



UNION INTERNATIONALE DES TÉLÉCOMMUNICATIONS

UIT-T

SECTEUR DE LA NORMALISATION
DES TÉLÉCOMMUNICATIONS
DE L'UIT

P.50

(09/99)

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

Appareils de mesures objectives

Voix artificielle

Recommandation UIT-T P.50

(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.

VOIX ARTIFICIELLE

Résumé

La "voix artificielle" décrite dans la présente Recommandation reproduit les caractéristiques de la parole humaine servant à caractériser des systèmes et dispositifs de télécommunication linéaires et non linéaires conçus pour la transduction ou la transmission de la parole. La voix artificielle est un signal qui est défini mathématiquement et qui reproduit les caractéristiques temporelles et spectrales de la parole qui ont une influence significative sur le fonctionnement des systèmes de télécommunication. Deux types de voix artificielle sont définis; ils reproduisent respectivement les caractéristiques spectrales de voix de femme et d'homme.

Source

La Recommandation UIT-T P.50, révisée par la Commission d'études 12 de l'UIT-T (1997-2000), a été approuvée le 30 septembre 1999 selon la procédure définie dans la Résolution n° 1 de la CMNT.

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, le terme *exploitation reconnue (ER)* désigne tout particulier, toute entreprise, toute société ou tout organisme public qui exploite un service de correspondance publique. Les termes *Administration*, *ER* et *correspondance publique* sont définis dans la *Constitution de l'UIT (Genève, 1992)*.

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 n'avait pas 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 2000

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 Introduction.....	1
2 Domaine d'application, objet et définition.....	1
2.1 Domaine d'application et objet.....	1
2.2 Définition.....	2
3 Terminologie.....	2
3.1 Voix artificielle électrique.....	2
3.2 Signal d'excitation de la bouche artificielle.....	2
3.3 Voix artificielle acoustique.....	2
4 Caractéristiques.....	3
4.1 Spectre moyen à long terme.....	3
4.2 Spectre à court terme.....	4
4.3 Distribution des amplitudes instantanées.....	4
4.4 Répartition du niveau de puissance des segments.....	4
4.5 Spectre de l'enveloppe de modulation.....	5
4.6 Convergence temporelle.....	6
5 Méthode d'émission de la voix artificielle.....	6
5.1 Signal de source d'excitation.....	7
5.2 Excitation glottale.....	7
5.3 Sons non voisés.....	7
5.4 Enveloppe de puissance.....	8
5.5 Filtre de mise en forme du spectre.....	9
6 Bibliographie.....	12
Annexe A – Caractéristiques du spectre à court terme de la voix artificielle.....	12

Recommandation P.50

VOIX ARTIFICIELLE

(Melbourne, 1988; modifiée à Helsinki, 1993, Genève, 1999)

1 Introduction

Le signal décrit ici reproduit les caractéristiques de la parole humaine servant à caractériser des systèmes et dispositifs de télécommunication linéaires et non linéaires conçus pour la transduction ou la transmission de la parole. On sait que pour certaines applications, par exemple, la mesure objective de l'équivalent pour la sonie, il est possible d'utiliser des signaux plus simples, par exemple, un bruit rose ou un bruit gaussien à spectre modelé; ces signaux ne peuvent être considérés comme des "voix artificielles" aux fins de la présente Recommandation.

La voix artificielle est un signal qui est défini mathématiquement et qui reproduit les caractéristiques temporelles et spectrales de la parole qui ont une influence significative sur le fonctionnement des systèmes de télécommunication. Deux types de voix artificielle sont définis; ils reproduisent respectivement les caractéristiques spectrales de voix de femme et d'homme.

Les caractéristiques temporelles et spectrales de la voix réelle mentionnées ci-après sont reproduites par la voix artificielle:

- a) spectre moyen à long terme;
- b) spectre à court terme;
- c) distribution des amplitudes instantanées;
- d) structure voisée et non voisée du signal "vocal";
- e) enveloppe syllabique.

L'Appendice I/P.50 se présente sous la forme d'un cédérom contenant des signaux d'essai utiles. Les signaux qui y sont enregistrés comprennent le signal décrit dans la Recommandation P.50 ainsi que d'autres signaux qui ont été jugés utiles par certaines Administrations. De plus, ce cédérom contient toute la base de données vocales numérisées qui a été utilisée pour mettre au point la Recommandation P.50. L'Appendice I/P.50 est publiée séparément.

2 Domaine d'application, objet et définition

2.1 Domaine d'application et objet

La voix artificielle vise à reproduire les caractéristiques de la voix réelle dans la bande comprise entre 100 Hz et 8 kHz. Elle peut servir à caractériser nombre d'appareils, tels que microphones à charbon, appareils téléphoniques à haut-parleur, codeurs non linéaires, dispositifs de protection contre les échos, compresseurs-extenseurs syllabiques et systèmes non linéaires en général.

La voix artificielle décrite dans la présente Recommandation est utilisée essentiellement pour l'évaluation objective des systèmes et dispositifs de traitement de la parole dans lesquels un signal monovoie à activité continue (c'est-à-dire sans repos) est suffisant pour mesurer les caractéristiques. A titre d'exemple, citons l'évaluation des codecs vocaux. Pour l'évaluation objective qui nécessite deux signaux avec des pauses (par exemple, évaluation de dispositifs avec détecteurs vocaux), il convient d'utiliser le signal de conversation artificiel décrit dans la Recommandation P.59.

L'utilisation de la voix artificielle à la place de la voix réelle offre deux avantages: elle est plus facile à émettre et présente moins de variations que les échantillons de voix réelle.

Naturellement, si l'on mesure un système déterminé, il faut tenir compte des caractéristiques du trajet de transmission qui le précède. Le signal d'essai réel produit doit alors représenter la convolution entre la voix artificielle et la réponse sur le trajet.

2.2 Définition

La **voix artificielle** est un signal mathématiquement défini qui reproduit toutes les caractéristiques de la voix humaine nécessaires pour caractériser les systèmes de télécommunication linéaires ou non linéaires, en vue d'obtenir une corrélation satisfaisante entre les mesures objectives et les mesures sur voix réelle.

3 Terminologie

La voix artificielle peut être produite sous la forme d'un signal électrique ou acoustique, selon le système ou l'appareil étudié (voies de communication, codeurs, microphones). Les définitions ci-après se rapportent aux signaux indiqués dans la légende de la Figure 1.

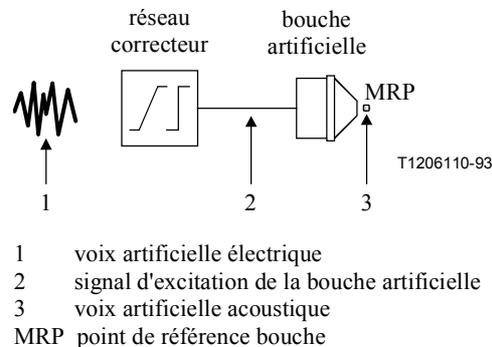


Figure 1/P.50

3.1 Voix artificielle électrique

Voix artificielle produite sous la forme d'un signal électrique pour tester ces voies de transmission ou d'autres dispositifs électriques.

3.2 Signal d'excitation de la bouche artificielle

Signal appliqué à la bouche artificielle afin d'obtenir la voix artificielle acoustique. Ce signal résulte de l'égalisation de la voix artificielle électrique, compte tenu de la caractéristique efficacité en fonction de la fréquence de la bouche.

NOTE – L'égalisation dépend du type de bouche utilisé et peut être effectuée électriquement ou mathématiquement au cours du processus de production du signal.

3.3 Voix artificielle acoustique

Signal acoustique au point de référence bouche (MRP, *mouth reference point*) de la bouche artificielle; ce signal doit être conforme aux mêmes spécifications de temps et de spectre que la voix artificielle électrique.

4 Caractéristiques

4.1 Spectre moyen à long terme

Le spectre moyen à long terme filtré en tiers d'octave de la voix artificielle est représenté à la Figure 2 et au Tableau 1, normalisé pour un niveau de pression acoustique à large bande de $-4,7$ dBPa.

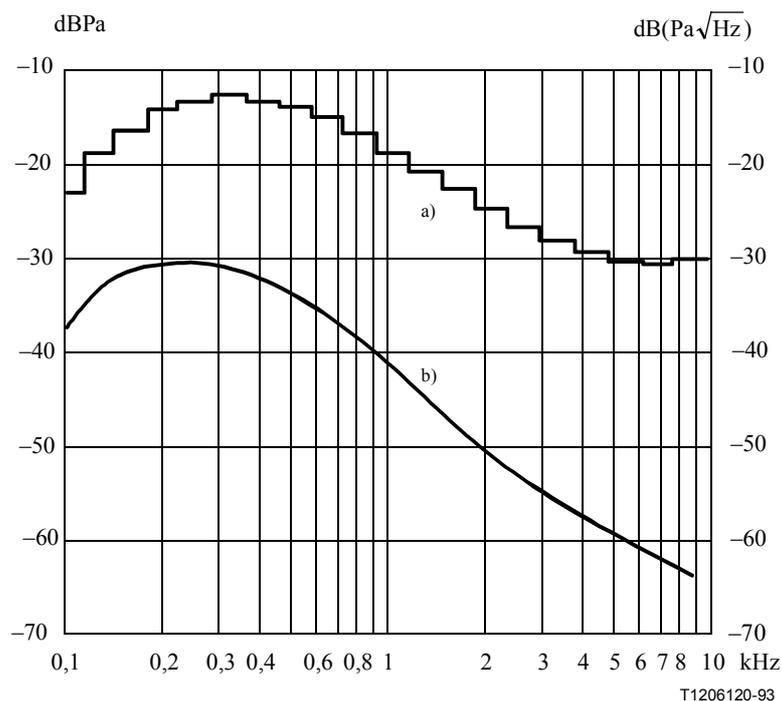
Les valeurs du spectre à long terme de la voix artificielle au point de référence bouche sont données par la formule:

$$S(f) = -376,44 + 465,439(\log_{10}f) - 157,745(\log_{10}f)^2 + 16,7124(\log_{10}f)^3 \quad (4-1)$$

où $S(f)$ est la densité spectrale en dB par rapport à une intensité sonore de 1 pW/m^2 par Hz à la fréquence f . La gamme de fréquences de définition est de 100 Hz à 8 kHz .

La courbe du spectre est représentée dans la Figure 2. Les valeurs de $S(f)$ aux fréquences définies par l'ISO correspondant à des tiers d'octave sont données dans la 4^e colonne du Tableau 1. Les tolérances sont indiquées dans la 5^e colonne du même tableau. Les tolérances au-dessous de 200 Hz ne s'appliquent qu'à la voix d'homme artificielle.

Le niveau de pression acoustique totale du spectre défini par la formule (4-1) est de $-4,7 \text{ dBPa}$. Cependant, ce spectre est valable également pour les niveaux compris entre $-19,7$ et $+10,3 \text{ dBPa}$. Cela signifie que le premier terme de la formule (4-1) peut être compris entre $-391,44$ et $-361,44$.



- a) spectre au 1/3 d'octave [Tableau 1, colonne (3)].
- b) densité du spectre [Tableau 1, colonne (3)-(2)].

Figure 2/P.50 – Spectre à long terme de la voix artificielle

4.2 Spectre à court terme

Les caractéristiques du spectre à court terme des voix artificielles d'homme et de femme sont décrites dans l'Annexe A.

4.3 Distribution des amplitudes instantanées

La probabilité de distribution de densité des amplitudes instantanées d'une voix artificielle est représentée à la Figure 3.

4.4 Répartition du niveau de puissance des segments

La répartition du niveau de puissance des segments de la voix artificielle, mesurée en fenêtres de temps de 16 ms est donnée par la Figure 4. Les limites de tolérance supérieure et inférieure sont également représentées.

NOTE – La limite de tolérance supérieure représente la répartition type du niveau de puissance des segments d'une conversation normale, alors que la limite inférieure représente la parole continue (phrases téléphonométriques).

Tableau 1/P.50 – Spectre à long terme de la voix artificielle

Fréquence au centre du 1/3 octave	Facteur de correction de la largeur de bande 10 $\log_{10}\Delta f$	Niveau de pression acoustique (1/3 octave)	Densité du spectre	Tolérance
(Hz) (1)	(dB) (2)	(dBPa) (3)	(dB) (3) – (2)	(dB)
100	13,6	-23,1	-36,7	-
125	14,6	-19,2	-33,8	+3, -6 ^{a)}
160	15,6	-16,4	-32,7	+3, -6 ^{a)}
200	16,6	-14,4	-31,7	+3, -6
250	17,6	-13,4	-31,7	±3,0
315	18,6	-13,0	-31,6	±3,0
400	19,6	-13,3	-32,9	±3,0
500	20,6	-14,1	-34,7	±3,0
630	21,6	-15,4	-37,7	±3,0
800	22,6	-17,0	-39,6	±3,0
1000	23,6	-18,9	-42,5	±3,0
1250	24,6	-21,0	-45,6	±3,0
1600	25,6	-23,0	-48,6	±3,0
2000	26,6	-25,1	-51,7	±3,0
2500	27,6	-26,9	-54,5	±3,0
3150	28,6	-28,6	-57,2	±3,0
4000	29,6	-29,8	-59,4	±6,0
5000	30,6	-30,6	-61,2	±6,0
6300	31,6	-30,9	-62,5	±6,0
8000	32,6	-30,5	-63,1	-

^{a)} Les tolérances indiquées s'appliquent au spectre à long terme de la voix d'homme et doivent aussi s'appliquer au bruit modelé de la parole. Toutefois, elles ne s'appliquent pas au spectre de la voix de femme, dont le contenu d'énergie dans cette gamme de fréquences est pratiquement négligeable.

4.5 Spectre de l'enveloppe de modulation

Le spectre du signal d'enveloppe de modulation est représenté par la Figure 5 et il devra être reproduit avec une tolérance de ± 5 dB sur la gamme entière des fréquences.

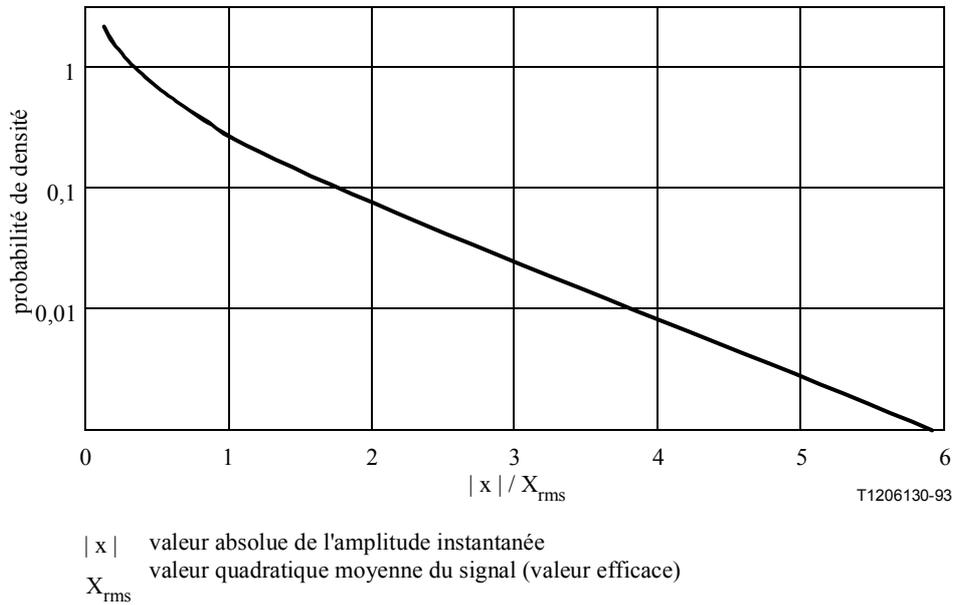


Figure 3/P.50 – Distribution des amplitudes instantanées

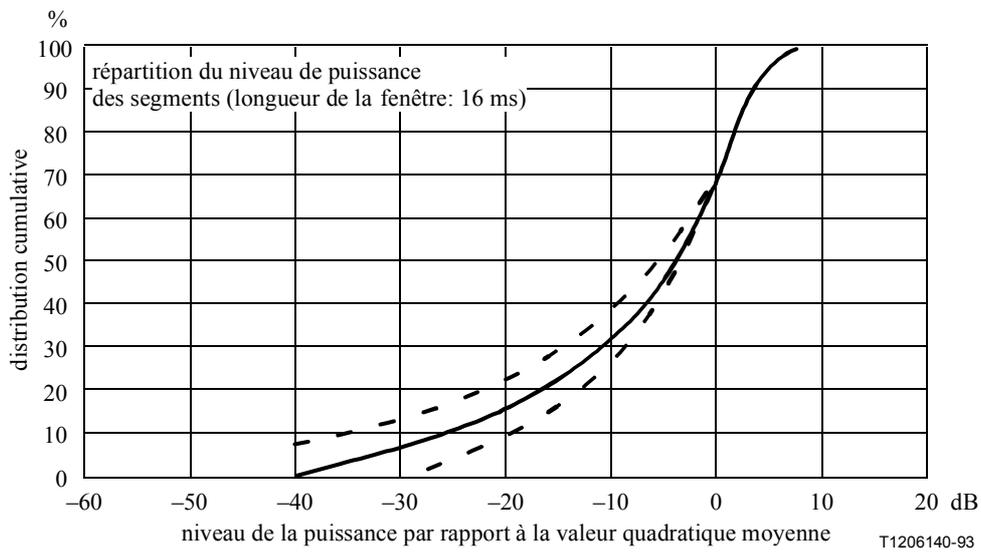


Figure 4/P.50 – Répartition du niveau de puissance des segments

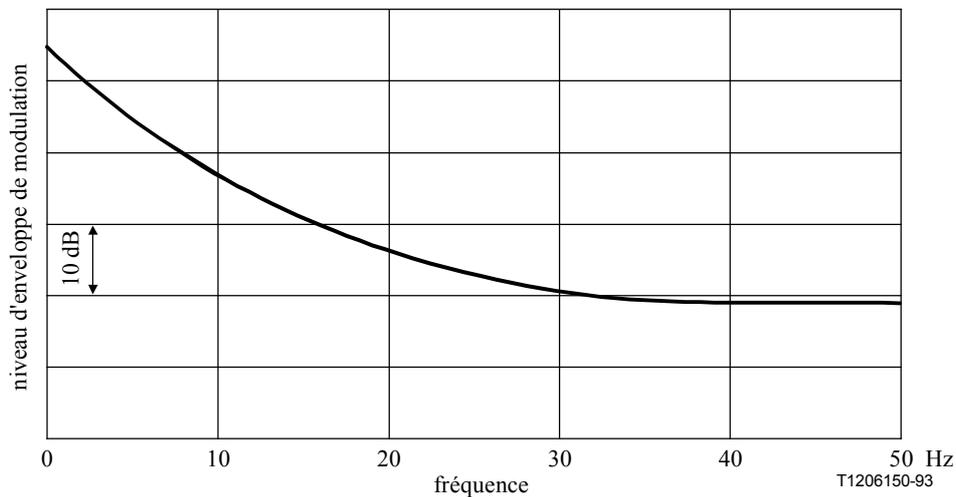


Figure 5/P.50 – Spectre de l'enveloppe de modulation

4.6 Convergence temporelle

La voix artificielle doit avoir des caractéristiques aussi proches que possible de la voix réelle. En particulier, il doit être possible d'obtenir les caractéristiques de spectre à long terme et de distribution d'amplitude sur 10 secondes.

5 Méthode d'émission de la voix artificielle

La Figure 6 montre un diagramme du processus d'émission de la voix artificielle. Elle est émise par l'application de deux types différents de signaux de source d'excitation: un signal d'excitation glottale et un bruit aléatoire, à un filtre de mise en forme du spectre variant en fonction du temps. La voix artificielle émise par le signal d'excitation glottale et par le bruit aléatoire correspond respectivement aux sons voisés et non voisés. La réponse en fréquence du filtre de mise en forme du spectre simule les caractéristiques de transmission du trajet vocal.

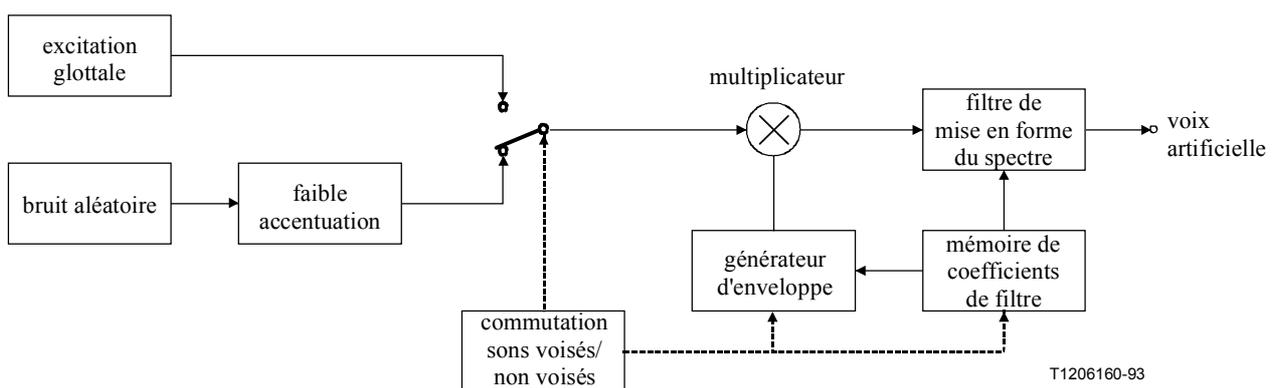


Figure 6/P.50 – Processus d'émission de la voix artificielle

5.1 Signal de source d'excitation

On obtient la voix artificielle en alternant de façon aléatoire quatre éléments unitaires de base, contenant chacun des segments voisés et non voisés. Alors qu'un élément unitaire commence par un son non voisé, suivi d'un son voisé, les trois autres éléments commencent par un son voisé suivi d'un son non voisé et se terminent par un son voisé de nouveau (voir aussi Figure 9). Le rapport de la durée du son non voisé T_{uv} à la durée totale des segments voisés T_v pour chaque élément unitaire est 0,25. La durée $T = T_{uv} + T_v$ des éléments unitaires varie conformément à la formule suivante:

$$T = -3,486 \ln(r)$$

où r est un nombre aléatoire distribué uniformément ($0,371 \leq r \leq 0,609$).

Les durées de sons voisés et non voisés des quatre éléments unitaires sont les suivantes:

Elément a: non voisé (T_{uv}) + voisé (T_v)

Elément b: voisé ($T_v/4$) + non voisé (T_{uv}) + voisé ($3 T_v/4$)

Elément c: voisé ($T_v/2$) + non voisé (T_{uv}) + voisé ($T_v/2$)

Elément d: voisé ($3 T_v/4$) + non voisé (T_{uv}) + voisé ($T_v/4$).

Les éléments unitaires seront répétés de façon aléatoire pendant au moins 10 secondes afin de satisfaire aux caractéristiques de la voix artificielle spécifiées à l'article 4.

5.2 Excitation glottale

Le signal d'excitation glottale est une forme d'onde périodique comme le montre la Figure 7. La fréquence maximale ($1/T_0$ dans la Figure 7) varie en fonction du diagramme de variation présenté à la Figure 8 pendant la période T_v . La valeur initiale de la fréquence maximale (F_s dans la Figure 8) est déterminée d'après les relations suivantes:

$F_s = F_c - 31,82 T_v + 39,4 R$ pour la voix d'homme artificielle;

$F_s = F_c - 51,85 T_v + 64,2 R$ pour la voix de femme artificielle.

où F_c et R indiquent respectivement la fréquence centrale et une variable aléatoire distribuée uniformément ($0 < R < 1$). F_c est de 128 Hz pour la voix d'homme artificielle et de 215 Hz pour la voix de femme artificielle. Dans le trapèze du diagramme de variation de la fréquence maximale, la partie de surface du trapèze placée au-dessus de F_c doit être égale à celle située au-dessous de F_c (partie hachurée dans la Figure 8). Pour les éléments b), c) et d) de la Figure 9 le diagramme de variation de la fréquence fondamentale s'applique à la combinaison des deux parties voisées, quel que soit l'endroit où le segment non voisé est inséré.

5.3 Sons non voisés

La fonction de transfert du filtre passe-bas situé après le générateur de bruit aléatoire (faible accentuation) est $1/(1 - (1 - \varepsilon) z^{-1})$, $\varepsilon = 0,001$, où z^{-1} indique le délai unitaire. Le générateur de bruit aléatoire génère un bruit blanc gaussien avec un processus aléatoire discret à moyenne nulle dont les échantillons ne sont pas corrélés.

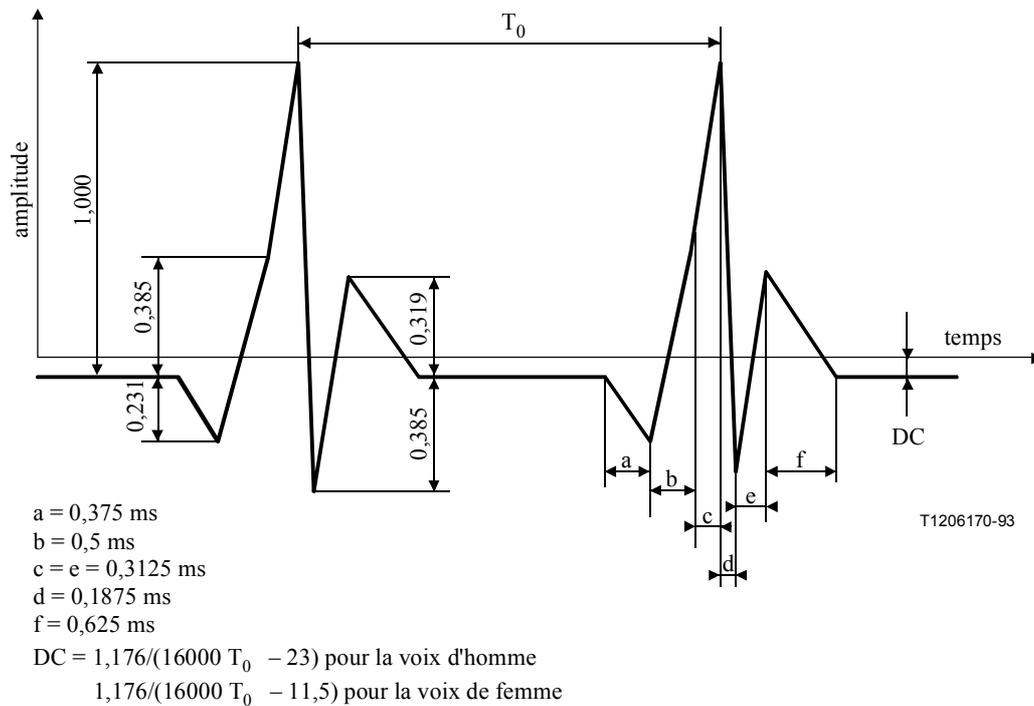


Figure 7/P.50 – Signal d'excitation glottale

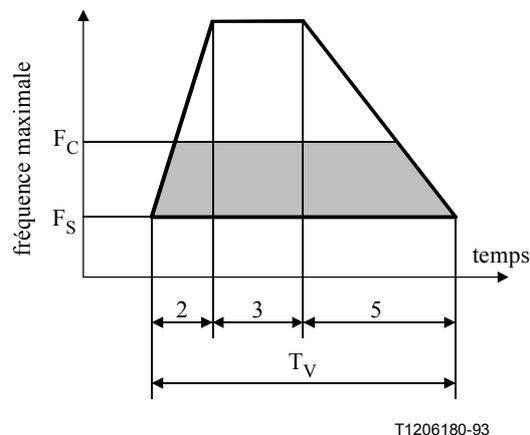


Figure 8/P.50 – Diagramme de variation de la fréquence maximale

5.4 Enveloppe de puissance

L'enveloppe de puissance de chaque élément unitaire du signal de source d'excitation est commandée de telle façon que la puissance des segments à court terme (évaluée sur des intervalles de 2 ms) de la voix artificielle varie en fonction des diagrammes présentés dans la Figure 9. On l'obtient en utilisant la relation suivante qui indique les signaux d'entrée et de sortie du filtre de mise en forme du spectre:

$$P_E = P_S \prod_{i=1}^{12} (1 - k_i^2)$$

où

P_E est la puissance d'entrée dans le filtre de mise en forme du spectre,

P_S est la puissance de sortie du filtre de mise en forme du spectre,

k_i est l'ième coefficient du filtre de mise en forme du spectre.

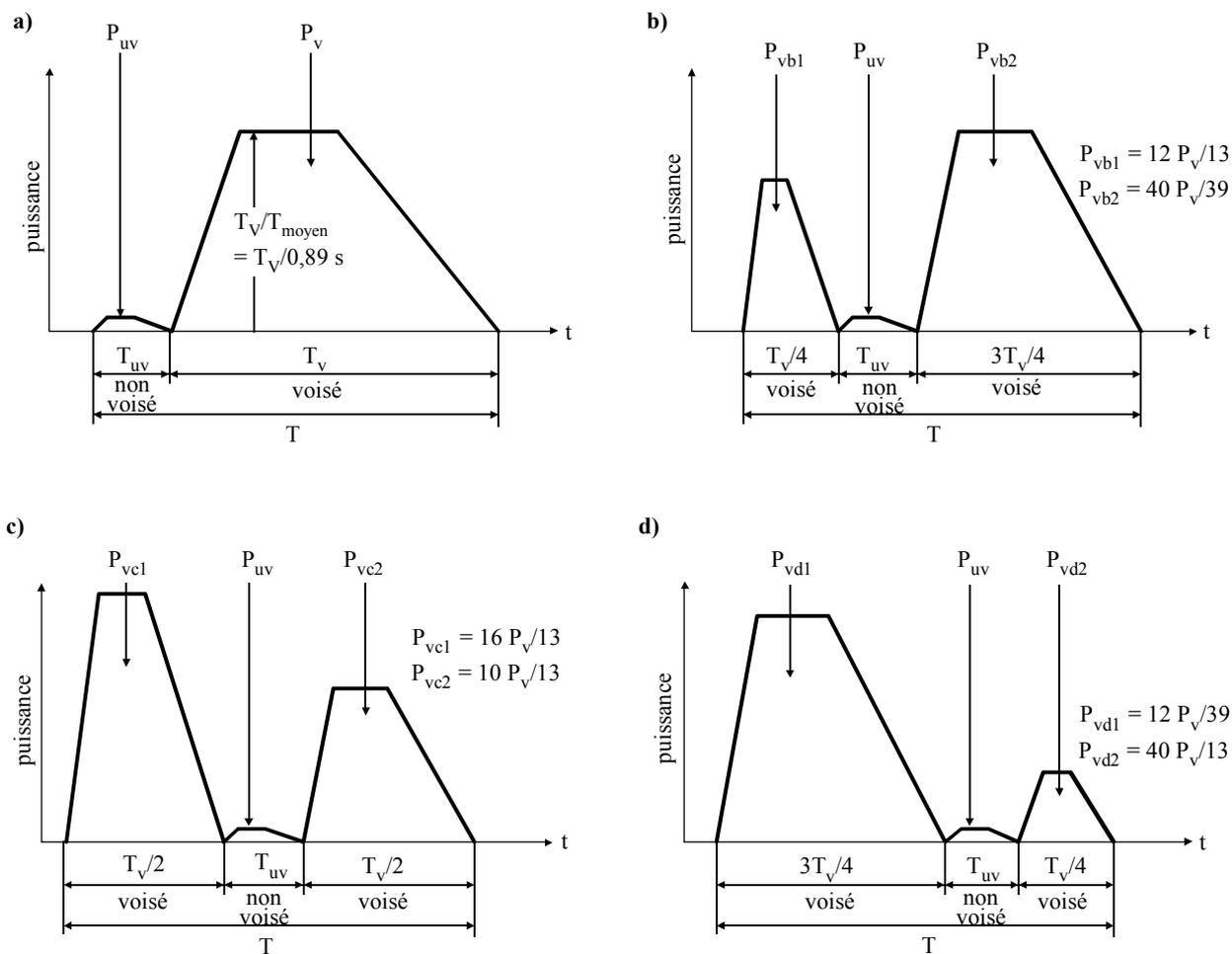
Les temps de montée, de maintien et de descente de chaque trapèze de la Figure 9 doivent être mutuellement liés par les mêmes coefficients de proportionnalité (2:3:5) du diagramme de variation de la fréquence maximale présenté à la Figure 8. Pour chaque élément unitaire, la puissance moyenne des sons non voisés (P_{uv}) doit être inférieure de 17,5 dB de la puissance moyenne des sons voisés (P_v).

5.5 Filtre de mise en forme du spectre

Le filtre de mise en forme du spectre a une structure en treillis du 12^e ordre comme le montre la Figure 10. Seize groupes de 12 coefficients de filtrage chacun ($k_1 - k_{12}$), sont définis; treize groupes seront utilisés pour émettre la partie voisée, et trois groupes seront utilisés pour émettre la partie non voisée. Ces coefficients sont indiqués dans le Tableau 2 à la fois pour les voix artificielles d'homme et de femme.

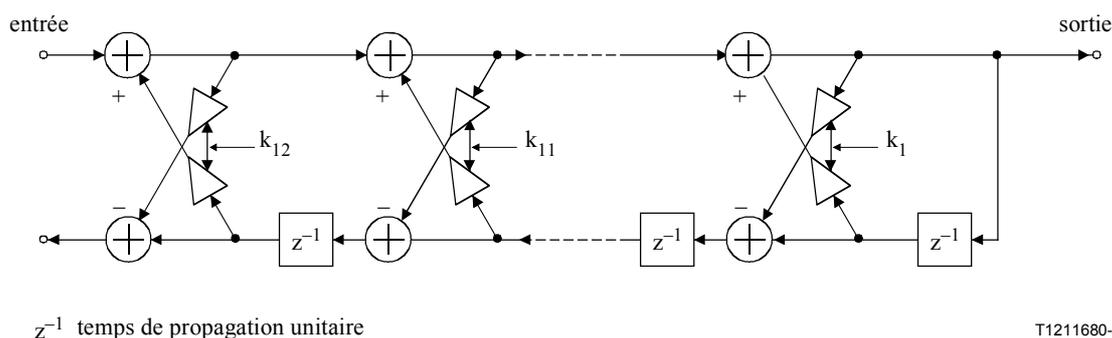
Les douze coefficients de filtre seront mis à jour toutes les 60 ms lors de l'émission du signal. Plus précisément, pendant chaque période de 60 ms, les coefficients de filtrage réels doivent être remis à jour toutes les 2 ms, par interpolation linéaire entre les deux ensembles de valeurs adoptés pour les intervalles suivants de 60 ms. Dans la partie son voisé, chacun des treize groupes de coefficients sera choisi au hasard toutes les 780 ms ($60 \text{ ms} \times 13$), et dans la partie son non voisé, chacun des trois groupes de coefficients sera choisi au hasard toutes les 180 ms ($60 \text{ ms} \times 3$).

NOTE – La mise en œuvre décrite du filtre de mise en forme doit être considérée comme un exemple; elle ne fait pas partie intégrante de la présente Recommandation. On peut utiliser une toute autre implantation assurant la même fonction de transfert. Un fréquence d'échantillonnage de 16 000 Hz correspond aux Tableaux 2a et 2b.



T1206190-93

Figure 9/P.50 – Diagrammes de variation de puissance à court terme des quatre éléments unitaires utilisés pour émettre la voix artificielle



T1211680-99

Figure 10/P.50 – Filtre de mise en forme du spectre

NOTE – L'accès de sortie du filtre situé dans la partie inférieure de la Figure 10 n'est pas utilisé, mais est indiqué pour faire ressortir la symétrie de la structure du filtre.

Tableau 2a/P.50 – Coefficients k_i pour la voix artificielle d'homme

	k_1	k_2	k_3	k_4	k_5	k_6	k_7	k_8	k_9	k_{10}	k_{11}	k_{12}	
Non voisé	1	-0,471	-0,108	0,024	-0,048	0,140	0,036	0,054	0,004	0,123	0,044	0,099	-0,003
	2	-0,284	-0,468	0,030	0,090	0,124	-0,020	0,087	0,067	0,131	0,011	0,076	-0,024
	3	-0,025	-0,496	-0,176	0,162	0,236	-0,012	0,068	0,001	0,096	0,029	0,086	-0,018
Voisé	1	0,974	0,219	0,025	-0,123	-0,132	-0,203	-0,103	-0,174	-0,079	-0,153	-0,010	-0,061
	2	0,629	-0,152	-0,138	-0,142	-0,118	-0,135	0,147	0,019	0,077	-0,040	0,029	-0,007
	3	0,599	-0,119	0,067	0,051	0,103	0,023	0,106	0,036	-0,006	-0,133	-0,052	-0,094
	4	0,164	-0,364	-0,248	-0,076	0,168	0,072	0,103	0,045	0,112	0,010	0,048	-0,034
	5	0,842	0,022	0,171	0,173	0,067	-0,057	0,089	-0,045	-0,039	-0,134	-0,034	-0,122
	6	0,933	-0,537	-0,137	-0,161	-0,216	-0,139	0,115	-0,042	0,027	-0,163	0,102	-0,107
	7	0,937	-0,413	0,132	-0,059	-0,103	-0,134	0,047	-0,115	-0,105	-0,097	0,039	-0,108
	8	0,965	-0,034	0,032	0,001	-0,107	-0,189	-0,057	-0,175	-0,109	-0,163	-0,003	-0,055
	9	0,870	-0,476	-0,016	-0,136	-0,125	-0,107	0,091	-0,008	0,021	-0,128	0,042	-0,069
	10	0,686	-0,030	0,178	0,197	0,155	-0,026	0,078	0,004	-0,001	-0,128	-0,004	-0,102
	11	0,963	-0,232	0,086	-0,018	-0,147	-0,192	-0,040	-0,179	-0,144	-0,133	0,042	-0,042
	12	0,930	-0,461	0,071	-0,144	-0,122	-0,096	0,034	-0,066	-0,021	-0,171	0,067	-0,091
	13	0,949	-0,334	0,143	-0,040	-0,112	-0,161	0,010	-0,156	-0,123	-0,119	0,049	-0,070

Tableau 2b/P.50 – Coefficients k_i pour la voix artificielle de femme

	k_1	k_2	k_3	k_4	k_5	k_6	k_7	k_8	k_9	k_{10}	k_{11}	k_{12}	
Non voisé	1	0,488	-0,388	0,145	0,053	0,122	0,027	0,135	0,035	0,080	0,017	0,068	0,028
	2	-0,093	-0,444	-0,102	0,121	0,154	0,009	0,102	-0,031	0,084	0,019	0,101	-0,020
	3	-0,709	-0,179	0,134	0,007	0,142	0,027	0,099	0,000	0,115	0,007	0,075	-0,037
Voisé	1	0,355	-0,247	-0,092	-0,043	0,032	0,046	0,113	-0,023	0,071	-0,030	-0,000	-0,116
	2	0,976	0,150	-0,062	-0,187	-0,172	-0,200	-0,122	-0,207	-0,054	-0,127	0,012	-0,111
	3	0,737	-0,324	-0,175	-0,197	-0,153	0,023	0,110	-0,018	0,040	-0,062	0,034	-0,091
	4	0,598	0,234	0,126	0,011	-0,005	-0,026	0,131	0,032	0,073	-0,063	0,011	-0,088
	5	0,808	0,118	0,262	0,139	0,063	-0,024	0,001	-0,184	-0,056	-0,100	0,014	-0,115
	6	0,914	-0,500	-0,051	-0,115	-0,211	-0,012	-0,077	-0,179	0,064	-0,102	0,037	-0,092
	7	0,933	-0,359	0,089	-0,107	-0,178	-0,050	-0,137	-0,206	0,046	-0,088	-0,004	-0,074
	8	0,966	-0,023	0,044	-0,105	-0,178	-0,195	-0,150	-0,233	-0,045	-0,092	0,029	-0,097
	9	0,870	-0,469	-0,244	-0,107	-0,140	-0,037	0,084	-0,131	0,021	-0,066	-0,003	-0,091
	10	0,673	-0,292	0,392	0,158	0,143	0,160	0,019	-0,281	-0,105	-0,195	-0,156	-0,185
	11	0,962	-0,191	0,030	-0,089	-0,207	-0,133	-0,141	-0,263	0,007	-0,054	0,014	-0,074
	12	0,879	-0,340	0,046	-0,049	-0,071	-0,024	-0,039	-0,188	0,017	-0,078	-0,014	-0,117
	13	0,941	-0,258	0,122	-0,073	-0,163	-0,089	-0,151	-0,250	0,025	-0,062	-0,006	-0,093

6 Bibliographie

- [1] Contribution CCITT COM XII-N° 76 (1982) – Voix artificielle (Italie).
- [2] Contribution CCITT COM XII-N° 108 (1982) – Projet partiel de texte pour le paragraphe 3 de la Recommandation P.51 (Rapporteur Spécial).
- [3] Contribution CCITT COM XII-N° 11 (1981) – Mesure des microphones à charbon au moyen de signaux vocaux réels et de signaux de test artificiels (Canada).
- [4] Contribution CCITT COM XII-N° 150 (1983) – Amélioration de la voix artificielle par insertion de sons fricatifs (Italie).
- [5] Contribution CCITT COM XII-N° 132 (1983) – Evaluation de la distorsion non linéaire au moyen de signaux vocaux simulés (Canada).

ANNEXE A

Caractéristiques du spectre à court terme de la voix artificielle

On émet la voix artificielle en sélectionnant de façon aléatoire chacun des seize diagrammes du spectre à court terme toutes les 960 ms (60 ms × 16 diagrammes). On obtient la densité du spectre de chaque diagramme à l'aide de la formule (A-1) et du Tableau A.1, le spectre à court terme du signal pendant l'intervalle de 60 ms qui existe entre deux sélections de diagrammes successives varie légèrement d'un diagramme à l'autre.

NOTE – Les diagrammes du spectre dans la formule (A-1) et le Tableau A.1 sont exprimés sous une forme de puissance normalisée.

$$\text{Densité du spectre } S_i(f) = \frac{1}{A_{i0} + 2 \sum_{j=1}^{12} A_{ij} [\cos(2\pi jf)]}, i = 1, 2, \dots, 16 \quad (\text{A-1})$$

Tableau A.1a/P.50 – Coefficients A_{ij} pour la voix artificielle d'homme

<i>i</i>	<i>j</i>												
	0	1	2	3	4	5	6	7	8	9	10	11	12
1	2,09230	-1,33222	1,32175	-1,14200	0,99352	-0,94634	0,72684	-0,63263	0,41196	-0,42858	0,22070	-0,19746	0,10900
2	9,34810	-8,55934	7,35732	-6,35320	5,33999	-4,47238	3,62417	-2,85246	2,12260	-1,49424	0,93988	-0,44998	0,12400
3	11,69068	-10,91138	9,46588	-8,11729	6,94160	-5,90977	4,95137	-3,89587	2,88750	-1,97671	1,14892	-0,50255	0,12100
4	12,56830	-11,81209	10,36030	-8,82879	7,37947	-6,01017	4,66740	-3,46913	2,42182	-1,60880	0,91652	-0,39648	0,12000
5	6,83438	-6,18275	5,59089	-4,71866	4,06004	-3,44767	2,65380	-2,12140	1,50334	-1,07904	0,64553	-0,31816	0,11500
6	12,37251	-11,52358	9,89962	-8,31774	6,99062	-5,86272	4,69809	-3,56806	2,53340	-1,70522	0,99232	-0,45403	0,13400
7	21,07637	-19,62125	16,56781	-13,67518	11,41379	-9,61940	7,93529	-6,32841	4,92443	-3,53539	2,09095	-0,86543	0,18100
8	30,77371	-29,17365	25,52254	-21,51978	17,80583	-14,30488	10,87190	-7,71572	5,14643	-3,20113	1,72149	-0,68054	0,14400
9	4,18618	-3,36611	3,36793	-2,92133	2,38452	-2,06047	1,57550	-1,34240	0,84994	-0,70462	0,38685	-0,21857	0,12100
10	14,12359	-13,14611	11,25804	-9,47510	7,97588	-6,70717	5,44803	-4,23843	3,10807	-2,12879	1,25096	-0,53230	0,12600
11	26,36971	-24,95984	21,80496	-18,41045	15,30642	-12,49415	9,84879	-7,40287	5,29262	-3,43906	1,84980	-0,71546	0,14800
12	11,50808	-10,74609	9,34328	-7,91953	6,66959	-5,54500	4,34328	-3,27036	2,33714	-1,61333	0,96597	-0,44666	0,13500
13	5,32020	-4,61998	4,29145	-3,62118	3,01310	-2,67071	2,13992	-1,72147	1,22163	-0,93163	0,53317	-0,28989	0,11900
14	20,61945	-19,39682	16,80034	-14,14817	11,84307	-9,78712	7,73534	-5,77921	4,06200	-2,66324	1,49831	-0,59887	0,12600
15	30,02641	-28,42244	24,75314	-20,70178	16,98199	-13,72247	10,81050	-8,20966	5,94148	-3,90501	2,11507	-0,81306	0,16400
16	27,62370	-26,17896	22,93678	-19,42253	16,18997	-13,17171	10,19859	-7,42299	5,07437	-3,21481	1,73980	-0,67818	0,14000

Tableau A.1b/P.50 – Coefficients A_{ij} pour la voix artificielle de femme

<i>i</i>	<i>j</i>												
	0	1	2	3	4	5	6	7	8	9	10	11	12
1	8,92953	-8,28905	7,23150	-6,06571	5,06663	-4,16883	3,34820	-2,64174	1,91152	-1,27122	0,74358	-0,35347	0,13100
2	9,11050	-8,29868	7,05018	-6,03862	5,02156	-4,15784	3,37442	-2,70084	2,04257	-1,41928	0,90339	-0,47240	0,14100
3	13,69058	-12,75539	10,87390	-8,98976	7,28838	-5,87662	4,78110	-3,72852	2,71831	-1,81828	1,06757	-0,48907	0,12900
4	6,95118	-6,29134	5,35757	-4,64910	3,87250	-3,23132	2,69856	-2,12367	1,58326	-1,13885	0,70609	-0,37778	0,12800
5	2,73454	-1,80664	1,95283	-1,79464	1,43897	-1,31656	0,93268	-0,87398	0,53694	-0,47562	0,24159	-0,15438	0,11800
6	10,82358	-10,07808	8,78565	-7,43643	6,14765	-5,10041	4,10027	-3,23241	2,34220	-1,54676	0,91918	-0,45059	0,14200
7	21,58481	-19,92676	16,21532	-12,43566	9,61057	-7,93982	6,97363	-6,00866	4,68271	-3,12797	1,70804	-0,67499	0,14200
8	23,73912	-22,20897	18,74416	-15,03715	11,99248	-9,85513	8,27112	-6,72826	4,94335	-3,10450	1,60004	-0,61090	0,12900
9	4,97162	-4,27705	4,01380	-3,38500	2,78457	-2,45010	1,98057	-1,63020	1,18104	-0,80108	0,51528	-0,29138	0,12500
10	13,37598	-12,45509	10,72295	-8,97928	7,35893	-6,05438	4,88819	-3,86108	2,85164	-1,88876	1,11490	-0,52260	0,13800
11	16,48817	-15,22287	12,62125	-10,23900	8,46966	-7,23692	6,24648	-5,09752	3,77465	-2,46950	1,37404	-0,57453	0,13200
12	18,22041	-17,17540	15,09489	-12,55171	10,24976	-8,45903	6,71874	-5,19063	3,52021	-2,10167	1,08066	-0,41880	0,14300
13	1,32602	-0,31718	0,44277	-0,47070	0,32935	-0,41555	0,25775	-0,32079	0,13791	-0,23640	0,10103	-0,10136	0,11500
14	16,90640	-15,73723	13,30151	-10,82887	8,78690	-7,34521	6,21516	-5,11100	3,80281	-2,43990	1,33506	-0,55971	0,12900
15	21,73895	-20,42432	17,51117	14,44152	11,79131	-9,66735	7,90433	-6,19508	4,41275	-2,75545	1,46525	-0,59916	0,14000
16	21,04832	-19,72714	16,81197	-13,70183	11,07189	-9,12707	7,57941	-6,08064	4,40471	-2,74320	1,43897	-0,58079	0,13300

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 et protocole Internet
Série Z	Langages et aspects informatiques généraux des systèmes de télécommunication