut utiliser le retard qui donne la corrélation maximale entre les deux signaux afin d'estimer cette valeur. Pour les signaux subissant une distorsion du temps de propagation de groupe, le retard qui donne la plus petite valeur de mesure PSQM est celui qu'il faut retenir.

Au cours du traitement des signaux, on ignore les zéros de tête et de queue dans le fichier de parole puis on calcule les points de départ et d'arrêt en détectant l'activité vocale sur la base du seul signal de source. Les algorithmes permettant de déterminer le premier et le dernier échantillon de conversation active sont les suivants.

Lorsque l'on détermine le début d'une conversation active dans un fichier, le premier échantillon à déclarer actif est celui dont l'effectif (c'est-à-dire la valeur absolue), plus les effectifs des quatre échantillons précédents, a une valeur totale d'au moins 200. (Aux fins des essais des quatre premiers échantillons pour déterminer le début d'une conversation, les échantillons qui précèdent le premier échantillon sont considérés comme ayant la valeur 0.)

Lorsque l'on détermine la fin d'une conversation active dans un fichier, le dernier échantillon déclaré actif est le dernier échantillon dont l'effectif (c'est-à-dire la valeur absolue), plus celui des quatre échantillons suivants, a une valeur totale d'au moins 200. (Aux fins des essais des quatre derniers échantillons pour déterminer la fin d'une conversation, les échantillons qui suivent le dernier échantillon sont considérés comme ayant la valeur 0.)

9.1.2 Échelonnement global

Après le processus d'alignement temporel, le signal codé $y[m]$ est échelonné afin de compenser le gain global du système. Le facteur d'échelonnement S_{global} est défini comme suit:

$$S_{global} = \sqrt{\frac{\sum_{\text{point de départ}}^{\text{point d'arrêt}} x^2[m]}{\sum_{\text{point de départ}}^{\text{point d'arrêt}} y^2[m]}}$$

Le signal codé $y[m]$ est ensuite multiplié par le facteur S_{global} .

9.1.3 Etalonnage global

Pour assurer une exactitude optimale de la mesure objective, il est nécessaire d'effectuer un étalonnage entre le niveau d'écoute et la sonie comprimée. Les valeurs indiquées dans le Tableau 4 sont fondées sur l'hypothèse qu'un niveau de pression acoustique de 0 dB est équivalent à une valeur maximale de 1,0 dans le domaine des puissances de fondamentale telles que calculées au 9.3.1 [c'est-à-dire $\max_j(Px_i[j]) = 1,0$ pour une trame donnée]. On suppose également que le niveau d'écoute optimal de 78 dBA est utilisé conjointement avec les fichiers vocaux qui ont un niveau de conversation active égal à -26 dBov, comme indiqué dans la Recommandation P.830.

Les étalonnages sont effectués avec une onde sinusoïdale de 1 kHz à un niveau acoustique de 40 dB (c'est-à-dire -64 dBov). Il est préférable d'utiliser, à cette fin, une onde sinusoïdale réelle (c'est-à-dire composée de valeurs non entières) afin d'éviter des phénomènes de quantification lors de l'application de la fonction d'étalonnage. Un niveau acoustique de 40 dB correspond à une amplitude zéro à crête de 29,54.

Le premier étalonnage consiste à échelonner à 10 000 la valeur maximale de la représentation de puissance fondamentale contenue dans l'onde d'étalonnage [c'est-à-dire, si la valeur $\max_j(Px_i[j]) = 1,0$ pour un niveau acoustique de 0 dB, cette valeur $\max_j(Px_i[j]) = 10\,000$ pour un niveau acoustique de 40 dB]. Ce facteur d'étalonnage, S_p , est calculé comme suit:

$$S_p = \frac{10\,000}{\max_j(Px_i[j])}$$

lorsque la valeur $Px_i[j]$ (voir 9.3.1) est calculée pour l'onde d'étalonnage. Dans une mise en œuvre de la mesure PSQM où la transformée FFT est échelonnée par un facteur n (comme dans la routine disponible sur le marché "four1" issue des *Numerical Recipes in C* [13],

$$S_p = 6,4661e^{-06}$$

Le deuxième étalonnage fixe à la valeur de 1,0 Sone la sonie comprimée de l'onde d'étalonnage, telle que calculée au 9.4. Ce facteur d'étalonnage est calculé comme suit:

$$S_l = \frac{1}{Lx_i}$$

lorsque l'on calcule Lx_i pour l'onde d'étalonnage. Si le premier étalonnage est effectué correctement, on obtient $S_l = 240,05$.

NOTE 1 – Il y a lieu que la tonalité d'étalonnage ne soit pas filtrée par la caractéristique de réception, F, ni qu'on lui ajoute du bruit spectral de Hoth avant le calcul des facteurs Lx_i et S_l . Cette exception ne vaut que pour les opérations d'étalonnage.

Si le niveau de conversation active dans le fichier numérique n'est pas égal à -26 dBov, ou si le niveau acoustique d'écoute n'est pas égal à 78 dB, il y a lieu d'échelonner les données d'entrée en conséquence.

NOTE 2 – Dans un fichier numérique en mots de 16 éléments binaires, le niveau 0 dBov est représenté par une composante constante de 32 767. Une onde sinusoïdale ayant une amplitude zéro à crête de 32 767 aura donc un niveau efficace de $-3,01$ dBov. Avec les hypothèses adoptées dans le présent-sous-paragraphe, cette valeur correspondra à un niveau acoustique d'environ 101 dB.

9.2 Mappage temps-fréquence

Le passage du domaine temporel au domaine fréquentiel est mis en œuvre au moyen d'une transformée de Fourier à court terme insérée dans une fenêtre de Hanning produisant une représentation du temps et de la fréquence ayant une résolution constante dans le domaine temporel comme dans le domaine fréquentiel.

9.2.1 Fenêtrage

Le signal source $x_i[n]$ et le signal codé $y_i[n]$ dans la trame i sont insérés dans une fenêtre de Hanning (\sin^2):

$$\begin{aligned} xw_i[n] &= w[n] \cdot x_i[n] \\ yw_i[n] &= w[n] \cdot y_i[n] \end{aligned}$$

où $w[n]$ est la fonction de fenêtrage.

La fonction de fenêtrage peut se calculer comme suit:

$$w[n] = 0,5 \left(1 - \cos \left(\frac{2\pi n}{Nf} \right) \right) \text{ pour } 0 \leq n \leq Nf - 1$$

Dans tout le paragraphe 9, tous les calculs sont définis sur une base de trame individuelle. On doit utiliser une longueur de trame de 256 échantillons pour une fréquence de 8 kHz et de 512 échantillons pour une fréquence de 16 kHz, ce qui correspond à peu près à la longueur de fenêtre de l'oreille humaine. Par ailleurs, les trames adjacentes doivent se chevaucher d'environ 50%.

9.2.2 Densité spectrale de puissance (SPD, *spectral power density*) échantillonnée

Les densités spectrales de puissance des signaux $xw_i[n]$ et $yw_i[n]$, représentées par $Px_i[k]$ et $Py_i[k]$, sont calculées au moyen des transformées de Fourier rapides (FFT, *fast fourier transform*) suivantes:

$$\begin{aligned}xw_i[n] &\Rightarrow FFT \Rightarrow X_i[k] \\yw_i[n] &\Rightarrow FFT \Rightarrow Y_i[k] \\Px_i[k] &= (\text{Re } X_i[k])^2 + (\text{Im } X_i[k])^2 \\Py_i[k] &= (\text{Re } Y_i[k])^2 + (\text{Im } Y_i[k])^2\end{aligned}$$

9.3 Prédistorion et filtrage des fréquences

Le présent sous-paragraphe décrit d'abord la prédistorion permettant de passer de l'échelle en hertz à l'échelle des bandes critiques, pour obtenir une représentation de la densité de puissance fondamentale échantillonnée trame par trame. La densité de puissance fondamentale échantillonnée dans le signal codé est échelonnée dans chaque trame. Ensuite, les signaux, de source et codé, sont tous les deux filtrés en bande téléphonique et le bruit spectral de Hoth est ajouté pour simuler l'environnement d'écoute. Finalement, les signaux sont filtrés par la fonction de transfert caractéristique entre l'oreille externe et l'oreille interne.

9.3.1 Densité de puissance fondamentale échantillonnée

L'index de fréquence k , exprimé en hertz, est transformé en index tonal j dans le domaine des bandes critiques, au moyen d'une prédistorion de l'échelle des fréquences. L'échelle des bandes critiques est d'abord subdivisée en bandes d'intervalle égal et, pour chaque bande, on calcule une valeur (échantillon) de densité de puissance fondamentale, à partir des échantillons (habituellement multiples) de densité de puissance spectrale dans la bande correspondante sur l'échelle en hertz. Les densités de puissance fondamentale échantillonnée, $Px_i^j[j]$ et $Py_i^j[j]$ pour la bande j dans la trame i sont données par les formules suivantes:

$$\begin{aligned}Px_i^j[j] &= S_p \cdot \frac{\Delta f_j}{\Delta z} \cdot \frac{1}{I_l[j] - I_f[j] + 1} \cdot \sum_{I_f[j]}^{I_l[j]} Px_i[k] \\Py_i^j[j] &= S_p \cdot \frac{\Delta f_j}{\Delta z} \cdot \frac{1}{I_l[j] - I_f[j] + 1} \cdot \sum_{I_f[j]}^{I_l[j]} Py_i[k]\end{aligned}$$

où $I_f[j]$ est l'index du premier échantillon et $I_l[j]$ celui du dernier sur l'échelle hertzienne pour la bande j , Δf_j étant la largeur de la bande j en hertz, Δz étant la largeur de chaque sous-bande dans le domaine des bandes critiques, et S_p étant le facteur d'étalonnage de la puissance fondamentale tel qu'indiqué au 9.1.3.

9.3.2 Échelonnement local

Les signaux source et codé doivent être échelonnés dans le cadre de chaque trame, afin de compenser les variations lentes du gain. Seules les composantes fréquentielles audibles sont prises en compte (au-dessus du seuil absolu d'audibilité pour chaque bande $P_{\theta}[j]$ définie dans le Tableau 4). Les

puissances totales des signaux source et codé dans une trame i , Px_i' et Py_i' , sont calculées à l'aide de la représentation après prédistorsion fréquentielle:

$$Px_i' = \sum_{j=1}^{Nb} Px_i'[j]$$

$$Py_i' = \sum_{j=1}^{Nb} Py_i'[j]$$

où Nb est le nombre total de bandes.

Lorsque les deux puissances, celle du signal source Px_i' et celle du signal codé Py_i' se trouvent au-dessus du niveau acoustique de 40 dB, la puissance du signal codé pour la bande j , $Py_i'[j]$, est multipliée par un facteur d'échelonnement S_i :

$$Py_i''[j] = S_i \cdot Py_i'[j]$$

où:

$$S_i = \frac{Px_i'}{Py_i'}$$

Lorsque soit la puissance du signal source Px_i' soit celle du signal codé Py_i' est inférieure à 40 dBA, la puissance du signal codé pour la bande j , $Py_i'[j]$, est multipliée par un facteur d'échelonnement S_{av} qui est la moyenne de tous les facteurs S_i calculés précédemment.

9.3.3 Filtrage de la bande téléphonique

On doit filtrer les composantes $Px_i'[j]$ et $Py_i''[j]$ au moyen des caractéristiques de réception appropriées à un combiné téléphonique:

$$PFx_i[j] = F[j] \cdot Px_i'[j]$$

$$PFy_i[j] = F[j] \cdot Py_i''[j]$$

où $F[j]$ est la réponse en fréquence dans la bande j des caractéristiques de réception d'un combiné. L'UIT-T recommande d'utiliser la courbe de réception de l'appareil IRS modifié, définies dans l'Annexe D/P.830 en tant que caractéristiques de fréquence de réception d'un combiné téléphonique. Les valeurs de $F[j]$ pour ces caractéristiques sont indiquées dans le Tableau 4.

9.3.4 Bruit spectral de Hoth

En utilisation téléphonique normale, le signal vocal est perturbé par des sons ambiants issus de l'environnement de réception. Dans le cadre de la mesure PSQM, cet effet est modélisé par adjonction d'un bruit spectral de Hoth au signal de source comme au signal codé. Le bruit spectral de Hoth [8] est ajouté à la densité de puissance fondamentale échantillonnée pour chaque valeur de j , au moyen de la fonction de densité de puissance spectrale comme indiqué dans la Recommandation P.800:

$$PHx_i[j] = H[j] \cdot PFX_i[j]$$

$$PHY_i[j] = H[j] \cdot PFY_i[j]$$

où $H[j]$ est la puissance du bruit de Hoth dans la bande j indiquée dans le Tableau 4.

NOTE – Toutes les validations de la méthode PSQM par l'UIT-T ont été effectuées avec un bruit de Hoth à un niveau de 45 dBA.

9.4 Prédistorsion d'intensité

Après le calcul des densités de puissance fondamentale échantillonnée, compte tenu du filtrage en bande téléphonique et de l'insertion du bruit spectral de Hoth, l'échelle des intensités est transformée en échelle des sonies de façon à obtenir une fonction de densité de sonie comprimée et échantillonnée.

Les densités de sonie comprimée et échantillonnée, $Lx_i[j]$ et $Ly_i[j]$, sont calculées à partir des densités de puissance fondamentale $PHx_i[j]$ et $PHy_i[j]$, au moyen d'une fonction de compression donnée comme suit par Zwicker [9]:

$$Lx_i[j] = S_l \cdot \left(\frac{P_0[j]}{0,5} \right)^\gamma \cdot \left[\left(0,5 + 0,5 \cdot \frac{PHx_i[j]}{P_0[j]} \right)^\gamma - 1 \right]$$

$$Ly_i[j] = S_l \cdot \left(\frac{P_0[j]}{0,5} \right)^\gamma \cdot \left[\left(0,5 + 0,5 \cdot \frac{PHy_i[j]}{P_0[j]} \right)^\gamma - 1 \right]$$

où $P_0[j]$ est le seuil interne indiqué dans le Tableau 4 et où S_l est le facteur d'étalonnage en sonie fondamentale tel qu'exposé au 9.1.3. Les valeurs négatives de $Lx_i[j]$ et de $Ly_i[j]$ sont mises à zéro.

La valeur optimale de γ , trouvée lors d'optimisations utilisant des bases de données construites par différentes expériences d'évaluation de la qualité de la parole, est de 0,001.

Les valeurs instantanées (totales) des sonies comprimées Lx_i et Ly_i (exprimées en sonies comprimés) sont calculées par sommation des densités comprimées et échantillonnées $Lx_i[j]$ et $Ly_i[j]$:

$$Lx_i = \sum_{j=1}^{Nb} Lx_i[j] \cdot \Delta z$$

$$Ly_i = \sum_{j=1}^{Nb} Ly_i[j] \cdot \Delta z$$

où Δz est la largeur de bande dans le domaine des bandes critiques. Les valeurs instantanées des sonies comprimées, Lx_i et Ly_i , sont utilisées lors de la modélisation cognitive.

9.5 Modélisation cognitive

Dans le contexte de la mesure PSQM, toutes les opérations qui ne peuvent pas être exécutées sur le signal source seul ou sur le signal codé seul sont définies comme étant des opérations cognitives. Quatre effets cognitifs sont examinés dans le présent sous-paragraphe:

- l'échelonnement en sonie;
- la densité échantillonnée du bruit perturbateur (cognitif interne);
- le traitement des asymétries;
- le traitement du bruit perturbateur y compris les intervalles de silence.

9.5.1 Échelonnement en sonie

Dans le contexte de la mesure PSQM, la densité de sonie comprimée et échantillonnée du signal codé est échelonnée, dans chaque trame, par rapport à la sonie du signal source:

$$Ly_i'[j] = Sl_i \cdot Ly_i[j]$$

où le facteur d'échelonnement Sl_i est calculé à partir des valeurs instantanées (totales) des sonies comprimées Lx_i et Ly_i :

$$Sl_i = \frac{Lx_i}{Ly_i}$$

Lorsque Lx_i ou Ly_i a une valeur inférieure à 0,02 (en sones comprimés), la composante Sl_i est mise à égalité avec 1.

9.5.2 Densité échantillonnée du bruit perturbateur

La densité échantillonnée du bruit perturbateur $N_i[j]$ dans la bande j et dans la trame i est calculée comme étant la différence absolue entre les composantes $Lx_i[j]$ et $Ly_i[j]$:

$$N_i[j] = |Ly_i[j] - Lx_i[j]| - 0,01$$

où le facteur 0,01 (en sones comprimés) représente le bruit cognitif interne. Si, en raison de ce facteur, la valeur de $N_i[j]$ devient négative, cette valeur de $N_i[j]$ est mise à égalité avec zéro.

9.5.3 Traitement des asymétries

Lorsqu'une nouvelle composante fréquentielle est introduite dans le signal de parole, la qualité subjective se révèle plus dégradée que lorsqu'une composante de sonie égale est négligée par le codec vocal. Cette asymétrie est plus évidente lors des intervalles de silence. Le bruit qui est présent dans le signal source peut être annulé, ce qui provoque une amélioration de la qualité. S'il n'y a pas de bruit au cours des intervalles de silence contenus dans le signal de source, toute différence entre signaux vocaux de source et codés conduira à une dégradation de la qualité.

En outre, lorsqu'une composante fréquentielle est négligée dans le signal de source (non codé par le codec), le signal restant reste un événement acoustique cohérent. Si une nouvelle composante fréquentielle indépendante (distorsion) est insérée dans le signal codé, le nouveau signal ainsi formé peut être décomposé en deux parties: le signal original et la distorsion. Cette décomposition du flux acoustique rend le bruit plus gênant.

L'effet d'asymétrie est quantifié par $C_i[j]$ et pris en compte dans le bruit perturbateur présent dans la trame i , N_i :

$$N_i = \sum_{j=1}^{Nb} N_i[j] \cdot C_i[j] \cdot \Delta z$$

où:

$$C_i[j] = \left(\frac{PHy_i[j] + 1}{PHx_i[j] + 1} \right)^{0,2}$$

avec $PHx_i[j]$ et $PHy_i[j]$ comme puissances des signaux source et codé (après filtrage par le système IRS et l'addition du bruit de Hoth), respectivement dans la trame i et dans la bande j . Lorsque $PHx_i[j]$ et $PHy_i[j]$ ont une valeur inférieure à 20 dB au-dessus du seuil absolu d'audibilité dans la bande j (c'est-à-dire si $PHx_i[j]$ et $PHy_i[j]$ ont une valeur inférieure à $100 * P_0[j]$), la composante $C_i[j]$ est mise à égalité avec 1. La valeur maximale de $C_i[j]$ a une limite supérieure de 2,0.

9.5.4 Traitement du bruit perturbateur, y compris les intervalles de silence

Dans la méthode PSQM, les intervalles de silence sont pris en compte au moyen d'un facteur de pondération, W_{sil} , qui dépend du contexte des expériences subjectives. Les trames de silence sont définies comme étant celles pour lesquelles le signal de source a une puissance totale de Px_i

(c'est-à-dire $\sum_j P_{x_i'}[j]$) au-dessous du niveau acoustique de 70 dB. Si le facteur d'étalonnage global, S_p , a été calculé correctement, le seuil de silence est $P_{x_i'} = 1,0 \cdot 10^7$. Les trames avec une puissance $P_{x_i'}$ au-dessous de cette valeur sont considérées comme étant silencieuses.

Les sonies moyennes du bruit, N_{spav} et N_{silav} , peuvent maintenant être calculées sur les trames de conversation active et sur les trames de silence, respectivement:

$$N_{spav} = \frac{1}{M_{sp}} \sum_{i \text{ pour les trames de conversation active}} N_i$$

$$N_{silav} = \frac{1}{M_{sil}} \sum_{i \text{ pour les trames de silence}} N_i$$

où M_{sp} est le nombre de trames de conversation actives et où M_{sil} est le nombre de trames de silence.

L'influence des intervalles de silence dépend directement de la longueur de ces intervalles. Si le signal vocal de source ne contient pas d'intervalles de silence, l'influence est nulle. Si le signal vocal de source contient un certain pourcentage de trames de silence, l'influence est proportionnelle à ce pourcentage. En utilisant un ensemble de conditions aux limites triviales, on peut montrer que la pondération correcte est la suivante:

$$N_{w_{sil}} = \frac{W_{sp} \cdot P_{sp}}{W_{sp} \cdot P_{sp} + P_{sil}} \cdot N_{spav} + \frac{P_{sil}}{W_{sp} \cdot P_{sp} + P_{sil}} \cdot N_{silav}$$

où p_{sil} est la fraction de trames de silence, p_{sp} la fraction de trames de conversation actives ($p_{sil} + p_{sp} = 1,0$), où W_{sil} est le facteur de pondération sur les intervalles de silence, où $W_{sp} = \frac{1 - W_{sil}}{W_{sil}}$, et où $N_{w_{sil}}$ est le bruit perturbateur corrigé par le facteur de pondération W_{sil} pour l'intervalle de silence.

Ce bruit perturbateur $N_{w_{sil}}$, appelé valeur PSQM dans les sous-paragraphes suivants, peut être utilisé pour prédire la qualité de parole perçue subjectivement, obtenue par la méthode d'évaluation par catégories absolues (ACR, *absolute category rating*) avec l'échelle de qualité d'écoute.

NOTE 1 – La valeur de la pondération $N_{w_{sil}}$ doit avoir une limite supérieure de 6,5.

NOTE 2 – Pour le matériel vocal comportant de longues périodes de silence, la pondération est différente de celle du matériel vocal comportant seulement de courtes périodes de silence. Par ailleurs, le bruit affectant l'enregistrement du matériel de source exerce également une influence sur la pondération de l'intervalle de silence. Pour le matériel vocal ne comportant pas d'intervalles de silence, la pondération n'est pas applicable et le terme $N_{w_{sil}}$ devient égal à N_{spav} . Un certain nombre de bases de données vocales ont été examinées afin de déterminer la pondération optimale sur les intervalles de silence. Ces bases de données se composaient de matériel vocal avec environ 50% d'intervalles de silence. La pondération optimale qui a été trouvée variait entre 0,0 et 0,5: [10], [11] et [12]. La détermination de la valeur de W_{sil} pour les paroles avec intervalles de silence est encore à l'étude. Un facteur de pondération de 0,2 est provisoirement recommandé pour le matériel vocal comportant environ 50% d'intervalles de silence.

10 Transformation de l'échelle de qualité objective en échelle de qualité subjective

La sortie de l'algorithme décrit au paragraphe 9, appelé valeur PSQM, indique le degré de dégradation de la qualité subjective en raison du codage de la parole. Lorsque l'estimation de la qualité subjective sur une échelle spécifique n'est pas nécessaire, par exemple lors de l'optimisation

des paramètres d'un codec ou lors de la simple comparaison des performances de codecs, la valeur de mesure PSQM est donc très utile par elle-même. Pour estimer la qualité subjective sur des échelles d'évaluation telles que les notes moyennes d'opinion (MOS, *mean opinion score*) et notes Q équivalentes, la mesure PSQM est cependant transformée comme décrit ci-dessous.

10.1 Notes moyennes d'opinion

Lors de l'évaluation subjective de la performance des codecs, la méthode d'évaluation ACR, faisant appel à l'échelle de qualité d'écoute spécifiée dans la Recommandation P.800 est souvent utilisée; elle exprime la qualité subjective en notes MOS. Etant donné que la relation entre notes MOS et mesures PSQM n'est pas nécessairement la même selon les différentes langues, il est difficile de déterminer une fonction unique permettant de transformer la valeur PSQM en valeur MOS estimée. En pratique, il est donc nécessaire de calculer d'avance de telles fonctions de transformation pour chaque langue et pour chaque essai subjectif particulier.

NOTE – La valeur absolue de la note MOS dépend du contexte de l'expérience subjective. La note MOS estimée qui est obtenue par une fonction de transformation prédéterminée prédit la qualité subjective perçue lors de l'expérience subjective avec un contexte équivalent à celui qui a été utilisé pour calculer la fonction de transformation.

Lorsque les résultats sont présentés dans le domaine des notes MOS estimées, il convient d'inclure dans le rapport la fonction de transformation entre la valeur PSQM et la valeur MOS.

10.2 Notes Q équivalentes

Il est difficile de comparer les notes MOS obtenues lors de différentes expériences subjectives car le jugement subjectif est affecté par les conditions d'expérience, par exemple l'étendue de l'échelle de qualité vocale utilisé dans l'expérience. Les valeurs Q équivalentes sont donc utilisées parfois comme échelle de qualité subjective. Ce sont les valeurs Q de l'appareil MNRU qui est défini dans la Recommandation P.810, pour lequel les notes MOS sont équivalentes à celles du signal vocal codé.

Dans la mesure objective, la valeur Q équivalente peut être estimée directement à partir des valeurs de la mesure PSQM pour le signal vocal codé et les conditions de l'appareil MNRU, sans transformation de la valeur PSQM en note MOS (voir la Figure 4). Lorsque les résultats sont présentés dans le domaine des valeurs Q équivalentes estimées, il y a lieu d'inclure dans le rapport la courbe des valeurs Q en fonction des valeurs PSQM, illustrée dans la Figure 4.

NOTE – Les valeurs Q équivalentes deviennent relativement peu fiables dans les régions haute et basse de leur répartition, parce que la courbe des valeurs Q en fonction des mesures PSQM devient presque plate dans ces régions. En conséquence, il convient de prendre des précautions lorsque l'on travaille dans le domaine Q avec des signaux vocaux de très haute ou de très basse qualité.

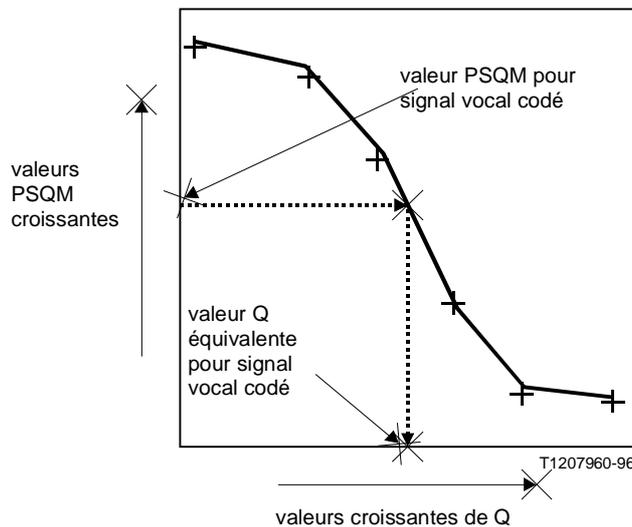


Figure 4/P.861 – Détermination de la valeur Q équivalente du signal vocal codé

11 Analyse des résultats

L'analyse des résultats des mesures objectives doit être effectuée sur la base des valeurs PSQM, des notes MOS estimées ou des valeur Q équivalentes estimées.

Pour chaque condition d'essai, il convient de calculer et de rapporter séparément les notes moyennes correspondant aux locuteurs et aux locutrices, ainsi que leur moyenne générale.

Le calcul d'un écart type distinct pour chaque condition d'essai n'est pas recommandé. Les limites de confiances doivent être évaluées par prise en compte de la variation de qualité objective selon les locuteurs, selon les phrases et selon les tests de signification effectués par des techniques conventionnelles d'analyse de variance.

NOTE – L'analyse statistique décrite ici est différente de celle qui est effectuée lors des évaluations subjectives où les moyennes de qualité subjective sont statistiquement évaluées par prise en compte des variations selon les sujets, les locuteurs et les phrases. Etant donné que la mesure PSQM ne peut pas donner une estimation des distributions des votes subjectifs mais qu'elle n'indique que leur valeur moyenne, il est impossible d'effectuer l'analyse selon les sujets. L'estimation des courbes de répartition des votes subjectifs est encore à l'étude. Lorsqu'une analyse selon les sujets sera nécessaire, il conviendra donc d'effectuer des expériences subjectives conformes à la Recommandation P.830.

Bibliographie

- [1] NTT: Transmission performance objective evaluation model for fundamental factors, *CCITT Contribution COM XII-174*, novembre 1983.
- [2] LALOU (J.): L'indice d'information: mesure objective de la qualité de transmission de la parole, *Annales des Télécommunications*, Volume 45, n° 1-2, p. 47-65, CNET/France, 1990.
- [3] BNR: Evaluation of non-linear distortion via the coherence function, *CCITT Contribution COM XII-60*, avril 1982.
- [4] KUBICHEK (R.F.), QUINCY (E.A.), KISER (K.L.): Speech Quality Assessment Using Expert Pattern Recognition Techniques, *Proceedings of the IEEE Pacific Rim Conference on Computers, Communication, and Signal Processing*, juin 1989.

- [5] Royal PTT, Netherlands: Measuring the quality of audio devices, *CCITT Contribution COM XII-114*, Genève, décembre 1991.
- [6] BEERENDS (J.G.), STEMERDINK (J.A.): A Perceptual Speech-Quality Measure Based on a Psychoacoustic Sound Representation, *J. Audio Eng. Soc.*, Vol. 42, n° 3, p. 115-123, mars 1994.
- [7] BEERENDS (J.G.): Modelling Cognitive Effects that Play a Role in the Perception of Speech Quality, *Speech Quality Assessment*, Workshop papers, Bochum, p. 1-9, novembre 1994.
- [8] Supplément n° 13 du CCITT aux Recommandations de la série P, *Spectres de bruit*, paragraphe 2, *Livre bleu*, tome V, UIT, Genève, 1988.
- [9] ZWICKER (Feldtkeller): *Das Ohr als Nachrichtenempfänger*, S. Hirzel Verlag, Stuttgart, 1967.
- [10] Royal PTT, The Netherlands: Correlation of a perceptual quality speech measure with the subjective quality of the CCITT LD-CELP (G.728) speech codec, *ITU-T Contribution COM 12-10*, Genève, mars 1993.
- [11] Royal PTT, The Netherlands: Correlation between the PSQM and the subjective results of ITU-T 8 kbit/s 1993 speech codec test, *ITU-T Contribution COM 12-31*, Genève, septembre 1994.
- [12] NTT: Review of validation tests for objective speech quality measures, *ITU-T Contribution COM 12-74*, Genève, mai 1996.
- [13] PRESS (W.H.) *et al.*: Numerical Recipes in C, The Art of Scientific Computing, *Cambridge University Press*, Cambridge, Massachusetts, 1988.

APPENDICE I

Contenu de la disquette accompagnant la Recommandation P.861

I.1 Introduction

La présente Recommandation comporte une disquette contenant des vecteurs de test et des valeurs calculées détaillées destinées à aider les lecteurs à réaliser leurs propres implémentations PSQM. Le présent appendice en décrit les fichiers.

I.2 Répertoire\test

Ce répertoire contient des fichiers servant aux tests de précision d'une mise en œuvre de la mesure PSQM. Les fichiers contenus dans ce répertoire sont les suivants:

<code>longs.cod</code>	<code>longs.src</code>	<code>outlong.txt</code>
<code>outshort.txt</code>	<code>shorts.cod</code>	<code>shorts.src</code>

Le contenu de ces fichiers est décrit ci-dessous.

<code>longs.cod</code>	Fichier de signaux vocaux codés qui est destiné à servir lors de l'étalonnage de la mesure PSQM. Octet de poids faible enregistré en premier et retard de 22 échantillons par rapport au fichier source de signaux vocaux.
<code>longs.src</code>	Fichier source de signaux vocaux utilisé pour créer le fichier <code>longs.cod</code> . Octet LSB enregistré en premier.

Outlong.txt	<p>Signal de sortie trame par trame pour le vecteur d'essai de longue durée. En plus des variables d'étalonnage (S_p et S_1), ce fichier contient les informations suivantes:</p> <p>facteur de normalisation global calculé (S_{global});</p> <p>points de départ et d'arrêt pour les deux fichiers de signaux vocaux, source et codé;</p> <p>bruit perturbateur trame par trame (N_i);</p> <p>indicateur trame par trame de trame silencieuse (1=silence, 0 bruit).</p>
Outshort.txt	<p>Valeurs pas à pas des variables pour le vecteur d'essai de courte durée. Ce fichier nécessite la configuration des valeurs suivantes dans le programme:</p> <p>retard = 0</p> <p>point de départ = 0</p> <p>point d'arrêt = 511</p> <p>facteur de normalisation $S_{global} = 1,0$</p>
<p>Les valeurs intermédiaires suivantes sont fournies dans le fichier:</p>	
	<p>séquence d'entrée ($x_i[n]$, $y_i[n]$)</p> <p>version fenêtrée de la séquence d'entrée ($xw_i[n]$, $yw_i[n]$)</p> <p>échantillons de densité de puissance spectrale ($Px_i[k]$, $Py_i[k]$)</p> <p>facteur de normalisation locale (S_i)</p> <p>échantillons de densité de puissance fondamentale ($Px'_i[j]$, $Py'_i[j]$)</p> <p>résultats du filtrage en bande téléphonique ($PFx_i[j]$, $PFy_i[j]$)</p> <p>résultats d'injection du bruit de Hoth ($PHx_i[j]$, $PHY_i[j]$)</p> <p>échantillons de densité de sonie comprimée ($Lx_i[j]$, $Ly_i[j]$)</p> <p>facteur de normalisation locale de la sonie (Sl_i)</p> <p>échantillons de densité du bruit perturbateur ($N_i[j]$)</p> <p>facteur d'influence de l'asymétrie ($C_i[j]$)</p> <p>bruit perturbateur (N_i)</p>
shorts.cod	<p>Fichier de signaux vocaux codés destinés à servir lors de l'étalonnage de la mesure PSQM. Octet LSB en premier et retardé de 0 échantillon par rapport au fichier vocal source.</p>
shorts.src	<p>Fichier vocal source utilisé pour créer le fichier shorts.cod. Octet LSB en premier.</p>

Mesure objective de la qualité des codecs vocaux fonctionnant en bande téléphonique (300-3400 Hz) au moyen de modules de normalisation de mesure (MNB, *measuring normalizing blocks*)

II.1 Introduction

L'UIT-T étudie des méthodes objectives de mesure de la qualité des paroles codées. Les performances de plusieurs mesures objectives de la qualité des codecs vocaux fonctionnant en bande téléphonique (300-3400 Hz) ont été comparées au cours de la période d'études 1993-1996. Les résultats des essais de validation ont permis de conclure que la méthode dite méthode de la mesure de la qualité de parole perçue (PSQM, *perceptual speech quality measure*) était celle qui évaluait le mieux la qualité subjective indépendamment des langues utilisées, des locuteurs ou des codecs. Il a également été démontré qu'en moyenne l'erreur d'estimation existant avec la méthode PSQM était comparable à la fiabilité statistique des données subjectives. La Recommandation P.861 fondée sur l'algorithme PSQM a été élaborée à la suite de cette étude.

La mesure objective de la qualité décrite dans le corps de la présente Recommandation est un outil puissant permettant d'estimer la qualité subjective des codecs vocaux. Toutefois, ainsi qu'il est décrit dans la section Domaine d'application de la présente Recommandation (voir le Tableau 1), les facteurs de qualité susceptibles d'être évalués selon la Recommandation P.861 sont limités.

Pour étendre le domaine d'application de la mesure objective de la qualité des paroles codées, l'UIT-T met actuellement au point une méthode de mesure objective de la qualité qui peut s'appliquer à l'évaluation de la parole dégradée par des erreurs dans la voie de transmission, notamment les erreurs sur les bits dans les réseaux mobiles, la perte de cellules dans les réseaux ATM, la perte de paquets en matière de téléphonie Internet, etc.

Le présent appendice décrit une méthode de mesure objective de la qualité dont l'algorithme détaillé est actuellement disponible pour l'UIT-T et devrait s'appliquer à l'évaluation des erreurs sur les bits et des effacements de trames. Il est également prévu que l'algorithme décrit dans le présent appendice étendra le domaine d'application de la mesure objective de la qualité du point de vue des types de codecs à traiter (Tableau II.1). Il convient de noter que les facteurs expérimentaux relevant du domaine d'application de la présente Recommandation devraient être évalués par la Recommandation P.861 même.

Certes, on peut trouver dans les documents de référence [1], [2] et [3] des renseignements sur les performances de la mesure objective décrite dans le présent appendice, mais d'autres informations sont nécessaires. Il convient en particulier d'examiner l'applicabilité de la mesure à d'autres types de dégradation de voie tels que la perte de cellules/paquets par rapport à d'autres mesures objectives de la qualité qui ont également été proposées dans les documents de référence [4], [5] et [6] de l'UIT-T. C'est pour cette raison que l'UIT-T ne recommande pour le moment aucun nouvel algorithme afin d'étendre le domaine d'application de la présente Recommandation mais donne un exemple de mesure objective de la qualité dont la portée peut être plus étendue que celle de la Recommandation P.861.

L'UIT-T continuera à comparer/valider les performances des mesures objectives de la qualité, y compris en ce qui concerne la méthode décrite dans le présent appendice. Il en découlera l'élaboration d'une nouvelle Recommandation applicable à l'évaluation des erreurs dans la voie de transmission et qui élargira le domaine d'application de la présente Recommandation.

II.2 Calcul de la mesure objective

Le présent sous-paragraphe décrit le calcul d'une mesure objective, qui est fondé sur des modules de normalisation de mesure (MNB, *measuring normalizing blocks*). Les modules MNB ont été définis en réponse à des observations établissant que les auditeurs s'adaptent et réagissent différemment aux déviations spectrales couvrant différentes échelles de temps et de fréquence. Ainsi, pour l'application estimation de la qualité de la parole, il est nécessaire d'effectuer une série d'analyses à différentes échelles de fréquence et de temps pour assurer une cohérence perceptive maximale dans une gamme étendue de types de distorsion. Au fur et à mesure que les déviations spectrales sont mesurées, il est nécessaire d'éliminer les déviations observées à une échelle donnée de manière à ne pas les compter à nouveau dans les déviations se produisant à d'autres échelles. On observe également que l'utilisation d'échelles décroissantes permet très probablement d'émuler les schémas d'adaptation et de réaction des auditeurs vis-à-vis des déviations spectrales. Cette observation a entraîné la définition d'une structure hiérarchique des modules MNB.

Tableau II.1/P.861 – Relation entre techniques de codage, facteurs expérimentaux et applications selon la présente Recommandation

Facteurs expérimentaux	Note
niveaux d'entrée des signaux de parole dans un codec	1
niveaux d'écoute lors d'expériences subjectives	3
variations selon le locuteur	1
locuteurs multiples simultanés	3
erreurs sur les bits et effacements de trames	2
perte de cellules et de paquets	3
débits lorsque le codec peut fonctionner selon plusieurs modes	1
transcodages	1
discordance de débit entre un codeur et un décodeur si un codec possède plusieurs modes de débit	3
bruit ambiant du côté émission	3
signaux d'information de couche Réseau à l'entrée d'un codec	3
signaux de musique à l'entrée d'un codec	3
temps de propagation	4
prédistorsion à court terme du signal audio	3
prédistorsion à long terme du signal audio	5
écrêtage temporel du signal vocal	3
écrêtage en amplitude du signal vocal	3
techniques de codage	
codecs temporels	1
codecs CELP et hybrides ≥ 4 kbit/s	1
codecs CELP et hybrides < 4 kbit/s	2
VOCODEURS	2
autres codeurs	2

Tableau II.1/P.861 – Relation entre techniques de codage, facteurs expérimentaux et applications selon la présente Recommandation (fin)

applications	Note
optimisation du codeur	1
évaluation du codeur	1
sélection du codeur	3
planification du réseau	6
essais actifs dans le réseau	7
dispositifs de mesure sans intrusion en service	4
<p>NOTE 1 – Si les conditions expérimentales ne comprennent que les éléments marqués 1, il convient d'utiliser l'algorithme décrit dans le corps de la présente Recommandation.</p> <p>NOTE 2 – Si les conditions expérimentales ne comprennent que les éléments marqués 2, ou une combinaison d'éléments marqués 1 et 2, l'algorithme décrit dans le présent appendice devrait donner des résultats satisfaisants.</p> <p>NOTE 3 – On ne dispose pas d'assez de renseignements au sujet de la précision de la mesure objective en ce qui concerne cette variable.</p> <p>NOTE 4 – Les mesures objectives sont réputées fournir des prédictions inexactes lorsqu'elles sont utilisées conjointement avec cette variable ou ne sont pas censées être utilisées avec cette variable.</p> <p>NOTE 5 – La mesure objective décrite dans le corps de la présente Recommandation est réputée fournir des prédictions inexactes en présence d'un dérapage assez important (plus de 10% de la longueur de trame). La possibilité d'appliquer la mesure décrite dans le présent appendice fera l'objet d'un complément d'étude, ainsi que la possibilité d'appliquer les deux mesures objectives en présence d'un dérapage peu important.</p> <p>NOTE 6 – Au prix de certaines précautions, on peut utiliser les mesures objectives pour certains objets de planification de réseau. Le lecteur doit prendre acte du fait qu'il existe d'importants facteurs de planification de réseau auxquels la présente Recommandation n'est pas applicable (voir la section "Facteurs expérimentaux" du présent tableau).</p> <p>NOTE 7 – Au prix de certaines précautions, on peut utiliser la méthode objective pour certains essais actifs sur le réseau. Le lecteur doit prendre acte du fait qu'il peut exister, dans une chaîne de connexion établie dans un réseau actif, des facteurs ou des techniques auxquels la présente Recommandation n'est pas applicable (voir les sections "Facteurs expérimentaux" et "Techniques de codage" du présent tableau).</p>	

Deux types de modules de normalisation de mesures sont examinés ici. Le premier est le module de normalisation de mesure temporelle (TMNB, *time measuring normalizing block*) et le second est le module de normalisation de mesure fréquentielle (FMNB, *frequency measuring normalizing block*). Chaque module prend en entrée la référence transformée au niveau perceptif [R(t,f)] et les signaux d'essai [T(t,f)] et les restitue en sortie accompagnés d'un ensemble de mesures. Ces deux modules sont définis par les Figures II.1 et II.2 respectivement. Le module TMNB est intégré dans une échelle de fréquence donnée, puis mesure les différences et normalise le signal d'essai à différents instants. Enfin, les parties positive et négative des mesures sont intégrées dans le temps. La réciproque est vraie pour un module FMNB. Celui-ci est intégré dans une certaine échelle de temps, puis mesure les différences et normalise le signal d'essai à différentes fréquences. Enfin, les parties positive et négative des mesures sont intégrées en fréquence. Par définition, les deux types de modules MNB sont équipotents. Cette importante caractéristique est illustrée à la Figure II.3 et signifie simplement qu'un deuxième passage à travers un module MNB donné n'altérera pas davantage le signal d'essai et donnera en résultat un vecteur de mesure composé de valeurs nulles. Cette équipotence des modules MNB permet de les mettre en cascade, ce qui ne les empêche pas de mesurer la déviation à une échelle de temps ou de fréquence donnée une fois et seulement une fois.

Pour mesurer les déviations spectrales à différentes échelles de temps et de fréquence, on a défini une structure hiérarchique des modules TMNB et FMNB fonctionnant à des échelles décroissantes (Figure II.4). Lorsqu'elle est utilisée comme mesure de distance conjointement avec une transformation perceptive (décrite ci-après), cette structure semble émuler avec efficacité les schémas d'adaptation et de réaction des auditeurs vis-à-vis des déviations spectrales. La structure indiquée donne 12 mesures. En raison de la nature hiérarchique de ces structures, les mesures provenant d'une couche autre que la couche supérieure ne signifient pas grand chose séparément, mais il a été établi qu'une combinaison linéaire de mesures était un bon indicateur de la distance perceptive entre les deux signaux. La valeur qui découle de cette combinaison linéaire est appelée distance acoustique (AD, *auditory distance*):

$$AD = \sum_{i=1}^N \text{weight}_i \cdot m_i .$$

La distance acoustique est une quantité positive. Lorsque les signaux de référence et d'essai sont analogues, la distance AD est faible. Celle-ci augmente à mesure que les signaux de référence et d'essai s'éloignent l'un de l'autre au niveau perceptif. Une fonction logistique ou une autre "fonction de limitation" peut être utilisée pour le mappage de la distance AD dans un intervalle fini. La distance AD présente ainsi la meilleure corrélation avec les évaluations de la qualité ou de la dégradation subjectives, qui couvrent généralement une gamme finie. Des facteurs de pondération weight_i ont été choisis pour maximiser cette corrélation.

NOTE – Le présent appendice contient suffisamment d'informations pour mettre en œuvre les algorithmes dans un langage de programmation. De nouvelles mises en œuvre peuvent être validées au moyen des renseignements figurant dans le fichier <ftp://ftp.its.blrdoc.gov/dist/voice/verify.zip>.

II.2.1 Spécifications des entrées et des sorties

Les données d'entrée de l'algorithme sont deux fichiers de parole appelés fichier de référence et fichier d'essai. Le fichier de référence contient une représentation numérique du signal de référence, qui est généralement la donnée d'entrée du dispositif à l'essai (DUT, *device under test*) et qui est désigné par x dans les équations. Le fichier d'essai contient une représentation numérique du signal d'essai, qui est généralement la donnée de sortie du dispositif à l'essai et est désigné par y dans les équations. Le taux d'échantillonnage est de 8000 échantillons par seconde et il est recommandé d'observer une précision d'au moins 16 bits par échantillon. Il est possible d'utiliser une précision moins élevée si l'utilisateur accepte la perte de sensibilité qui y est liée. En outre, on peut appliquer une précision et des taux d'échantillonnage plus élevés (par exemple 16 000 échantillons par seconde) à condition de veiller à ce que la transformation en domaine fréquentiel produise les résultats indiqués au II.2.4. De même, si des fichiers à taux d'échantillonnage plus élevés sont disponibles, il est possible de définir un échantillonnage moins fréquent à l'aide des programmes disponibles dans la Bibliothèque logicielle de l'UIT-T et publiés dans la Recommandation G.191 (1996). Les fichiers d'entrée doivent contenir au moins une seconde de communication vocale en bande téléphonique. (Les fichiers qui contiennent uniquement des pauses en conversation naturelle ne sont pas utiles.) La durée des fichiers utilisés dans la mise au point des algorithmes est comprise entre 3 et 9 secondes.

L'algorithme produit une valeur de sortie unique non négative appelée distance acoustique (AD). La distance AD est une estimation de la distance perceptive entre les signaux de référence et d'essai. Ainsi, lorsque le dispositif à l'essai et la configuration d'essai sont transparents, les signaux de référence et d'essai seront identiques et la distance AD sera nulle. A mesure que le dispositif à l'essai introduit de plus en plus de distorsion, les signaux de référence et d'essai s'éloigneront l'un de l'autre au niveau perceptif et la distance AD augmentera.

II.2.2 Temps de propagation

On suppose que les deux fichiers sont de même longueur et sont synchronisés. Autrement dit, tout temps de propagation dans le dispositif à l'essai ou dans la configuration d'essai a été supprimé. Si ces temps de propagation sont connus a priori, ils peuvent être supprimés au moyen d'une temporisation appropriée pendant l'acquisition de données. Si ces temps de propagation ne sont pas connus a priori, ils peuvent être supprimés à l'aide de la technique décrite à la Section 7 de la norme ANSI T1 sur le temps de propagation, la synchronisation et la mesure de la fréquence de trame des communications multimédias (Norme ANSI T1.801.04-1997), puis supprimés par modification d'un fichier ou des deux.

II.2.3 Préparation des signaux

Les données de référence sont introduites dans le vecteur x et les données d'essai dans le vecteur y . La valeur moyenne est ensuite supprimée de chacune des $N1$ entrées effectuées dans chaque vecteur:

$$x(i) = x(i) - \frac{1}{N1} \cdot \sum_{j=1}^{N1} x(j), \quad y(i) = y(i) - \frac{1}{N1} \cdot \sum_{j=1}^{N1} y(j), \quad 1 \leq i \leq N1.$$

NOTE – On utilise la notation *Matrice (i) = Matrice (i) +/- /·/÷ Facteur de normalisation* dans l'ensemble du présent appendice. Bien que non appropriée du point de vue mathématique, elle indique comment cet algorithme pourrait être codé. La matrice initiale est modifiée (normalisée) de la manière indiquée et la valeur modifiée est utilisée dans les calculs ultérieurs.

Ce processus supprime toute composante continue pouvant être présente dans les signaux d'essai et de référence. (La composante continue d'un signal est inaudible.)

Ensuite, chaque vecteur est normalisé à un niveau commun de valeur quadratique moyenne:

$$x(i) = x(i) \cdot \left[\frac{1}{N1} \sum_{j=1}^{N1} x(j)^2 \right]^{-1/2}, \quad y(i) = y(i) \cdot \left[\frac{1}{N1} \sum_{j=1}^{N1} y(j)^2 \right]^{-1/2}, \quad 1 \leq i \leq N1$$

Ce processus supprime à peu près tout gain fixe dans le dispositif à l'essai ou dans la configuration d'essai. Un gain fixe n'influera donc pas sur les valeurs de la distance AD produites par cet algorithme.

II.2.4 Transformation en domaine fréquentiel

Les signaux sont ensuite transformés en domaine fréquentiel au moyen de la transformée de Fourier rapide (FFT). La taille de trame est de 16 ms (128 échantillons pour des signaux vocaux échantillonnés à 8000 Hz), et le chevauchement de trame est de 50%. Tout échantillon se trouvant au-delà de la dernière trame entière est ignoré. Chaque trame d'échantillons est multipliée (échantillon par échantillon) par la fenêtre Hamming de longueur 128:

$$w(i) = 0,54 - 0,46 \cos\left(\frac{2\pi(i-1)}{127}\right), \quad 1 \leq i \leq 128.$$

Après multiplication par la fenêtre de Hamming, chaque trame est transformée en vecteur de domaine fréquentiel de 128 points au moyen de la transformée FFT. Pour chaque trame, le carré de l'effectif des échantillons de fréquence 1 à 65 (du niveau DC au niveau de Nyquist) est retenu. Les résultats sont mémorisés dans les matrices X et Y . Ces matrices contiennent 65 lignes et $N2$ colonnes, $N2$ étant le nombre de trames extraites des $N1$ échantillons initiaux contenus dans x et y .

Il convient de noter que dans ces matrices, les lignes représentent l'axe des fréquences (indexée par i dans la description algorithmique) et les colonnes l'axe des temps (indexée par j).

L'échelonnement de la transformée FFT n'étant pas normalisé, il convient de veiller à ce que la transformée FFT utilisée dans cet algorithme soit échelonnée de manière à satisfaire à la condition suivante: lorsqu'une trame de 128 échantillons à valeur réelle égale à 1 est introduite dans la transformée FFT sans fenêtrage, la valeur complexe du segment DC résultant de la transformée FFT doit être $128 + 0 \cdot j$.

II.2.5 Sélection de trame

Seules les trames qui correspondent aux seuils d'énergie ou qui les dépassent dans X et Y sont utilisées dans le calcul de la distance AD. Pour X, ce seuil d'énergie est fixé à 15 dB au-dessous de l'énergie de la trame de crête dans X:

$$xenergy(j) = \sum_{i=1}^{65} X(i, j), \quad xthreshold = 10^{\frac{-15}{10}} \cdot \max_j(xenergy(j)).$$

Pour Y, le seuil d'énergie est fixé à 35 dB au-dessous de l'énergie de la trame de crête dans Y:

$$yenergy(j) = \sum_{i=1}^{65} Y(i, j), \quad ythreshold = 10^{\frac{-35}{10}} \cdot \max_j(yenergy(j)).$$

Les trames qui correspondent à ces deux seuils d'énergie ou qui les dépassent sont retenues:

$$\{xenergy(j) \geq xthreshold\} \text{ AND } \{yenergy(j) \geq ythreshold\} \Rightarrow \text{la trame } j \text{ est retenue.}$$

Si une trame contient un ou plusieurs échantillons égaux à zéro, elle est supprimée de X et de Y. Ces matrices contiennent maintenant 65 lignes et N3 colonnes, N3 étant le nombre de trames qui ont été retenues. Si $N3 = 0$, les fichiers d'entrée ne contiennent pas de signaux appropriés et l'algorithme se termine.

Il convient de noter que les niveaux de seuil diffèrent pour X et Y. Cela permet d'inclure dans Y des trames dont la puissance est réduite en raison d'un artefact significatif au niveau perceptif (par exemple un écrêtage temporel) mais qui sont toujours suffisamment puissantes pour être comparées avec succès au signal vocal initial.

II.2.6 Approximation de la sonie perçue

Chacun des échantillons du domaine fréquentiel dans X et Y est maintenant logarithmiquement transformé et donne une approximation de la sonie perçue:

$$X(i, j) = 10 \cdot \log_{10}(X(i, j)), \quad Y(i, j) = 10 \cdot \log_{10}(Y(i, j)), \quad 1 \leq i \leq 65, 1 \leq j \leq N3.$$

II.2.7 Module de normalisation de mesure fréquentielle (FMNB)

Un module FMNB est appliqué à X et à Y à l'échelle de temps la plus longue qui est disponible et qui est définie par la longueur des fichiers d'entrée. Quatre mesures sont extraites et enregistrées dans le vecteur de mesure m. Ces mesures couvrent les limites inférieure et supérieure de la bande du signal vocal en bande téléphonique (là où il est le plus aisé de détecter le bruit et les changements de réponse en fréquence). Les vecteurs temporaires f1, f2 et f3 sont utilisés:

$$f1(i) = \frac{1}{N3} \sum_{j=1}^{N3} Y(i, j) - \frac{1}{N3} \sum_{j=1}^{N3} X(i, j), 1 \leq i \leq 65, \text{ Mesure}$$

$$f2(i) = f1(i) - f1(17), 1 \leq i \leq 65, \text{ Normalisation de la mesure à 1 kHz}$$

$$Y(i, j) = Y(i, j) - f2(i), 1 \leq i \leq 65, 1 \leq j \leq N3, \text{ Normalisation de Y}$$

$$f3(i) = \frac{1}{4} \sum_{j=1}^4 f2(1 + 4 \cdot (i - 1) + j), 1 \leq i \leq 16, \text{ Lissage de la mesure}$$

$$m(1)=f3(1), m(2)=f3(2), m(3) = f3(13), m(4), \text{ Sauvegarde de 4 mesures}$$

II.2.8 Calcul des modules de normalisation de mesure temporelle

Dans la structure MNB, la partie médiane de la bande est assujettie à deux niveaux additionnels de découpage de bande binaire, ce qui donne des bandes larges d'environ 2 à 3 Bark. Les parties supérieure et inférieure de la bande sont traitées chacune une fois par un module TMNB séparé. Enfin, une mesure résiduelle est effectuée. La structure compte au total 9 modules TMNB, une représentation graphique étant donnée à la Figure II.4. La structure MNB produit 8 mesures qui viennent s'ajouter à celles déjà produites par le module FMNB initial (décrit au II.2.7 du présent appendice). Les variables temporaires t0, t1, ... t9 sont utilisées.

TMNB-0 (partie inférieure de la bande, large de 1,9 Bark):

$$t0(j) = \frac{1}{5} \sum_{i=2}^6 Y(i, j) - \frac{1}{5} \sum_{i=2}^6 X(i, j), 1 \leq j \leq N3, \text{ Mesure}$$

$$Y(i, j) = Y(i, j) - t0(j), 2 \leq i \leq 6, 1 \leq j \leq N3, \text{ Normalisation de Y}$$

$$m(5) = \frac{1}{N3} \sum_{j=1}^{N3} \max(t0(j), 0), \text{ Sauvegarde de la partie positive de la mesure}$$

TMNB-1 (milieu de la bande, couche supérieure, large de 10 Bark):

$$t1(j) = \frac{1}{36} \sum_{i=7}^{42} Y(i, j) - \frac{1}{36} \sum_{i=7}^{42} X(i, j), 1 \leq j \leq N3, \text{ Mesure}$$

$$Y(i, j) = Y(i, j) - t1(j), 7 \leq i \leq 42, 1 \leq j \leq N3, \text{ Normalisation de Y}$$

$$m(6) = \frac{1}{N3} \sum_{j=1}^{N3} \max(t1(j), 0), \text{ Sauvegarde de la partie positive de la mesure}$$

$$m(7) = -\frac{1}{N3} \sum_{j=1}^{N3} \min(t1(j), 0), \text{ Sauvegarde de la partie négative de la mesure}$$

TMNB-2 (partie supérieure de la bande, large de 3 Bark):

$$t2(j) = \frac{1}{23} \sum_{i=43}^{65} Y(i, j) - \frac{1}{23} \sum_{i=43}^{65} X(i, j), 1 \leq j \leq N3, \text{ Mesure}$$

$$Y(i, j) = Y(i, j) - t2(j), 43 \leq i \leq 65, 1 \leq j \leq N3, \text{ Normalisation de } Y$$

$$m(8) = \frac{1}{N3} \sum_{j=1}^{N3} \max(t2(j), 0), \text{ Sauvegarde de la partie positive de la mesure}$$

TMNB-3 (milieu de la bande, couche médiane, large de 5 Bark):

$$t3(j) = \frac{1}{12} \sum_{i=7}^{18} Y(i, j) - \frac{1}{12} \sum_{i=7}^{18} X(i, j), 1 \leq j \leq N3, \text{ Mesure}$$

$$Y(i, j) = Y(i, j) - t3(j), 7 \leq i \leq 18, 1 \leq j \leq N3, \text{ Normalisation de } Y$$

$$m(9) = \frac{1}{N3} \sum_{j=1}^{N3} \max(t3(j), 0), \text{ Sauvegarde de la partie positive de la mesure}$$

TMNB-4 (milieu de la bande, couche médiane, large de 5 Bark):

$$t4(j) = \frac{1}{24} \sum_{i=19}^{42} Y(i, j) - \frac{1}{24} \sum_{i=19}^{42} X(i, j), 1 \leq j \leq N3, \text{ Mesure}$$

$$Y(i, j) = Y(i, j) - t4(j), 19 \leq i \leq 42, 1 \leq j \leq N3, \text{ Normalisation de } Y$$

TMNB-5 (milieu de la bande, couche inférieure, large de 2,5 Bark):

$$t5(j) = \frac{1}{5} \sum_{i=7}^{11} Y(i, j) - \frac{1}{5} \sum_{i=7}^{11} X(i, j), 1 \leq j \leq N3, \text{ Mesure}$$

$$Y(i, j) = Y(i, j) - t5(j), 7 \leq i \leq 11, 1 \leq j \leq N3, \text{ Normalisation de } Y$$

$$m(10) = \frac{1}{N3} \sum_{j=1}^{N3} \max(t5(j), 0), \text{ Sauvegarde de la partie positive de la mesure}$$

TMNB-6 (milieu de la bande, couche inférieure, large de 2,5 Bark):

$$t6(j) = \frac{1}{7} \sum_{i=12}^{18} Y(i, j) - \frac{1}{7} \sum_{i=12}^{18} X(i, j), 1 \leq j \leq N3, \text{ Mesure}$$

$$Y(i, j) = Y(i, j) - t6(j), 12 \leq i \leq 18, 1 \leq j \leq N3, \text{ Normalisation de } Y$$

TMNB-7 (milieu de la bande, couche inférieure, large de 2,5 Bark):

$$t7(j) = \frac{1}{10} \sum_{i=19}^{28} Y(i, j) - \frac{1}{10} \sum_{i=19}^{28} X(i, j), 1 \leq j \leq N3, \text{ Mesure}$$

$$Y(i, j) = Y(i, j) - t7(j), 19 \leq i \leq 28, 1 \leq j \leq N3, \text{ Normalisation de } Y$$

$$m(11) = \frac{1}{N3} \sum_{j=1}^{N3} \max(t7(j), 0), \text{ Sauvegarde de la partie positive de la mesure}$$

TMNB-8 (milieu de la bande, couche inférieure, large de 2,5 Bark):

$$t8(j) = \frac{1}{14} \sum_{i=29}^{42} Y(i, j) - \frac{1}{14} \sum_{i=29}^{42} X(i, j), 1 \leq j \leq N3, \text{ Mesure}$$

$$Y(i, j) = Y(i, j) - t8(j), 29 \leq i \leq 42, 1 \leq j \leq N3, \text{ Normalisation de } Y$$

Mesure résiduelle:

$$t9(i, j) = Y(i, j) - X(i, j), 1 \leq i \leq 65, 1 \leq j \leq N3, \text{ Mesure résiduelle}$$

$$m(12) = \frac{1}{N364} \sum_{i=2}^{65} \sum_{j=1}^{N3} \max(t9(i, j), 0), \text{ Sauvegarde de la partie positive de la mesure résiduelle}$$

Il convient de noter que si deux mesures (partie positive et partie négative) avaient été retenues pour chacun des 9 modules TMNB de la structure, il en découlerait en tout 18 mesures. La nature hiérarchique de la structure MNB, ainsi que l'équipotence qui caractérise le module MNB, donne une dépendance linéaire entre ce 18 mesures. Seules 7 mesures MNB linéairement indépendantes sont disponibles. Ces mesures combinées avec l'unique mesure résiduelle et les 4 mesures FMNB donnent un nombre total de 12 mesures.

II.2.9 Combinaison linéaire de mesures pour la structure MNB

Les 12 mesures sont maintenant combinées linéairement en vue de produire une valeur de distance AD. Les coefficients de pondération utilisés dans cette combinaison linéaire sont indiqués au Tableau II.2:

$$AD = \sum_{i=1}^{12} m(i) \cdot \text{weight}_i.$$

Il convient de noter que lorsque les 12 mesures sont toutes nulles, la distance AD est également nulle.

Tableau II.2/P.861 – Coefficients de pondération utilisés pour la combinaison linéaire de la structure MNB

i	weight_i
1	0,0000
2	-0,0023
3	-0,0684
4	0,0744
5	0,0142
6	0,0100
7	0,0008
8	0,2654
9	0,1873
10	2,2357
11	0,0329
12	0,0000

NOTE – Il peut sembler inutile de calculer les valeurs de mesure et de les multiplier par un coefficient de pondération nul (à savoir weight₁ et weight₁₂). Le maintien de ce calcul s'explique cependant pour deux raisons. Premièrement, l'ensemble de mesures m₁ à m₁₂ représente l'ensemble complet des valeurs de mesure linéairement indépendantes disponibles dans la structure MNB. Deuxièmement, cet ensemble de coefficients de pondération weight_i constitue les meilleurs coefficients pouvant être utilisés pour les applications et les conditions énoncées dans le Tableau II.1. Pour d'autres applications et conditions, il peut être nécessaire d'appliquer une autre fonction de pondération ayant des valeurs non nulles.

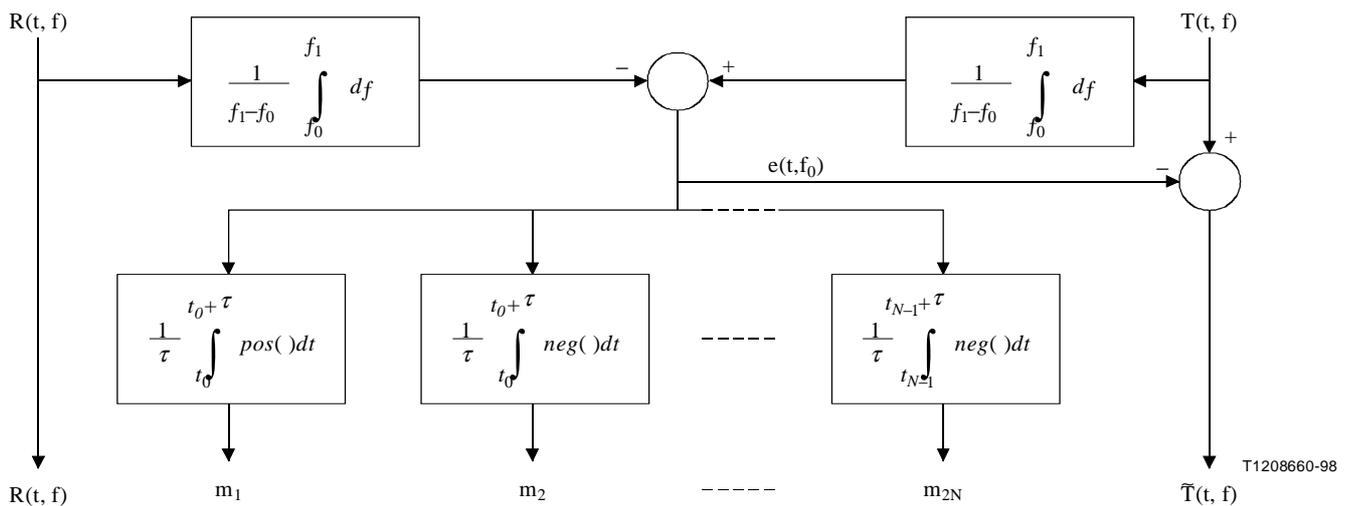


Figure II.1/P.861 – Modules de normalisation de mesures temporelle (TMNB)

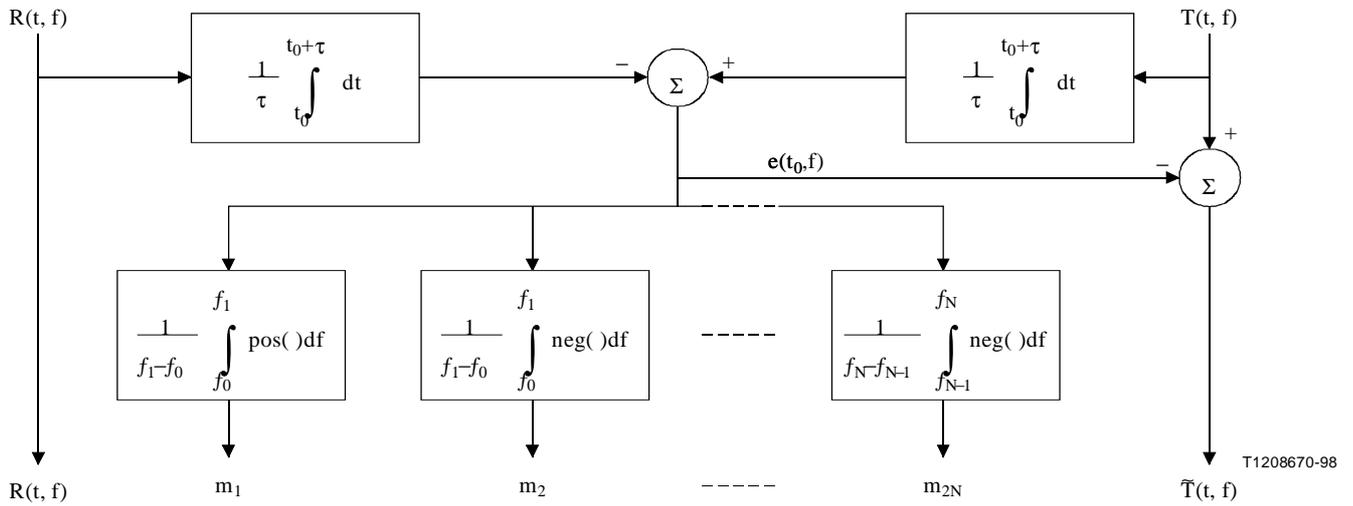
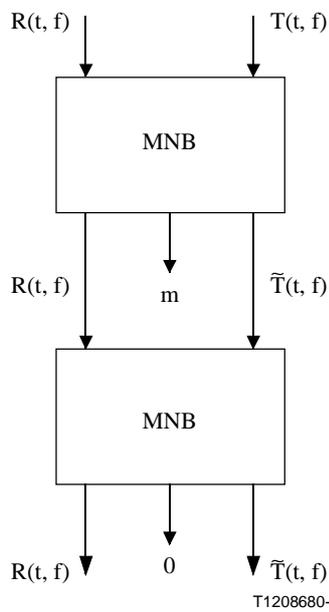


Figure II.2/P.861 – Modules de normalisation de mesure fréquentielle (FMNB)



Si $MNB(R(t, f), T(t, f)) = (R(t, f), \tilde{T}(t, f), \underline{m})$,
 Alors $MNB(R(t, f), \tilde{T}(t, f)) = (R(t, f), \tilde{T}(t, f), \underline{0})$

Figure II.3/P.861 – Les modules MNB sont équipotents

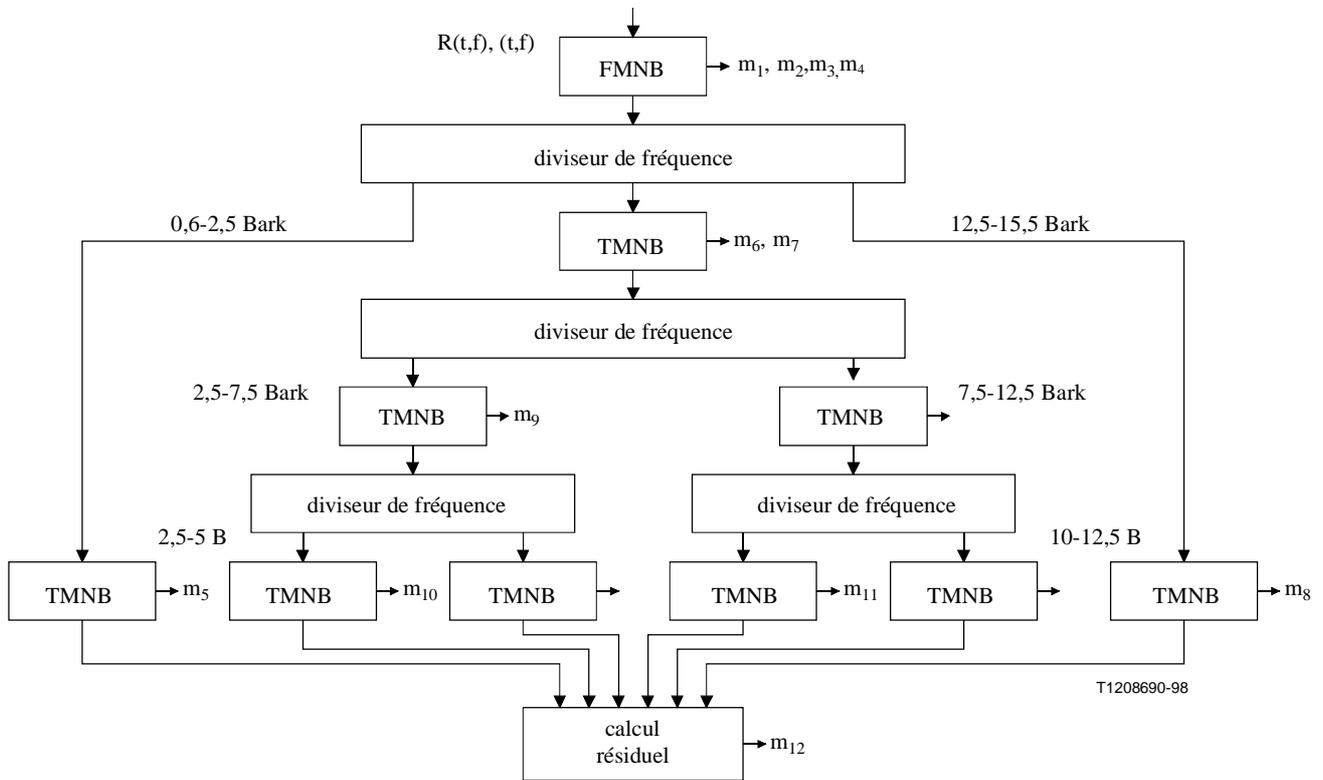


Figure II.4/P.861 – Structure hiérarchique des modules MNB

Bibliographie

- [1] Contribution COM 12-25 de l'UIT-T: Summary of available performance results for the MNB voice quality measurement algorithm, Etats-Unis, février 1998.
- [2] Contribution COM 12-26 de l'UIT-T: Comparison of two methods for objective assessment of the quality of coded speech, AT&T, février 1998.
- [3] Contribution COM 12-27 de l'UIT-T: Results of objective voice quality experiments, Etats-Unis, février 1998.
- [4] Contribution COM 12-20 de l'UIT-T: Improvement of the P.861 Perceptual Speech Quality Measure, KPN, février 1998.
- [5] Contribution COM 12-21 de l'UIT-T: An experimental investigation of the accumulation of perceived error in time-varying speech distortions, BT, février 1998.
- [6] Contribution COM 12-34 de l'UIT-T: TOSQA – Telecommunication objective speech quality assessment, FRG, février 1998.

SERIES DES RECOMMANDATIONS UIT-T

Série A	Organisation du travail de l'UIT-T
Série B	Moyens d'expression: définitions, symboles, classification
Série C	Statistiques générales des télécommunications
Série D	Principes généraux de tarification
Série E	Exploitation générale du réseau, service téléphonique, exploitation des services et facteurs humains
Série F	Services de télécommunication non téléphoniques
Série G	Systèmes et supports de transmission, systèmes et réseaux numériques
Série H	Systèmes audiovisuels et multimédias
Série I	Réseau numérique à intégration de services
Série J	Transmission des signaux radiophoniques, télévisuels et autres signaux multimédias
Série K	Protection contre les perturbations
Série L	Construction, installation et protection des câbles et autres éléments des installations extérieures
Série M	RGT et maintenance des réseaux: systèmes de transmission, de télégraphie, de télécopie, circuits téléphoniques et circuits loués internationaux
Série N	Maintenance: circuits internationaux de transmission radiophonique et télévisuelle
Série O	Spécifications des appareils de mesure
Série P	Qualité de transmission téléphonique, installations téléphoniques et réseaux locaux
Série Q	Commutation et signalisation
Série R	Transmission télégraphique
Série S	Equipements terminaux de télégraphie
Série T	Terminaux des services télématiques
Série U	Commutation télégraphique
Série V	Communications de données sur le réseau téléphonique
Série X	Réseaux pour données et communication entre systèmes ouverts
Série Y	Infrastructure mondiale de l'information
Série Z	Langages de programmation