



UNION INTERNATIONALE DES TÉLÉCOMMUNICATIONS

UIT-T

SECTEUR DE LA NORMALISATION
DES TÉLÉCOMMUNICATIONS
DE L'UIT

P.501

(08/96)

SÉRIE P: QUALITÉ DE TRANSMISSION
TÉLÉPHONIQUE

Appareils de mesures objectives

Signaux d'essai à utiliser en téléphonométrie

Recommandation UIT-T P.501

(Antérieurement Recommandation du CCITT)

RECOMMANDATIONS UIT-T DE LA SÉRIE P
QUALITÉ DE TRANSMISSION TÉLÉPHONIQUE

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.501

SIGNAUX D'ESSAI A UTILISER EN TELEPHONOMETRIE

Résumé

La présente Recommandation décrit les signaux d'essai qui s'appliquent à différents usages en téléphonométrie. Elle indique une grande variété de signaux d'essai en commençant par les signaux de faible complexité jusqu'aux signaux très complexes qui comportent de nombreux paramètres types caractérisant la parole. Outre les signaux techniques tels que les ondes sinusoïdales ou le bruit, des signaux plus proches de la parole y sont décrits.

La présente Recommandation spécifie les principes de la formation du signal pour chaque type de signal d'essai. Les caractéristiques propres telles que les spectres de densité de puissance, les fonctions de densité de probabilité ou les réponses du filtre de conformation y sont indiquées.

La présente Recommandation offre un aperçu général des applications types des signaux d'essai décrits. Cet aperçu est en fait une directive qui donne des règles d'application générales. Toutefois, la description détaillée des applications se trouve dans les différentes Recommandations qui indiquent les méthodes de mesure pour telle ou telle application.

Afin d'éviter les problèmes inhérents à la formation des signaux d'essai qui y sont décrits, la présente Recommandation est complétée par un CD-ROM où tous les signaux d'essai sont enregistrés.

Source

La Recommandation UIT-T P.501, élaborée par la Commission d'études 12 (1993-1996) de l'UIT-T, a été approuvée le 30 août 1996 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 (Helsinki, 1^{er}-12 mars 1993).

Dans certains secteurs de la technologie 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.

© UIT 1996

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

TABLE DES MATIÈRES

	Page
1	Champ d'application..... 1
2	Références..... 1
3	Définitions et abréviations 1
4	Récapitulatif des signaux d'essai et des applications types décrits dans la présente Recommandation 2
5	Types de signal d'essai..... 5
5.1	Signaux différents de la voix humaine (totalement artificiels)..... 5
5.1.1	Signaux déterministes..... 5
5.1.2	Signaux aléatoires..... 7
5.1.3	Combinaison de signaux aléatoires et déterministes 7
5.2	Signaux semblables à la voix humaine 8
5.2.1	Signaux de source composite (signaux composés en temps) 8
5.2.2	Bruit modulé semblable à la voix humaine 22
5.2.3	Signaux composés en fréquence (technique de la tonalité test)..... 22
5.2.4	Signaux composés complexes 23
5.3	Signaux vocaux..... 29
	Bibliographie..... 30

Recommandation P.501

SIGNAUX D'ESSAI A UTILISER EN TELEPHONOMETRIE

(Genève, 1996)

1 Champ d'application

La présente Recommandation décrit toute une gamme de signaux d'essai susceptibles d'être utilisés à des fins diverses en téléphonométrie, depuis des signaux de peu de complexité jusqu'à des signaux de grande complexité faisant intervenir un nombre élevé de paramètres types de la voix humaine. C'est ainsi qu'elle décrit, outre des signaux techniques comme les ondes sinusoïdales ou le bruit sinusoïdal, des signaux s'apparentant plus à la voix humaine.

En outre, elle présente les règles générales d'application type des signaux d'essai décrits, le détail des modalités d'application étant exposé dans les différentes Recommandations relatives aux méthodes de mesure correspondant aux différentes applications.

Pour éviter tout problème lors de la réalisation des signaux d'essai décrits, la présente Recommandation est complétée par un CD-ROM où tous les signaux d'essai sont stockés.

2 Références

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

- Recommandation UIT-T P.50 (1993), *Voix artificielle*.
- Recommandation UIT-T P.59 (1993), *Voix conversationnelle artificielle*.
- Recommandation UIT-T P.340 (1996), *Caractéristiques de transmission des postes téléphoniques mains-libres*.

3 Définitions et abréviations

La présente Recommandation définit les termes suivants et utilise les abréviations suivantes.

- 3.1 facteur de crête:** rapport puissance de crête/puissance efficace d'un signal.
- 3.2 signal de source composite (CSS, *composite source signal*):** signal composé en temps de différents éléments.
- 3.3 transformée de Fourier rapide (FFT, *fast fourier transform*)**
- 3.4 fonction de transfert de modulation (MTF, *modulation transfer function*):** signal de modulation, dérivé de l'enveloppe d'un signal d'essai.

3.5 processus markovien de modélisation vocale (MSMP, *markov speech model process*): signal de simulation vocale utilisant des chaînes de Markov préconditionnées pour la génération d'un signal semblable à la parole, tenant compte du processus réel de génération de la parole.

3.6 séquence de bruit pseudo-aléatoire (séquence PN): bruit pseudo-aléatoire au contenu fréquentiel défini, dérivé par transformée de Fourier inverse d'un spectre fréquentiel prédéterminé.

3.7 générateur de parole artificielle (SSG, *simulated speech generator*): signal présentant des propriétés proches de la voix humaine, constitué en tenant compte du processus de génération de la voix réelle.

3.8 indice de transmission de la parole (STI, *speech transmission index*): indice d'intelligibilité de la parole en particulier dans des conditions de réverbération, dérivé de la mesure de la fonction de transfert de modulation.

4 Récapitulatif des signaux d'essai et des applications types décrits dans la présente Recommandation

Systèmes linéaires, invariables par rapport au temps (par exemple, téléphones avec combiné)									
	Onde sinusoïdale (Note 1)	Bruit (Note 2)	CSS (Note 3)	Tonalité d'essai	SSG	Rec. P.50	Rec. P.59	MSMP	Voix
équivalents pour la sonie	x	x	x	x	x	x	x	x	x
réponses en fréquence	x	x	x	x	x	x	x	x	x
effet local pour l'auditeur/le locuteur	x	x	x	x	x	x	x	x	x
distorsion harmonique	x		x						
distorsion		x	x						
signaux hors bande	x		x						
mesures de niveau	x	x	x	x	x	x	x	x	x
mesures de propagation	x	x	x		(x)	(x)		(x)	(x)
mesures d'écho		x	x	x	x	x	x	x	x
x applicable (x) applicable avec certaines précautions NOTES 1 Y compris onde sinusoïdale modulée et spectres de Fourier. 2 Y compris bruit rose, bruit blanc et bruit commuté. 3 Y compris diverses combinaisons de son voisé et de signaux de mesure (séquence PN, sinusoidal, etc.).									

Systèmes non linéaires et/ou variables par rapport au temps									
	Onde sinusoïdale (Note 1)	Bruit (Note 2)	CSS (Note 3)	Tonalité test	SSG	Rec. P.50	Rec. P.59	MSMP	Voix
équivalents pour la sonie (valeurs sur longue durée)		(x)	(x)		(x)	(x)	(x)	(x)	(x)
équivalents pour la sonie (valeurs sur courte durée)		(x)	x						
réponses en fréquence (valeurs sur longue durée)		(x)	(x)		(x)	(x)	(x)	(x)	(x)
réponses en fréquence (valeurs sur courte durée)		(x)	x						
effet local pour l'auditeur/le locuteur (valeurs sur longue durée)		(x)	(x)	(x)	(x)	(x)	(x)	(x)	(x)
distorsion harmonique			(x)						
distorsion		(x)	(x)		(x)	(x)	(x)	(x)	(x)
signaux hors bande			(x)						
mesures de niveau		(x)	(x)		(x)	(x)	(x)	(x)	(x)
mesures de propagation		(x)	(x)						
mesures d'écho		(x)	(x)		(x)	(x)	(x)	(x)	(x)
<p>x applicable</p> <p>(x) applicable avec certaines précautions</p> <p>NOTES</p> <p>1 Y compris onde sinusoïdale modulée et spectres de Fourier.</p> <p>2 Y compris bruit rose, bruit blanc et bruit commuté ainsi que MTF.</p> <p>3 Y compris diverses combinaisons de son voisé et de signaux de mesure (séquence PN, sinusoïdal, etc.).</p> <p>4 Les valeurs sur longue durée correspondent au comportement établi des systèmes. Les valeurs sur courte durée correspondent au comportement dynamique des systèmes.</p>									

Téléphones mains libres									
	Onde sinusoïdale (Note 1)	Bruit (Note 2)	CSS (Note 3)	Tonalité test	SSG	Rec. P.50	Rec. P.59	MSMP	Voix
équivalents pour la sonie (valeurs sur longue durée)		x	x	(x)	x	x	x	x	x
équivalents pour la sonie (valeurs sur courte durée)		(x)	x						
réponses en fréquence (valeurs sur longue durée)		x	x	(x)	x	x	x	x	x
réponses en fréquence (valeurs sur courte durée)		(x)	x						
distorsion harmonique			x						
distorsion		(x)	x						
signaux hors bande			x						
mesures de niveau		x	x	(x)	x	x	x	x	x
mesures de propagation		x	x						
caractéristiques de commutation	(x)	(x)	x	(x)	(x)	(x)	(x)	(x)	(x)
mesures de réverbération		x	x						
caractéristiques du trajet d'écho		(x)	(x)		(x)	(x)		(x)	(x)
<p>x applicable</p> <p>(x) applicable avec certaines précautions</p> <p>NOTES</p> <p>1 Y compris onde sinusoïdale modulée et spectres de Fourier.</p> <p>2 Y compris bruit rose, bruit blanc et bruit commuté ainsi que bruit modulé (MTF).</p> <p>3 Y compris diverses combinaisons de son voisé et de signaux de mesure (séquence PN, sinusoïdal, etc.).</p> <p>4 Les valeurs sur longue durée correspondent au comportement établi des systèmes. Les valeurs sur courte durée correspondent au comportement dynamique des systèmes.</p>									

Annuleurs d'écho									
	Onde sinusoïdale (Note 1)	Bruit (Note 2)	CSS (Note 3)	Tonalité test	SSG	Rec. P.50	Rec. P.59	MSMP	Voix
mesures de niveau		(x)	x	(x)	x	x	x	x	x
mesures de propagation		(x)	x						
caractéristiques de commutation	(x)	(x)	x		(x)	(x)	(x)	(x)	(x)
performance avec bruit de fond		(x)	x	(x)	x	x	x	x	x
affaiblissement de l'écho, TCL		(x)	(x)	(x)	x	x	x	x	x
performance avec double conversation		(x)	x		(x)	(x)	(x)	(x)	(x)
x applicable (x) applicable avec certaines précautions NOTES 1 Y compris onde sinusoïdale modulée et spectres de Fourier. 2 Y compris bruit rose, bruit blanc et bruit commuté ainsi que bruit modulé (MTF). 3 Y compris diverses combinaisons de son voisé et de signaux de mesure (séquence PN, sinusoïdal, etc.).									

5 Types de signal d'essai

Différents signaux d'essai présentant un degré de complexité divers ont été évalués pour différents types d'application. On a pu déterminer, en fonction de leur complexité, les différents groupes de signaux suivants.

5.1 Signaux différents de la voix humaine (totalement artificiels)

Constituant les signaux de mesure classiques, ces signaux peuvent être divisés en signaux déterministes et en signaux aléatoires. Les signaux déterministes peuvent être définis par une formule qui les décrit complètement dans la dimension temporelle ou dans le domaine fréquentiel tandis que les signaux aléatoires peuvent être décrits par leurs statistiques et par leur spectre sur longue durée.

5.1.1 Signaux déterministes

5.1.1.1 Description

- onde sinusoïdale
- onde sinusoïdale modulée
- spectre généré par transformée de Fourier

Les signaux déterministes utilisés pour des mesures en téléphonométrie sont normalement décrits dans le domaine fréquentiel. La description générale d'un signal sinusoïdal qui pourrait éventuellement être modulé en amplitude ou en fréquence est donnée par la formule suivante:

$$s(t) = [A + \mu_{am} \cdot \cos(2\pi t \cdot f_{am})] \cdot \cos(2\pi t \cdot f_0 + \mu_{fm} \cdot \sin(2\pi t \cdot f_{fm})) \quad [5-1]$$

avec

- A amplitude du signal
- μ_{am} facteur de modulation de la modulation d'amplitude
- f_{am} fréquence de modulation de la modulation d'amplitude
- f_0 fréquence de la porteuse
- μ_{fm} facteur de modulation de la modulation de fréquence
- f_{fm} fréquence de modulation de la modulation de fréquence
- t temps

Le balayage fréquentiel linéaire est décrit par la formule:

$$s(t) = A(f_0 + st) \cdot e^{j(\pi t(f_0 + st) + \phi_0)} \quad [5-2]$$

avec

- s vitesse de balayage
- f_0 fréquence initiale

Le balayage logarithmique est décrit par la formule:

$$s(t) = A(f_0 \cdot 10^{t/Td}) \cdot e^{j((2\pi f_0 Td / \ln 10) \cdot 10^{t/Td} + \phi_0)} \quad [5-3]$$

avec

- Td durée de balayage d'une octave.

Un signal composé de plusieurs ondes sinusoïdales peut être décrit par son équivalent de Fourier et peut s'exprimer comme suit:

$$s(t) = \sum_n A_n \cdot \sin(2\pi t \cdot f_n + \phi_n) \quad [5-4]$$

avec

- A_n amplitude de la composante en fréquence n
- f_n fréquence de la composante en fréquence n
- ϕ_n phase de la composante en fréquence n
- t temps

Un signal de bruit pseudo-aléatoire, comme celui décrit au 5.2.1.1, peut être considéré comme une forme spéciale d'un signal se composant de plusieurs ondes sinusoïdales et adapté selon la transformée de Fourier rapide (FFT). Alors que les signaux sinusoïdaux discrets s'appliquent normalement à des niveaux fixés, indépendants de la fréquence, les signaux se composant de plusieurs ondes sinusoïdales et les signaux sinusoïdaux balayés s'appliquent souvent avec une pondération fréquentielle qui correspond mieux au spectre de la voix humaine.

5.1.1.2 Application

On peut toujours utiliser des signaux déterministes pour déterminer les caractéristiques de transfert d'un système linéaire invariable par rapport au temps principalement dans le domaine fréquentiel. Normalement, des signaux de ce type sont utilisés pour déterminer la distorsion harmonique et la distorsion d'intermodulation. L'avantage de ces signaux est qu'ils permettent de déterminer facilement les paramètres d'un système simplement en procédant à des mesures de niveau.

L'avantage du balayage logarithmique est que la résolution fréquentielle effective se rapproche davantage de la résolution fréquentielle de l'oreille humaine.

Le balayage linéaire permet de contrôler et de traiter le résultat des mesures directement dans le domaine fréquentiel ainsi que dans la dimension temporelle à l'aide, par exemple, de techniques FFT. En outre, et en particulier, il permet de supprimer le bruit parasite et d'isoler les échos électriques ou acoustiques.

Une attention particulière doit être apportée à tous les signaux discrets ou se composant de plusieurs ondes sinusoïdales destinés à être employés pour mesurer des dispositifs à l'aide de techniques adaptatives. En effet, la fonction d'autocorrélation du signal de mesure ne devrait pas périodiquement être à l'intérieur de la fenêtre de traitement du dispositif soumis à l'essai.

5.1.2 Signaux aléatoires

5.1.2.1 Description

Bruit erratique

Le bruit erratique peut être déterminé par ses caractéristiques statistiques, par le spectre de sa densité de puissance sur longue durée, par sa fonction de densité de probabilité unidimensionnelle et bidimensionnelle ou simplement en fonction du temps si le signal aléatoire est échantillonné. Les signaux suivants sont le plus souvent utilisés en téléphonométrie:

Bruit blanc

- Caractéristiques fréquentielles: les limites extrêmes du spectre de densité de puissance généré sont définies par l'application et le spectre de densité de puissance sur longue durée est le plus souvent décrit.
- Densité de probabilité: distribution normalement gaussienne avec un facteur de crête de $11 \text{ dB} \pm 1 \text{ dB}$.

Bruit rose

- Caractéristiques fréquentielles: le spectre de densité de puissance du signal montre une diminution de 3 dB par octave; les limites extrêmes du spectre de densité de puissance généré sont définies par les applications; le spectre de densité de puissance sur longue durée est le plus souvent décrit.
- Densité de probabilité: distribution normalement gaussienne avec un facteur de crête de $11 \text{ dB} \pm 1 \text{ dB}$.

Pour être utilisés en téléphonométrie, ces signaux sont souvent modulés par un signal de modulation de type rectangulaire (modulation ON/OFF). Les paramètres de modulation types sont: ON pendant 250 ms et OFF pendant 150 ms. La modulation type de la voix humaine est ainsi simulée selon un procédé très simple.

5.1.2.2 Application

On utilise couramment des signaux aléatoires pour des systèmes linéaires invariables par rapport au temps afin de déterminer des niveaux large bande ou des niveaux dans des bandes d'octaves fractales. En outre, ils peuvent être utilisés pour déterminer les caractéristiques de transfert dans le domaine fréquentiel, comme la réponse en fréquence ou les équivalents pour la sonie (*LR, loudness rating*).

Quand on utilise des signaux aléatoires aux fins de mesure, les temps d'intégration doivent toujours être longs (normalement $>10 \text{ s}$).

5.1.3 Combinaison de signaux aléatoires et déterministes

5.1.3.1 Description

En modulant du bruit erratique avec un signal déterministe, on peut obtenir la fonction de transfert de modulation (MTF) d'un système. Si l'on prend, par exemple, un signal de bruit d'une intensité

moyenne $\bar{I}(t)$ et on le module avec un signal sinusoïdal de fréquence f selon un indice de modulation $m=1$, on obtient un signal:

$$I(t) = \bar{I}(t) \cdot (1 + \cos(2\pi ft)) \quad [5-5]$$

5.1.3.2 Application

Si on applique ce signal à un système, il est possible de mesurer la fonction de transfert de modulation (MTF) [2] en mesurant l'indice de modulation $m(f)$ du signal de sortie du système. En mesurant la MTF et en l'interprétant judicieusement, il est possible dans une certaine mesure de prédire le degré d'intelligibilité de la voix humaine (voir 5.2.2).

5.2 Signaux semblables à la voix humaine

Les signaux semblables à la voix humaine qui existent revêtent différents degrés de complexité, celle-ci étant toujours liée aux paramètres types de la voix simulée par lesdits signaux, lesquels peuvent se subdiviser dans les catégories ci-dessous.

Tous les signaux semblables à la voix humaine sont constitués essentiellement – à côté d'autres paramètres – de sons voisés et non voisés. En règle générale, il faut qu'un téléphone ou tout autre dispositif de traitement de la voix utilisant des détecteurs de courants vocaux soit activé ou soit maintenu en activité en présence de ces sons voisés ou non voisés alors que d'autres signaux peuvent empêcher ou causer une interruption de la transmission.

5.2.1 Signaux de source composite (signaux composés en temps)

5.2.1.1 Description

a) Généralités

Lors de la mise au point du signal de source composite, trois éléments ont été jugés particulièrement importants:

- signal voisé pour reproduire les propriétés de la voix humaine;
- signal déterministe pour mesurer les fonctions de transfert, sans erreurs statistiques, avec spectre de la densité de puissance constant pour le signal d'excitation dans le domaine fréquentiel à mesurer;
- "signal" de pause, produisant une modulation d'amplitude.

Les caractéristiques suivantes sont obtenues:

- i) courte période de mesure;
- ii) possibilité d'injection simultanée du signal d'essai dans les directions du locuteur et de l'auditeur (opération duplex).

Le principe fondamental de l'utilisation d'un tel signal d'essai est de mettre le dispositif soumis à l'essai dans un état défini et reproductible pendant la période de mesure et de garantir que ses fonctions de transfert ne seront pas modifiées pendant les mesures proprement dites (signal quasi stationnaire). A cette fin, le signal de source composite est composé des éléments suivants:

- 1) signal voisé produit par le signal de "voix artificielle" conforme à la Recommandation P.50

Le signal voisé du CSS doit activer d'éventuels détecteurs de parole dans des systèmes commandés par des signaux vocaux. Si le signal voisé a été choisi, c'est parce qu'il est prévisible qu'à l'avenir tous les dispositifs de transmission de la parole répondront rapidement à un tel son. Ce signal doit activer le dispositif soumis à l'essai pour la direction de transmission à mesurer. Comme la durée, le commencement et la fin du signal voisé sont

connus précisément, ce signal peut également être utilisé pour mesurer le temps de commutation pour la direction de transmission à l'essai. Le temps de commutation et le temps de propagation de tout le système peuvent être déterminés conformément à la Recommandation P.340 au moyen de la forme du signal dans la dimension temporelle. La durée du signal est de 50 ms;

2) signal de bruit pseudo-aléatoire

Le signal de mesure est le signal de bruit pseudo-aléatoire injecté après le son voisé de la parole artificielle. Il peut avoir certaines caractéristiques d'un signal de bruit. L'amplitude du signal transformé selon la formule de Fourier est constante dans le domaine de fréquence, mais sa phase change. Aux fins de mesures, seule l'amplitude de la fonction de transfert présente habituellement un intérêt; la phase n'est pas aussi importante, mais elle peut cependant être définie elle aussi.

Le signal est produit comme suit:

Pour commencer, un spectre complexe est produit dans le domaine de fréquence selon l'équation suivante:

$$H(k) = W(k) \cdot e^{j i_k \cdot \pi} ; k = -M/2, \dots, M/2, \text{ sans } 0; i_k \in \{+1, 0\}, i_k = -i_{-k} \text{ aléatoire} \quad [5-6]$$

L'indice M est adapté à la taille de la FFT choisie, par exemple 2048 points. La formule indique que la puissance du spectre complexe produit est constante pour toutes les fréquences si W(k) est égal à 1 pour toutes les fréquences, alors que la phase peut être π ou 0 pour chaque fréquence, correspondant à une séquence aléatoire. Cependant, afin d'obtenir une pondération différente dans le domaine de fréquence, W(k) peut facilement être ajusté de façon à produire différents spectres pendant la durée de la séquence de bruit pseudo aléatoire. Ce spectre sera alors transposé dans la dimension temporelle au moyen de la transformée de Fourier inverse et produira le signal suivant:

$$S(n) = \frac{1}{M} \sum_{k=-M/2, k \neq 0}^{M/2} H(k) \cdot e^{j 2\pi \cdot n \cdot k / M} , n = -M/2, \dots, M/2 - 1; \quad [5-7]$$

Ainsi, le signal produit est limité dans le temps (selon la longueur choisie pour la transformée de Fourier) et adapté précisément à la taille de la FFT choisie. Si on désire une séquence de temps plus longue, le signal peut être répété. Cette méthode permet d'obtenir des séquences temporelles de n'importe quelle longueur.

La durée de ce signal de mesure est d'environ 200 ms si M, le taux d'échantillonnage et le nombre de répétitions sont choisis de façon appropriée.

NOTE 1 – Normalement, pour des systèmes dotés de paramètres très variables par rapport au temps – par exemple, systèmes utilisant des techniques de compression-extension – la durée de la FFT devrait être brève pour permettre d'obtenir une bonne estimation, en peu de temps, de la fonction de transfert variable par rapport au temps. Pour des systèmes incorporant des techniques adaptatives, tels qu'annuleurs d'écho ou de bruit, la valeur de M devra peut-être être plus élevée (proche des 200 ms de la durée du signal) pour que la fonction d'autocorrélation du signal de mesure ne soit pas périodiquement à l'intérieur de la fenêtre de traitement du dispositif soumis à l'essai.

NOTE 2 – En lieu et place des signaux de bruit pseudo-aléatoires, on peut éventuellement utiliser pour des applications spéciales d'autres signaux, tels que des séquences M (séquences de longueur maximale) ou d'autres séquences pourvues de fonctions d'autocorrélation parfaites. Pour d'autres applications, comme par exemple les mesures de distorsion, le signal PN peut être remplacé par des signaux appropriés, comme des ondes sinusoïdales ou du bruit bande étroite.

3) Pause

La pause a deux buts. Il est nécessaire de faire une pause initiale avant d'appliquer tout signal de mesure afin de ramener les systèmes à fonctions de transfert variables dans le temps vers un état initial défini. A cet effet, la pause devrait être aussi longue que possible (>1 s). Si, toutefois, le système doit être mis dans un état activé permanent (flot de paroles continu), les pauses intermédiaires doivent être plus courtes (environ 100 ms) afin de produire une modulation d'amplitude adaptée au signal de source composite.

La pause de la séquence CSS doit être comprise entre 100 et 150 ms.

NOTE 3 – La pause peut être plus longue dans le cas de mesures où, après la mise en marche, le phénomène à observer doit être de longue durée; dans pareil cas, elle peut durer plusieurs secondes selon les besoins des mesures.

Pour obtenir une séquence de longue durée dépourvue de décalage la séquence CSS répétée doit être inversée en amplitude (déplacement de phase de 180°).

b) *Calcul et analyse utilisant un signal de source composite*

Lorsqu'on utilise un signal de source composite pour effectuer des mesures, il est possible de répéter la séquence signal voisé – signal pseudo-aléatoire – pause, c'est-à-dire qu'après la pause la séquence recommence par un signal voisé. Il est ainsi possible de produire des séquences de n'importe quelle longueur.

1) Principe d'étalonnage acoustique et électrique, niveaux des signaux d'essai

Après avoir créé une séquence selon la procédure mentionnée ci-dessus, il est possible d'utiliser ce signal comme un signal de mesure standard, par exemple comme un bruit rose commuté. L'étalonnage en niveaux (acoustique et électrique) s'effectue en utilisant la totalité de la séquence comprenant les signaux voisés, les séquences de bruit pseudo-aléatoire et les pauses. En principe, il est possible d'utiliser un dispositif de mesure standard du rapport puissance de crête/puissance efficace, d'une largeur de bande de 20 kHz, fonctionnant avec une intégration "rapide". Toutefois, la méthode préférée consiste à recourir, pour les calculs de niveaux, à une analyse par FFT, avec les paramètres de calcul suivants:

- le taux d'échantillonnage doit être conforme à celui choisi pour la génération du signal (de préférence 44,1 kHz ou 48 kHz);
- la durée de la FFT doit être conforme à celle choisie pour la génération du signal;
- le fenêtrage doit être de type rectangulaire;
- absence de chevauchement;
- intégration pendant la **totalité de la séquence (répétée)**, c'est-à-dire signaux voisés – séquences de bruit pseudo-aléatoire – pauses;
- calcul du niveau à partir du spectre de densité de puissance dérivée du calcul par FFT (intégration des niveaux sur tous les composants fréquentiels).

2) Paramètres d'analyse

Pour la mesure des fonctions de transfert, pour les directions de réception et d'émission, pour les équivalents pour la sonie, etc., la séquence signal voisé – séquence de bruit pseudo-aléatoire – pause est là aussi répétée. Le niveau de la séquence complète est corrigé de telle façon que le niveau total mesuré corresponde à celui qui a été prescrit et qui est décrit ci-dessus.

Toutes les mesures (ou analyses) s'effectuent **uniquement pendant la séquence de bruit pseudo-aléatoire**. Concernant l'analyse de tous les paramètres de transmission dans le domaine fréquentiel et les équivalents pour la sonie, le signal mesuré et transformé selon la formule de Fourier doit toujours être rapporté au signal d'entrée, utilisant les mêmes

paramètres d'analyse. Dans le cas de mesures acoustiques, ces signaux d'entrée sont mesurés au MRP en vue de mesurer les caractéristiques d'émission. Dans le cas de mesures électriques, le signal mesuré et transformé selon la formule de Fourier est rapporté au signal d'entrée transmis sur le trajet de transmission soit numérique soit analogique, ce qui peut être fait à l'aide d'une mesure sur deux canaux ou de techniques permettant de stocker le signal d'entrée analysé (signal de mesure mesuré au MRP ou à interface numérique). Aux fins de l'analyse, on utilise les paramètres suivants:

- une fréquence d'échantillonnage conforme à celle choisie pour la génération du signal;
- une durée de la FFT conforme à celle utilisée pour la génération du signal, appliquée uniquement à la séquence de bruit pseudo-aléatoire;
- un fenêtrage de type rectangulaire;
- un chevauchement autorisé entre 0 et 99,9%. Le même chevauchement doit être appliqué pour le signal de mesure à l'entrée de l'objet soumis à l'essai (MRP ou interface numérique) et pour le signal mesuré à la sortie de la direction d'émission ou de réception de l'objet soumis à l'essai;
- rapporter le signal, transformé selon la formule de Fourier et mesuré soit à la sortie de la direction d'émission ou de réception, au signal transformé selon la formule de Fourier au point d'excitation correspondant (MRP, entrée numérique ou analogique).

Dans d'autres cas, il faut éventuellement disposer de signaux sinusoïdaux ou de signaux de bruits différents pour mesurer différents paramètres, par exemple la distorsion. Dans pareil cas, la séquence de bruit pseudo-aléatoire est remplacée par le signal correspondant, par exemple par un signal sinusoïdal ou par un signal de bruit bande étroite. Les niveaux des signaux de source composite (séquence complète répétée), comprenant les différents types de signal de mesure, sont calculés suivant la procédure décrite ci-dessus. Les mesures sont effectuées à l'aide du signal de mesure compris dans le signal de source composite et des calculs décrits dans la norme correspondante.

5.2.1.2 Réalisation pratique d'un signal de source composite en vue de mesures jusqu'à 20 kHz

a) Signal voisé destiné à simuler les propriétés de la voix humaine

La durée du signal est de 48,62 ms. Pendant cet intervalle, n'importe quel détecteur de parole devrait avoir reconnu le signal vocal et activé le système. Le signal voisé peut se décrire comme une séquence de mots de 16 bits d'une fréquence d'échantillonnage de 44,1 kHz.

Le Tableau 1 de 134 mots doit être répété 16 fois pour activer le système soumis à l'essai pendant une durée de 48,62 ms.

TABLEAU 1/P.501

Echantillons (valeurs ASCII) de la partie 1 du CSS (à lire par colonnes)

-76	2098	3116	2930	2392	1824	1306	-3462	-7492	-2806	-626
112	2244	3158	2866	2410	1772	1170	-4024	-6414	-2844	-456
298	2360	3180	2808	2430	1742	968	-4590	-5334	-2888	-298
472	2456	3180	2764	2444	1750	702	-5154	-4428	-2898	-130
628	2538	3168	2728	2460	1760	394	-5716	-3772	-2846	
776	2626	3146	2686	2472	1762	76	-6298	-3360	-2698	
916	2730	3132	2632	2452	1736	-244	-6912	-3128	-2460	
1068	2824	3122	2572	2398	1684	-594	-7556	-3002	-2166	
1234	2904	3108	2496	2300	1624	-968	-8194	-2924	-1846	
1398	2964	3096	2432	2178	1572	-1384	-8719	-2870	-1544	
1572	2996	3076	2382	2068	1516	-1846	-8998	-2830	-1274	
1752	3032	3038	2362	1976	1460	-2356	-8898	-2800	-1032	
1932	3072	2992	2368	1892	1390	-2898	-8378	-2792	-818	

b) Signal de bruit pseudo-aléatoire

Les paramètres correspondant à la séquence de bruit pseudo-aléatoire sont:

fréquence d'échantillonnage de 44,1 kHz, mots de 16 bits, longueur de la transformée de Fourier de 2048 points.

$$H(k) = \begin{cases} W(k) \cdot e^{j i_k \pi}; & k = -928, \dots, +928 \text{ sans } 0, i_k \{+1, 0\}, \text{aléatoire } i_k = -i_{-k} \\ 0 & \text{autre} \end{cases} \quad [5-8]$$

Selon la formule [5-8], le signal en temps est calculé suivant la transformée de Fourier inverse. Cette séquence est répétée 4307 fois pour obtenir une durée de 200 ms destinée à la séquence de mesure de bruit pseudo-aléatoire. Le facteur de crête de la séquence de bruit pseudo-aléatoire est de 11 dB \pm 1 dB.

Suivant la résolution en fréquence de 21,533 Hz (44,1 kHz/2048), il existe 928 valeurs de FFT dans la gamme de fréquences comprise entre 0 et 20 kHz. Chaque valeur $W(k)$ est égale à 152 680 et est calculée de telle sorte que les niveaux dans une largeur de bande de 20 kHz soient les mêmes pour le signal voisé et pour la séquence de bruit pseudo-aléatoire.

c) Pause

La pause est utilisée suivant les modalités exposées dans la description générale du signal de source composite. La durée de la pause est de 101,38 ms pour que le signal dure exactement 350 ms.

NOTE – Par un suréchantillonnage ou un sous-échantillonnage approprié, il est possible d'obtenir d'autres fréquences d'échantillonnage pour la séquence décrite. Le filtre d'interpolation utilisé pour le suréchantillonnage ou le sous-échantillonnage doit se rapprocher d'un filtre idéal rectangulaire. L'affaiblissement de réjection doit être supérieur à 60 dB, et l'ondulation dans la bande passante inférieure à $\pm 0,2$ dB.

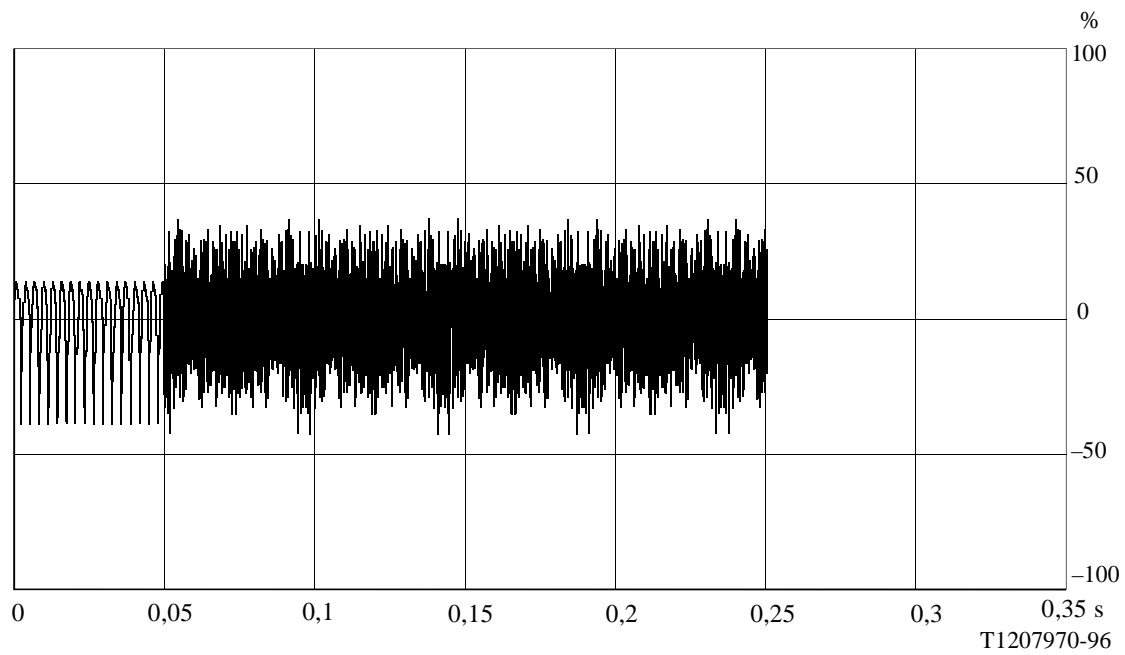


FIGURE 1/P.501
Signal de source composite, signal en temps

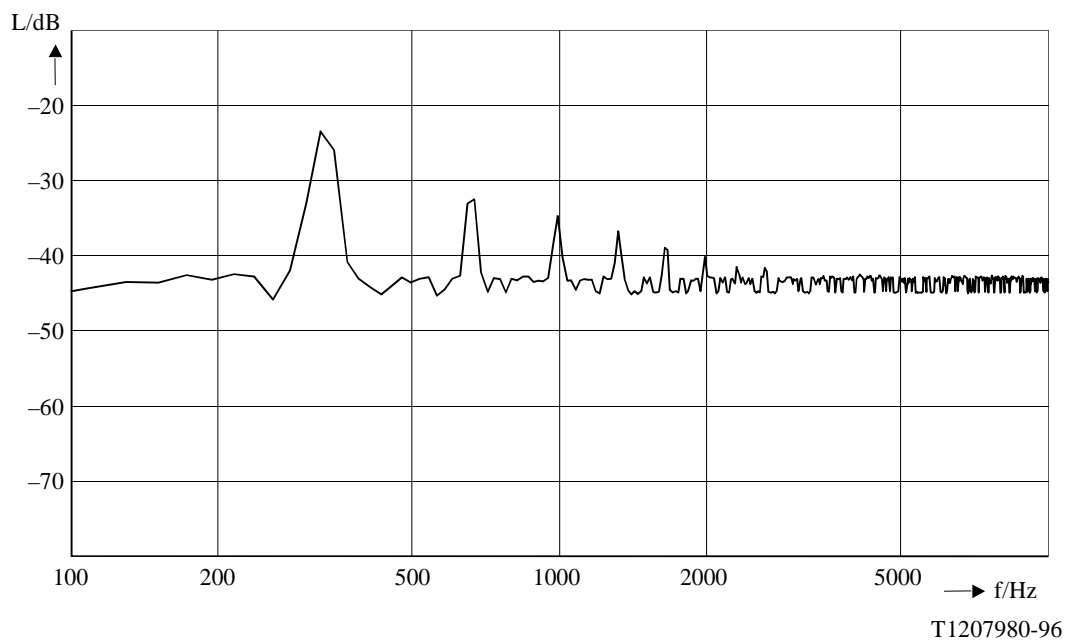


FIGURE 2/P.501
**Spectre de densité de puissance du signal de source composite
 (fenêtre d'analyse de Hanning)**

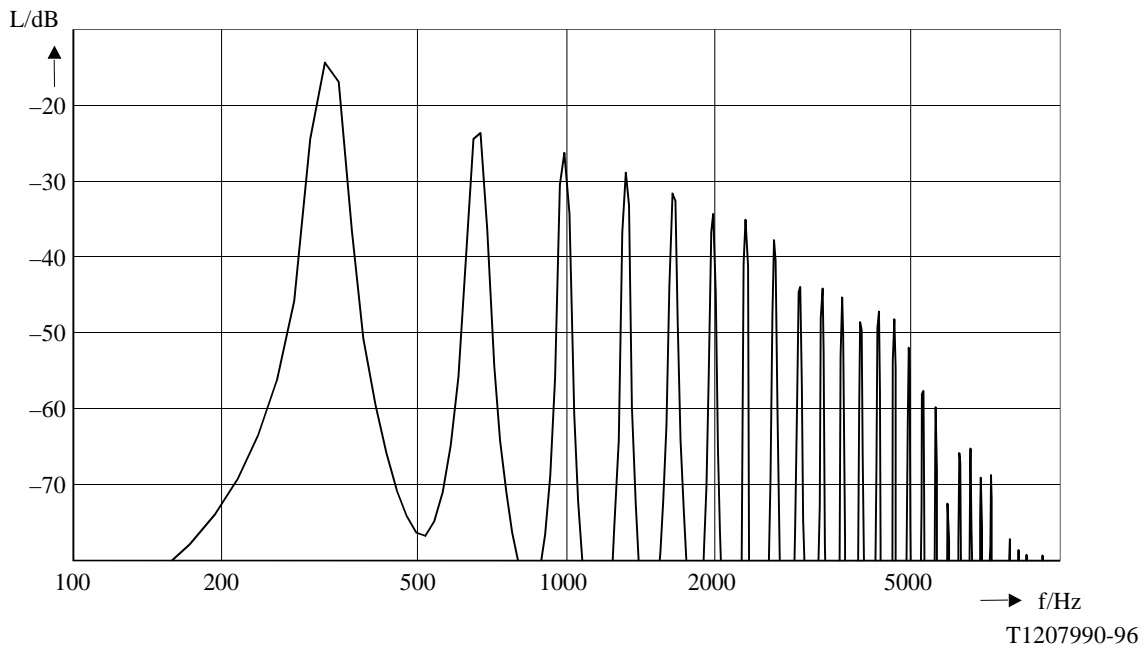


FIGURE 3/P.501
Spectre de densité de puissance du signal voisé
(fenêtre d'analyse de Hanning)

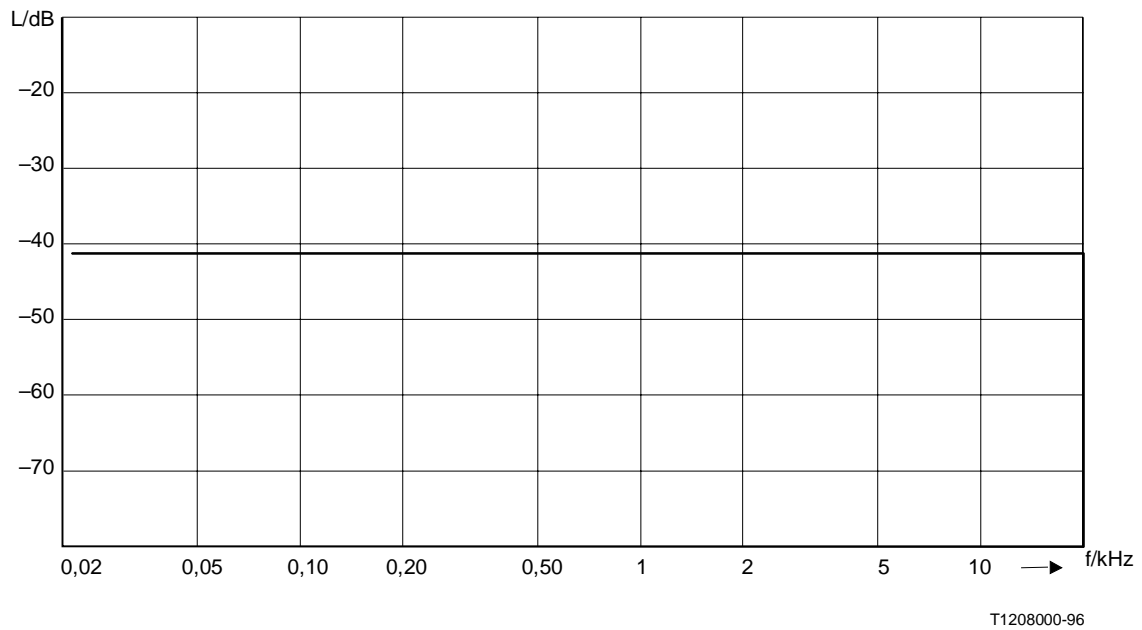


FIGURE 4/P.501
Spectre de densité de puissance de la séquence de bruit pseudo-aléatoire
(fenêtre d'analyse de caractéristique rectangulaire)

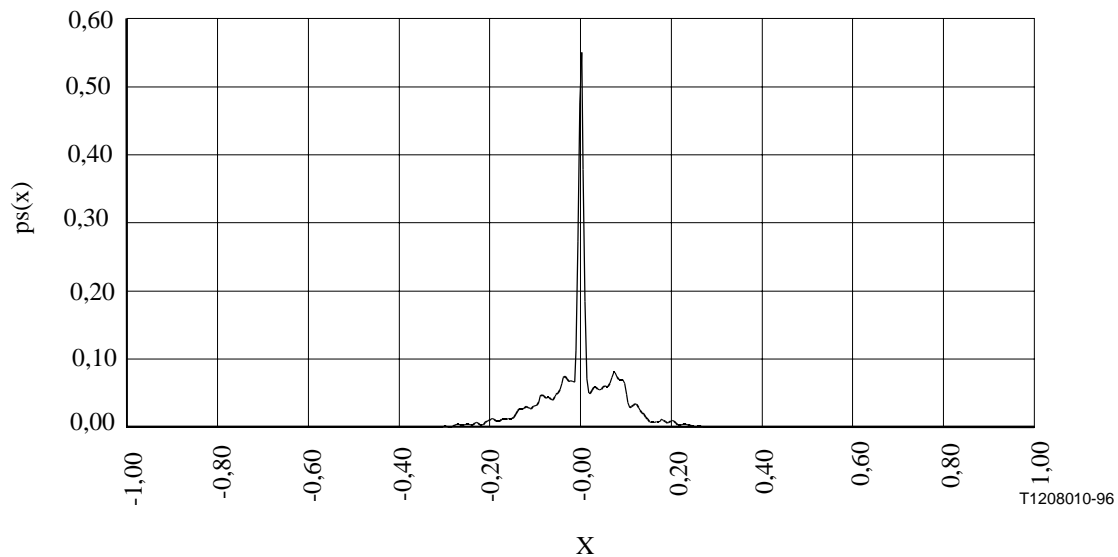


FIGURE 5/P.501

**Fonction de densité de probabilité du signal de source composite
suivant 5.2.1.2**

5.2.1.3 Application

Le signal décrit ci-dessus peut être appliqué à des systèmes se comportant de façon non linéaire et variable par rapport au temps, mais pouvant être considérés, pendant la brève durée de la mesure, comme des systèmes quasi stationnaires. Il est possible de déterminer dans le domaine fréquentiel des paramètres comme la réponse en fréquence, les équivalents en sonie, etc., ainsi que dans la dimension temporelle des paramètres tels que les instants de mise sous tension. Ces mêmes paramètres peuvent être également déterminés lorsqu'un signal de mesure de distorsion est inséré en lieu et place de la séquence de bruit pseudo-aléatoire (signaux sinusoïdaux ou bruit bande étroite).

Le signal de source composite constitue, d'un point de vue général, une classe de signaux. Si pour une application particulière il faut accroître la partie du signal voisé, il est éventuellement possible d'en répéter la séquence jusqu'à ce que l'on obtienne la longueur du signal requis. La même procédure peut être appliquée à la séquence de bruit pseudo-aléatoire et à la pause. Si l'on souhaite procéder à des applications particulières de ce type, il conviendrait de décrire dans la documentation correspondante la procédure de composition du signal.

Dans le cas de systèmes adaptifs dont les propriétés de transmission changent en fonction des caractéristiques du signal, l'utilisation d'un signal faiblement corrélé présente un intérêt. Pour ces systèmes, la durée de la transformée de Fourier doit être portée à environ 200 ms (par exemple, FFT de 8192 points au lieu de 2048 points), les paramètres de génération et d'analyse du signal devant être corrigés en conséquence: $k = -3715, \dots, 3715$ aléatoire sans 0.

5.2.1.4 Signal de source composite limité à une bande et pourvu d'un spectre de densité de puissance semblable à la voix humaine

5.2.1.4.1 Description

a) *Signal de source composite pour le cas où un seul locuteur parle*

1) Signal voisé à bande limitée

Le Tableau 1 donne les valeurs ASCII du signal voisé décrit au 5.2.1.2, limité dans la bande comprise entre 200 Hz et 3,6 kHz. Compte tenu d'une fréquence d'échantillonnage de 44,1 kHz, les 134 valeurs ASCII ont une durée de 3,04 ms. Le Tableau 2 doit être lu en colonnes:

TABLEAU 2/P.501

Valeurs ASCII du signal voisé limité à une bande donnée

-76	2098	3116	2930	2392	1824	1306	-3462	-7492	-2806	-626
112	2244	3158	2866	2410	1772	1170	-4024	-6414	-2844	-456
298	2360	3180	2808	2430	1742	968	-4590	-5334	-2888	-298
472	2456	3180	2764	2444	1750	702	-5154	-4428	-2898	-130
628	2538	3168	2728	2460	1760	394	-5716	-3772	-2846	
776	2626	3146	2686	2472	1762	76	-6298	-3360	-2698	
916	2730	3132	2632	2452	1736	-244	-6912	-3128	-2460	
1068	2824	3122	2572	2398	1684	-594	-7556	-3002	-2166	
1234	2904	3108	2496	2300	1624	-968	-8194	-2924	-1846	
1398	2964	3096	2432	2178	1572	-1384	-8719	-2870	-1544	
1572	2996	3076	2382	2068	1516	-1846	-8998	-2830	-1274	
1752	3032	3038	2362	1976	1460	-2356	-8898	-2800	-1032	
1932	3072	2992	2368	1892	1390	-2898	-8378	-2792	-818	

Les valeurs du signal voisé dans la gamme de fréquences comprise entre 200 Hz et 3,6 kHz sont là aussi calculées de telle sorte que la valeur du rapport puissance de crête/puissance efficace du signal voisé et la séquence de bruit pseudo-aléatoire soient égales. Cette dernière est répétée 16 fois pour obtenir une durée de 48,62 ms.

2) Signal de bruit pseudo-aléatoire généré à l'aide d'une FFT de 2048 points

Les paramètres correspondant à la séquence de bruit pseudo-aléatoire sont:

fréquence d'échantillonnage de 44,1 kHz, mot de 16 bits, longueur de la transformée de Fourier de 2048 points.

$$H(k) = \begin{cases} W(k) \cdot e^{j i_k \pi}; & k = -928, \dots, +928 \text{ sauf } 0, i_k \{+1, 0\}, \text{ aléatoire } i_k = -i_{-k} \\ 0 & \text{autre} \end{cases} \quad [5-9]$$

Conformément à la formule [5-7] ci-dessus, le signal en temps est calculé par transformée de Fourier inverse. La séquence est répétée 4307 fois pour obtenir une durée de 200 ms pour la séquence de mesure de bruit pseudo-aléatoire. Le facteur de crête de la séquence de bruit pseudo-aléatoire est de 11 dB \pm 1 dB.

Pour une résolution en fréquence de 21,5 Hz (44,1 kHz/2048), il existe 928 valeurs FFT dans la gamme de fréquences comprise entre 0 et 20 kHz. Chaque valeur $W(k)$ (avant filtrage) est égale à 152 680 et est calculée de telle sorte qu'à l'intérieur d'une largeur de bande de 20 kHz les niveaux du signal voisé et de la séquence de bruit pseudo-aléatoire soient identiques.

3) Signal de bruit pseudo-aléatoire généré à l'aide d'une FFT à 8192 points

Conformément à la formule [5-7] ci-dessus, le signal en temps est calculé par transformée de Fourier inverse. La séquence est répétée 1077 fois pour obtenir une durée de 200 ms pour la séquence de mesure de bruit pseudo-aléatoire. Le facteur de crête de la séquence de bruit pseudo-aléatoire est de $11 \text{ dB} \pm 1 \text{ dB}$.

Etant donné une résolution en fréquence de 5,4 Hz (44,1 kHz/8192), il existe 3715 valeurs FFT dans la gamme de fréquences comprise entre 0 et 20 kHz. Chaque valeur $W(k)$, avant filtrage, est égale à 305 360 et est calculée de telle sorte qu'à l'intérieur d'une largeur de bande de 20 kHz les niveaux du signal voisé et de la séquence de bruit pseudo-aléatoire soient identiques.

Pour obtenir la même valeur de rapport puissance de crête/puissance efficace pour la séquence de bruit pseudo-aléatoire limitée à une bande, il faut appliquer la fonction de filtrage indiquée dans la Figure 6. Le filtre est choisi de telle sorte que les niveaux de la séquence de bruit pseudo-aléatoire filtrée et non filtrée soient égaux.

NOTE 1 – Par un suréchantillonnage ou un sous-échantillonnage approprié, il est possible d'obtenir d'autres fréquences d'échantillonnage pour la séquence décrite. Le filtre d'interpolation utilisé pour ce suréchantillonnage ou ce sous-échantillonnage doit se rapprocher d'un filtre idéal rectangulaire. L'affaiblissement de réjection de bande doit être supérieur à 60 dB et l'ondulation dans la bande passante inférieure à $\pm 0,2 \text{ dB}$.

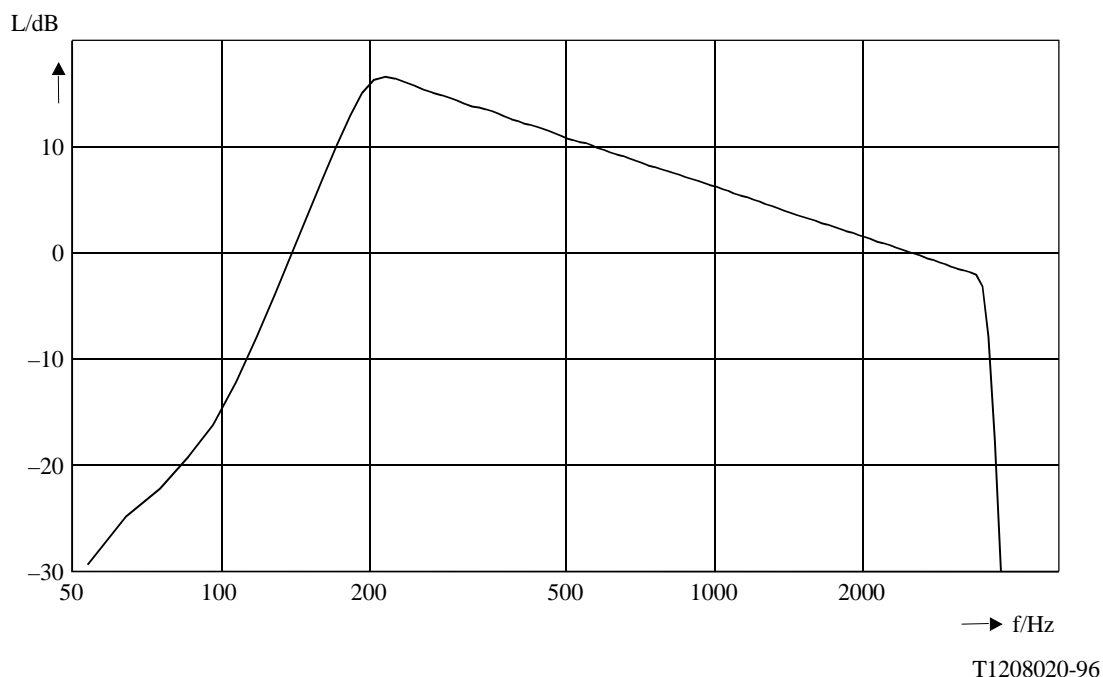


FIGURE 6/P.501

Fonction de transfert du filtre pour la bande limitant la séquence de bruit pseudo-aléatoire

TABLEAU 3/P.501

Tableau des fréquences centrales de filtrage

50 Hz	100 Hz	200 Hz	215 Hz	500 Hz	1 kHz	2,85 kHz	3,6 kHz	3,66 kHz	3,68 kHz
-25,8 dB	-12,8 dB	17,4 dB	17,8 dB	12,2 dB	7,2 dB	0 dB	-2 dB	-20 dB	-30 dB

Pour des systèmes adaptifs, comme les annuleurs d'écho, il est éventuellement préférable d'utiliser une séquence de bruit pseudo-aléatoire plus longue pour ne pas avoir de signaux de mesure corrélés à l'intérieur de la fenêtre d'adaptation. Pour ces systèmes, la longueur de la FFT doit être portée à 8192 points lorsqu'on utilise une fréquence d'échantillonnage de 44,1 kHz comme il est indiqué ci-dessus.

b) *Signal de source composite à bande limitée à double locuteurs*

La séquence de double parole est générée de la même façon que le signal correspondant au cas où un seul locuteur parle, mais les durées du signal voisé et de la pause sont légèrement différentes pour obtenir une situation de double parole type: application de deux signaux au même instant, présence d'un signal uniquement dans un seul canal, présence de signaux voisés dans les deux directions et présence de signaux voisés et de signaux non voisés au même instant dans les différents canaux. La corrélation entre le signal correspondant au cas d'un locuteur unique et le signal de double parole est faible. On choisit pour ce faire un signal voisé différent d'une hauteur fondamentale (*pitch frequency*) différente ainsi qu'un signal de bruit aléatoire en lieu et place de la séquence de bruit pseudo-aléatoire. La durée du signal voisé est de 72,69 ms, celle du signal de bruit aléatoire de 200 ms et celle de la pause de 127,31 ms, soit une durée totale de 400 ms.

1) Signal voisé

Pour le signal voisé dans le cas de la double parole, on a choisi une fréquence fondamentale différente de celle du signal voisé correspondant au cas du locuteur unique. Les valeurs du signal voisé dans le cas de la double parole sont données dans le Tableau 4. Le niveau de ce son est là encore identique à celui du signal correspondant au cas du locuteur unique. Compte tenu d'une fréquence d'échantillonnage de 44,1 kHz, on obtient 229 valeurs ASCII, représentant 5,19 ms. Le Tableau 4 ci-dessous doit être lu en colonnes.

TABLEAU 4/P.501

Valeurs ASCII du signal voisé limité à une bande dans le cas de double parole

-198	1146	-8292	4827	5853	1422	-1293	-810	-690	-1052	-621
-112	871	-8715	5094	5715	1224	-1302	-793	-724	-1043	-560
-9	560	-9077	5344	5560	1026	-1293	-767	-767	-1043	-509
103	233	-9370	5594	5387	819	-1267	-741	-793	-1052	-457
233	-121	-9542	5827	5215	603	-1250	-698	-819	-1060	-397
388	-491	-9542	6043	5043	388	-1233	-672	-845	-1060	-345
543	-871	-9361	6215	4879	181	-1224	-638	-853	-1060	-276
724	-1250	-8956	6344	4732	9	-1224	-603	-871	-1052	-207
896	-1638	-8327	6413	4586	-181	-1224	-595	-879	-1034	-112
1060	-2043	-7465	6422	4439	-328	-1224	-586	-888	-1017	
1233	-2465	-6396	6379	4276	-448	-1215	-595	-896	-991	
1388	-2896	-5163	6310	4086	-543	-1198	-603	-922	-957	

TABLEAU 4/P.501

Valeurs ASCII du signal voisé limité à une bande dans le cas de double parole

1517	-3345	-3827	6215	3870	-629	-1172	-621	-948	-931	
1638	-3819	-2448	6120	3629	-707	-1129	-629	-974	-905	
1747	-4310	-1103	6051	3370	-784	-1077	-938	-1009	-888	
1810	-4810	155	6000	3086	-871	-1026	-638	-1026	-862	
1845	-5319	1293	5991	2801	-948	-974	-638	-1052	-845	
1845	-5836	2241	5991	2534	-1026	-922	-638	-1069	-819	
1802	-6353	3034	6000	2267	-1112	-888	-638	-1077	-793	
1707	-6853	3655	6008	2034	-1181	-871	-638	-1069	-767	
1569	-7353	4138	5991	1819	-1241	-845	-647	-1060	-724	
1379	-7836	4517	5939	1612	-1276	-828	-664	-1060	-672	

Pour obtenir la durée requise de 72,69 ms, il faut répéter 14 fois les valeurs.

2) Bruit erratique

Le bruit erratique est choisi comme étant un bruit blanc de type gaussien limité à 20 kHz. Le facteur de crête du signal est égal à 12 ± 1 dB. La valeur du rapport puissance de crête/puissance efficace du bruit erratique limité dans une bande est choisie de telle sorte qu'elle est identique à celle du signal voisé.

Pour limiter le bruit erratique entre 200 Hz et 3,6 kHz, on utilise la fonction de filtrage indiquée dans la Figure 6, ce qui donne la même valeur de rapport puissance de crête/puissance efficace que pour le bruit erratique limité dans une bande.

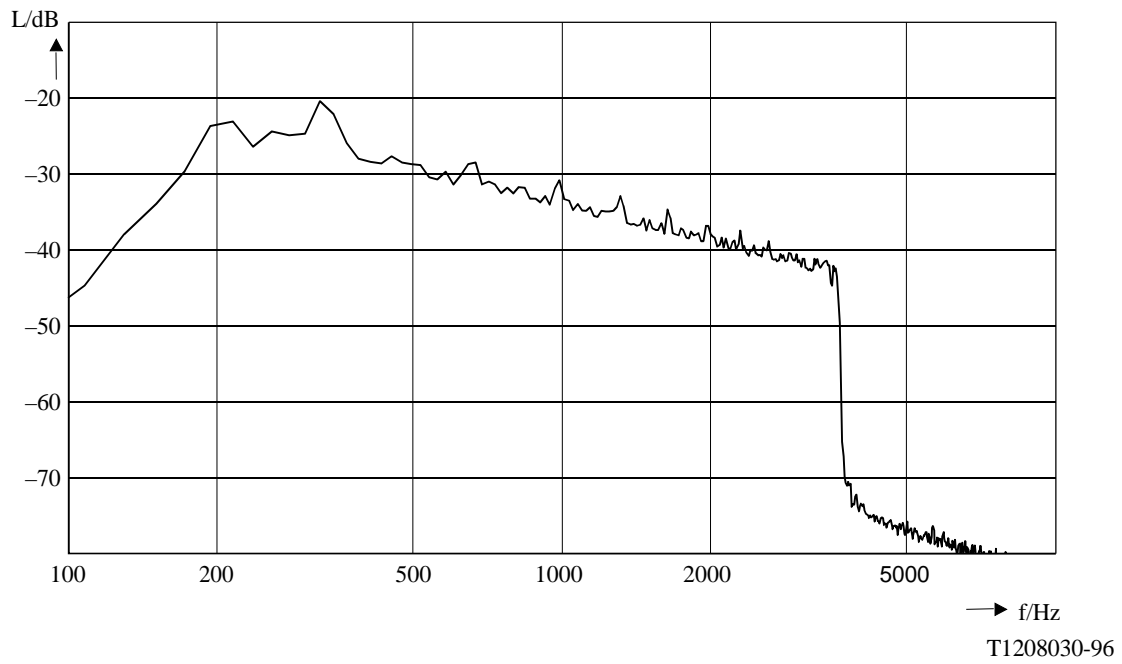


FIGURE 7/P.501

Spectre de densité de puissance du signal de source composite limité dans une bande (signal correspondant au cas du locuteur unique, fenêtre d'analyse de Hanning)

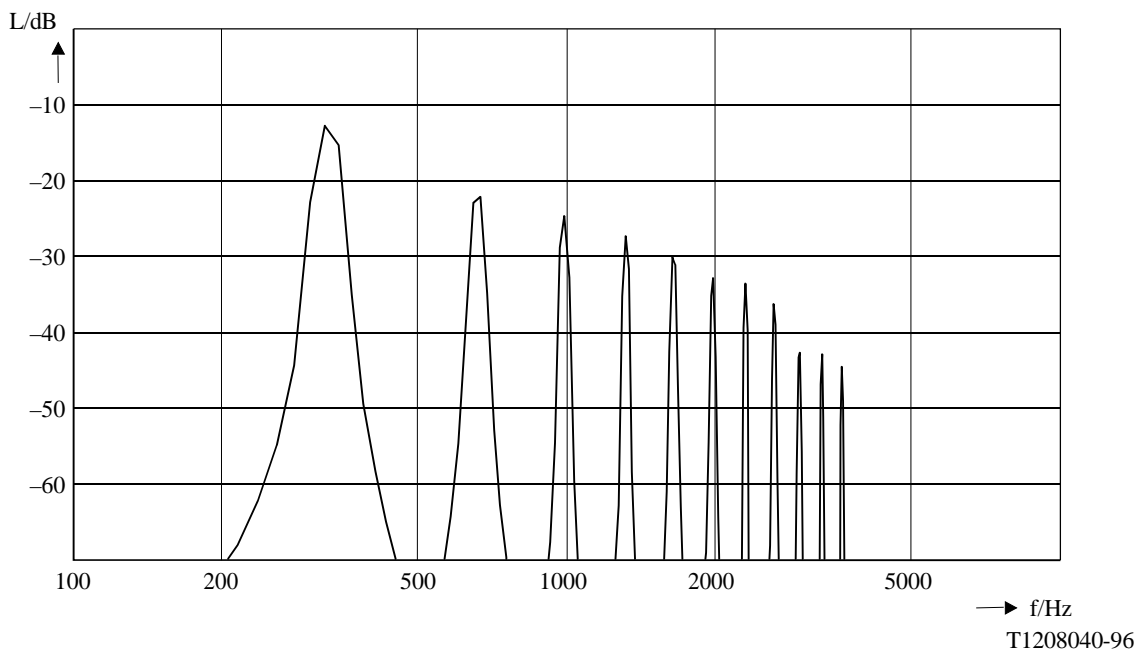


FIGURE 8/P.501

Spectre de densité de puissance du signal voisé limité dans une bande (signal correspondant au cas du locuteur unique, fenêtre d'analyse de Hanning)

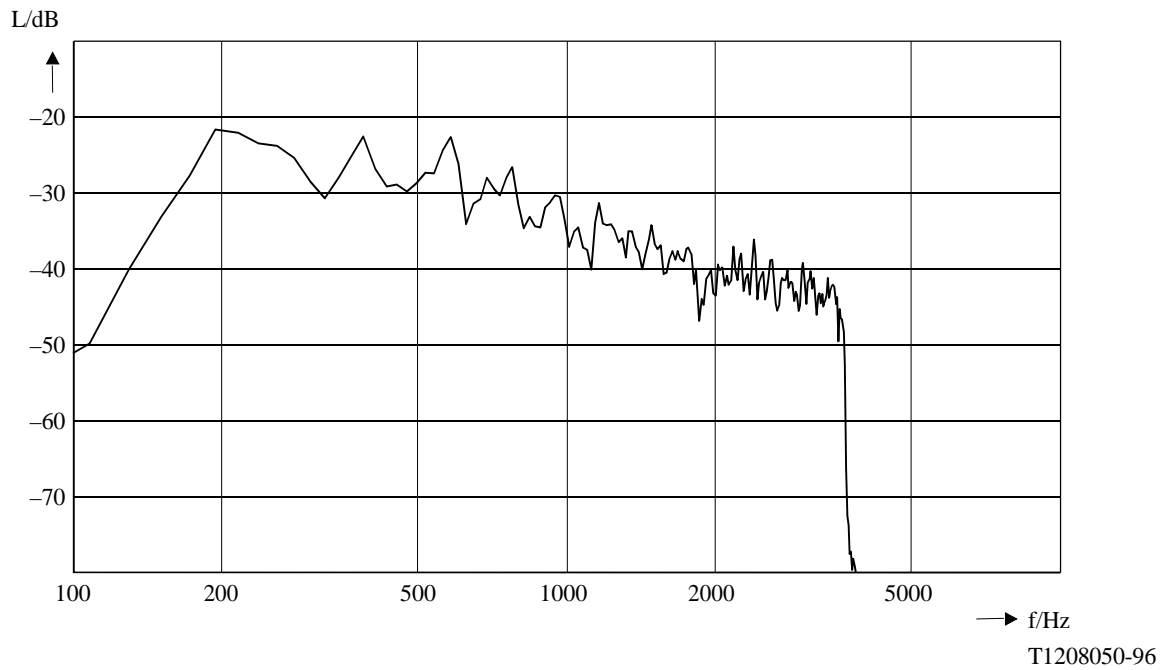


FIGURE 9/P.501

Spectre de densité de puissance du signal de source composite limité dans une bande et correspondant au cas de la double parole (fenêtre d'analyse de Hanning)

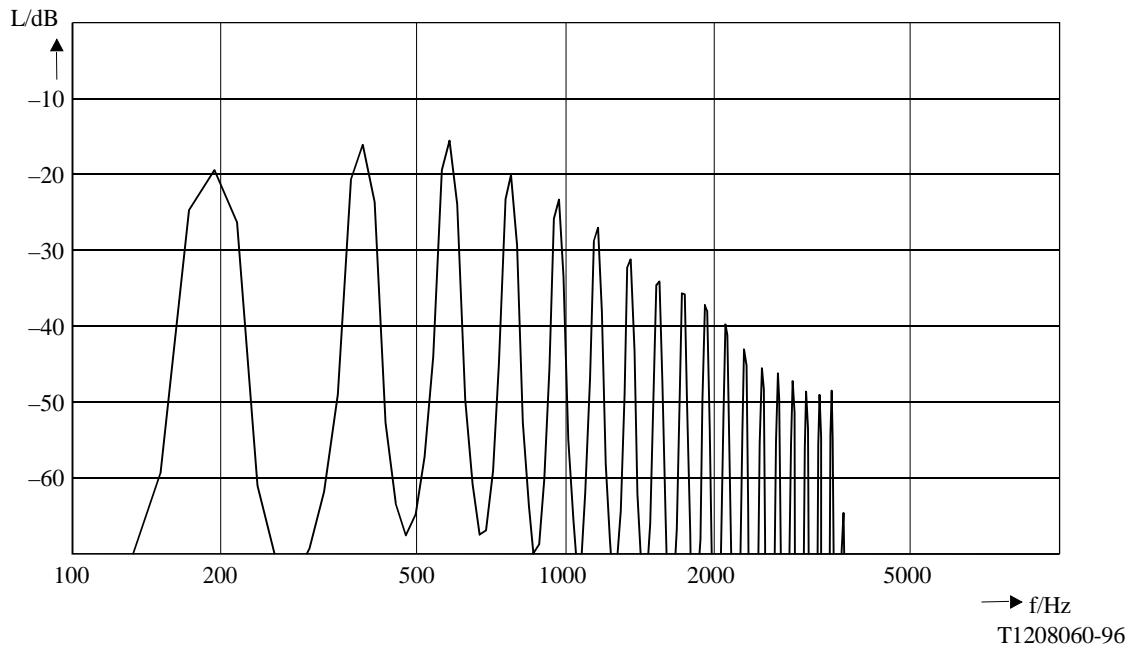


FIGURE 10/P.501

Spectre de densité de puissance du signal voisé limité dans une bande correspondant au cas de la double parole (fenêtre d'analyse de Hanning)

NOTE 2 – En procédant à un échantillonnage en amont ou en aval approprié, il est possible d'obtenir d'autres fréquences d'échantillonnage pour la séquence décrite. Le filtre d'interpolation utilisé pour cette opération d'échantillonnage en amont et en aval devrait se rapprocher d'un filtre idéal de caractéristique rectangulaire. L'affaiblissement d'élimination de bande devrait être supérieur à 60 dB et l'ondulation de la bande passante inférieure à $\pm 0,2$ dB.

5.2.1.4.2 Application

L'application des signaux de source composite limités dans une bande donnée aux deux cas du locuteur unique et de la double parole convient à toutes les mesures consacrées à des systèmes limités dans une bande, fonctionnant de manière non linéaire et variable par rapport au temps et exigeant le spectre de densité de puissance de longue durée type de signaux vocaux. L'application type correspond à la mesure d'anneaux d'écho de la voix humaine dans le réseau. Pour toutes les mesures unidirectionnelles, on utilisera le signal de source composite limité dans une bande correspondant au cas du locuteur unique; pour des mesures effectuées dans des conditions de double parole, on utilisera le signal correspondant dans les deux directions alors que le signal pour le locuteur unique sera transmis en direction de l'extrémité la plus lointaine.

5.2.2 Bruit modulé semblable à la voix humaine

5.2.2.1 Description

Comme il est indiqué au 5.1.3, la MTF peut être utilisée pour mesurer l'intelligibilité de la parole d'un système. En modulant un bruit filtré sur une octave, il est possible d'obtenir la MTF de différentes octaves. En pondérant correctement les indices de modulation de chaque octave, il est possible d'obtenir, sur différentes fréquences de modulation, un indice de transmission de la parole ($0 \leq STI \leq 1$) dont la corrélation à l'intelligibilité de la parole d'un système est élevée. Cet indice STI peut se mesurer à l'aide d'un signal se composant d'un certain nombre de bandes de bruit simultanément modulées. Le spectre de densité de puissance sur longue durée est choisi de telle sorte qu'il soit égal au spectre de densité de puissance de la parole. En appliquant une modulation idoine, on crée un signal qui reflète les caractéristiques temporelles d'une conversation continue. L'indice STI s'est révélé être un bon outil de prédiction de l'intelligibilité de la parole pour une large gamme de distorsions.

5.2.2.2 Application

L'indice de transmission de la parole (STI) peut être utilisé pour mesurer l'intelligibilité de la parole qui est altérée par les distorsions suivantes:

- bruit;
- filtrage dans la bande passante;
- écrêtage des pointes et de façon plus générale une vaste catégorie de distorsions non linéaires;
- commande automatique de gain;
- réverbération.

5.2.3 Signaux composés en fréquence (technique de la tonalité test)

5.2.3.1 Description

Pour déterminer les caractéristiques de transmission de systèmes téléphoniques dynamiquement variables, il est éventuellement nécessaire d'appliquer un signal de conditionnement (semblable à la voix humaine) idoine simultanément avec un signal d'essai analytique approprié. Il est donc indispensable que:

- le signal analytique soit appliqué à un niveau auquel son incidence sur le comportement dynamique du téléphone soumis à l'essai soit insignifiant. Pour satisfaire cette condition, il faut éventuellement procéder à une intégration longue afin d'obtenir une précision de mesure suffisante. Pour des systèmes variables dynamiquement, cette opération conduit à une caractéristique de transfert "moyenne" qui ne tient pas compte des effets de courte durée;
- la corrélation entre le signal de conditionnement et le signal d'essai analytique soit maintenue à un niveau minimum. Cette opération peut souvent s'effectuer par une simple séparation spectrale des deux signaux.

En règle générale, la relation entre l'état réel du système soumis à l'essai et le signal de mesure n'est pas évidente parce que celui-ci n'est pas corrélé au signal d'activation.

5.2.3.2 Application

Le plus souvent, cette méthode est utilisée pour déterminer les caractéristiques moyennes (sur longue durée) d'un système. Si par exemple on doit mesurer la réponse en fréquence moyenne dans des conditions d'exploitation réalistes (présence de réverbération et de bruit), on peut recourir éventuellement à une série de balayages sinusoïdaux linéaires. Souvent le signal analytique sera une tonalité unique qui pourra elle aussi être éventuellement utilisée, par exemple pour mesurer la variation de gain temporelle produite par le signal de conditionnement et mesurée à la fréquence de la tonalité unique.

5.2.4 Signaux composés complexes

5.2.4.1 Générateur de parole artificielle (SSG)

5.2.4.1.1 Description

1) Description générale

Pour générer un signal dont la répartition des amplitudes s'approche de celle de la parole, on module un signal principal de répartition gaussienne par un signal de modulation réalisé spécialement, comme il est indiqué dans la Figure 11. Le signal résultant est mis en forme de telle façon qu'il se rapproche du spectre de fréquence sur longue durée de la parole, comme il est indiqué dans la Figure 12.

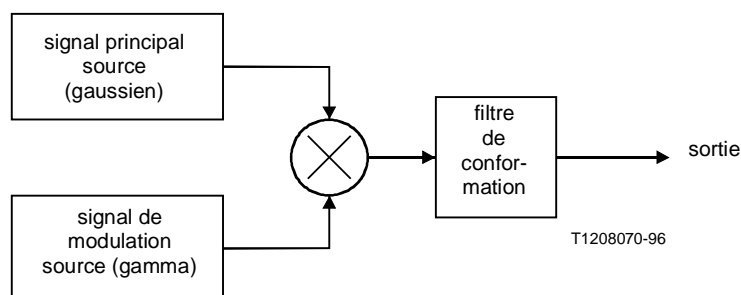


FIGURE 11/P.501

Schéma fonctionnel d'un générateur de parole artificielle

2) Signal principal

Le signal principal se compose de huit segments de bruit pseudo-aléatoire de 1024 points. Chaque segment possède le même spectre d'amplitude, mais un spectre de phase différent, la phase étant rendue aléatoire à l'intérieur des segments et entre eux de façon uniforme entre 0 et 360° afin de rendre aléatoire l'interaction entre les produits d'intermodulation des composantes spectrales en relation harmonique. Les segments ont chacun une durée de 80 ms et sont fusionnés les uns avec les autres par l'intermédiaire d'une fenêtre à cosinus biaisé avec insertion entre eux d'un segment de

liaison supplémentaire de 80 ms. La fermeture en fondu du segment précédent et l'ouverture en fondu simultanée du segment suivant éliminent les phénomènes transitoires qui se produiraient aux frontières des segments. Le signal principal complet se compose donc de huit segments pseudo-aléatoires avec intercalage de huit segments de liaison, d'une durée de 80 ms chacun, d'où une durée totale de 1,28 s. La conformation en fréquence à laquelle il est souhaitable de procéder pour s'approcher d'un spectre de parole moyen est fournie par un simple filtre disposé à la sortie.

3) *Signal de modulation*

Les mesures montrent qu'une distribution gamma de paramètre $m = 0,545$ constitue une bonne approximation de la répartition des amplitudes instantanées d'une conversation continue. Les caractéristiques syllabiques peuvent être représentées par la réponse d'un filtre passe-bas qui est pratiquement plate jusqu'à environ 4 Hz (pouvant être considérée comme le point -3 dB) pour ensuite décroître de -6 dB par octave.

La forme d'onde finale du signal de modulation a été dérivée empiriquement de la distribution gamma. En faisant varier la période de cette impulsion d'une manière pseudo-aléatoire et en ajustant son rapport temps de montée/temps de descente, on obtient une approximation satisfaisante du spectre de l'enveloppe de modulation de la voix réelle.

4) *Signal combiné*

Pour allonger le temps de répétition du signal final et pour étaler de façon plus égale les maxima du signal de modulation sur la séquence répétée du signal de type gaussien, on a choisi comme rapport entre les fréquences d'horloge d'échantillonnage des deux signaux le rapport de 4/255; c'est ainsi que la fréquence d'horloge du signal principal est de 12 800 Hz et celle du signal de modulation d'environ 200,8 Hz. Les durées de répétition sont de: 1,28 s pour le signal de type gaussien, 10,2 s pour le signal de modulation et 326,4 s pour le signal modulé final.

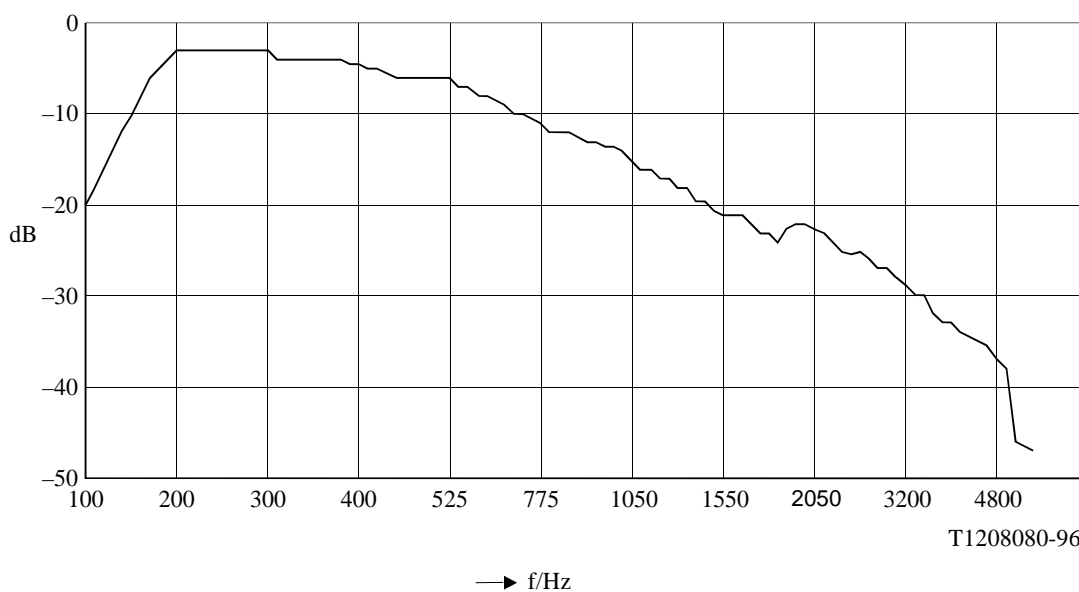


FIGURE 12/P.501

Spectre généré par le simulateur de parole artificielle

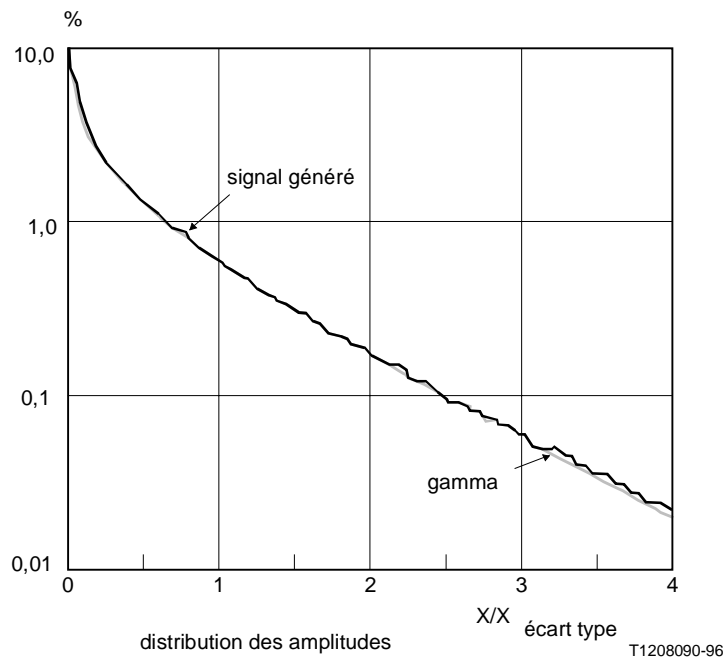


FIGURE 13/P.501

Distribution des amplitudes du simulateur de parole artificielle

5.2.4.1.2 Application

Les signaux SSG peuvent éventuellement être utilisés lorsqu'on a besoin pour la mesure d'une séquence de signaux vocaux "type". En effet, comme le signal préconisé dans la Recommandation P.50, le signal SSG représente la parole en générant des paramètres types de la voix naturelle selon un processus défini. Par rapport à la voix réelle aucune langue spécifique, aucune voix spécifique ne sont simulées.

En règle générale, il faut disposer de temps d'intégration longs (>10 s) si l'on veut calculer à partir des mesures des paramètres tels que la réponse en fréquence ou les équivalents en sonie.

5.2.4.2 Voix artificielle UIT-T P.50

Le signal composé, complexe, semblable à la voix humaine le plus utilisé en téléphonométrie est le signal de voix artificielle, décrit dans la Recommandation P.50.

5.2.4.3 Voix conversationnelle artificielle UIT-T P.59

La Recommandation P.59 décrit un signal de voix conversationnelle artificielle, composé et complexe, offrant des salves de paroles, etc., ainsi qu'une séquence de double parole.

5.2.4.4 Modélisation de la parole suivant des chaînes discrètes de Markov (MSMP)

5.2.4.4.1 Description

1) Généralités

Les paragraphes ci-après présentent brièvement la version actuelle d'un procédé de modélisation de la parole suivant une technique MSMP (Markov speech-model process, procédé de modélisation de la parole suivant des chaînes de Markov) proposée comme signal d'essai pour des applications large bande dans le traitement des signaux vocaux. Ce signal d'essai est en fait une extension du modèle connu utilisé pour des applications bande étroite (MSIRP) [3]. Comme des publications [3] et [4]

l'ont mentionné, certains écarts ont été constatés entre, d'une part, le spectre sur longue durée et le spectre de l'enveloppe du signal du MSIRP et, d'autre part, ceux d'une conversation téléphonique naturelle (300 Hz - 3400 Hz). Par suite des améliorations apportées au modèle, le spectre de l'enveloppe du signal du MSIRP et celui du MSMP sont mieux adaptés à ceux de la parole en bande étroite ou en large bande.

2) *Procédure de génération du processus de modélisation de la parole suivant les chaînes de Markov (MSMP)*

La procédure de génération du MSMP est présentée dans le schéma fonctionnel de la Figure 14. Comme il y est indiqué, le MSMP est obtenu par le produit d'un processus de type gaussien $n(t)$ et d'un processus $s(t)$; il est ainsi possible de faire varier séparément la fonction de densité de probabilité (PDF, *probability density function*) en amplitude et l'autocorrélation du processus consécutif.

La variation dans le temps des propriétés de ce processus est commandée par des chaînes de Markov préconditionnées, ce qui permet d'obtenir des structures naturelles de formant et de hauteur des sons. Quant à la détermination de la nature voisée (v) ou non voisée (uv) d'une trame de 20 ms de durée, elle est effectuée par la chaîne de Markov (mc) qui fixe la valeur de la hauteur des sons dans cette trame (mc-pitch). Cette chaîne de Markov préconditionnée produit une séquence naturelle de 33 valeurs différentes de hauteur des sons, dont l'une est égale à 0 Hz, ce qui indique que la trame en cours est non voisée. Suivant ce processus de détermination, la chaîne de Markov se généralise et produit une séquence naturelle d'indices qui permet de choisir l'un des 50 filtres de formants correspondant à cette trame (mc-formant). Cette chaîne de Markov généralisée (mc-formant) fonctionne comme la partie cachée d'une HMM, qui produit une séquence naturelle de termes de gain et indique l'énergie sur courte durée du processus de production (mc-energy).

La branche supérieure détermine la PDF du processus qui s'ensuit. Un filtre passe-bas, excité par un bruit blanc aléatoire, pondéré, de type gaussien, produit un processus aléatoire lentement variable. Les facteurs de pondération sont obtenus à partir de mc-energy et sont constants pendant chaque trame. A ce stade, il nous faut signaler une propriété particulière des processus de production: leur PDF en amplitude est en effet généralement symétrique, alors que la conversation naturelle a une PDF en amplitude asymétrique. On utilise en conséquence deux mappages non linéaires différents pour obtenir la PDF souhaitée du processus $s(t)$ et selon le signe du processus $n(t)$ on passe d'un résultat à l'autre de ces mappages non linéaires; on obtient ainsi un processus aléatoire pourvu de la PDF en amplitude souhaitée, PDF formée de telle façon qu'en multipliant le processus $s(t)$ par un processus de type gaussien on obtient la PDF souhaitée de la voie naturelle.

La branche inférieure pilote l'autocorrélation du processus $x(t)$. Dans le cas d'une trame non voisée, le filtre de formants est excité par un processus non corrélé de type gaussien caractérisé par une variance moyenne et unitaire nulle. Dans le cas de trames voisées, l'excitation pourrait être modélisée par des impulsions périodiquement espacées. Il faut tenir compte du fait que l'application d'une fréquence de crête (pitch) provoque une structure quasi périodique dans le processus résultant, d'où une contradiction *a priori* étant donné que ce processus doit être aléatoire. On peut apporter une solution de compromis en utilisant un filtre en peigne doté de la fonction de transfert

$$H(z) = \alpha / (1 - a_k z^{-k_0})$$

On obtient ainsi un processus de type gaussien dont le spectre possède une structure quasi linéaire comme l'excitation du filtre de formants pendant les trames voisées. Il est à noter que la voix naturelle n'a de son côté qu'une structure en lignes spectrales approximative, qui correspond bien à la structure spectrale du produit du filtre en peigne. Les distances entre ces lignes spectrales représentent la fréquence de crête (pitch), qui est déterminée par k_0 . La valeur de k_0 est modifiée par mc-pitch, mais est constante pendant chaque trame voisée. La commutation entre régions voisée et non voisée s'effectue en deux étapes, pour obtenir une transition régulière; c'est ainsi que le facteur

de netteté a_k du filtre en peigne utilisé est égal à 0,6 pour les première et dernière trames de chaque région et de 0,95 pour tous les autres cas. Enfin, les filtres de formants sont déterminés par un jeu de 50 filtres en treillis du 16ème ordre. Les coefficients sont empruntés à un tableau de références qui est optimisé en fonction de l'algorithme généralisé de Lloyd [5]. Pour assurer le passage harmonieux entre les filtres de formants, les coefficients de filtrage utilisés sont actualisés toutes les 2 ms par interpolation linéaire entre les deux ensembles de coefficients de filtrage de trames voisines.

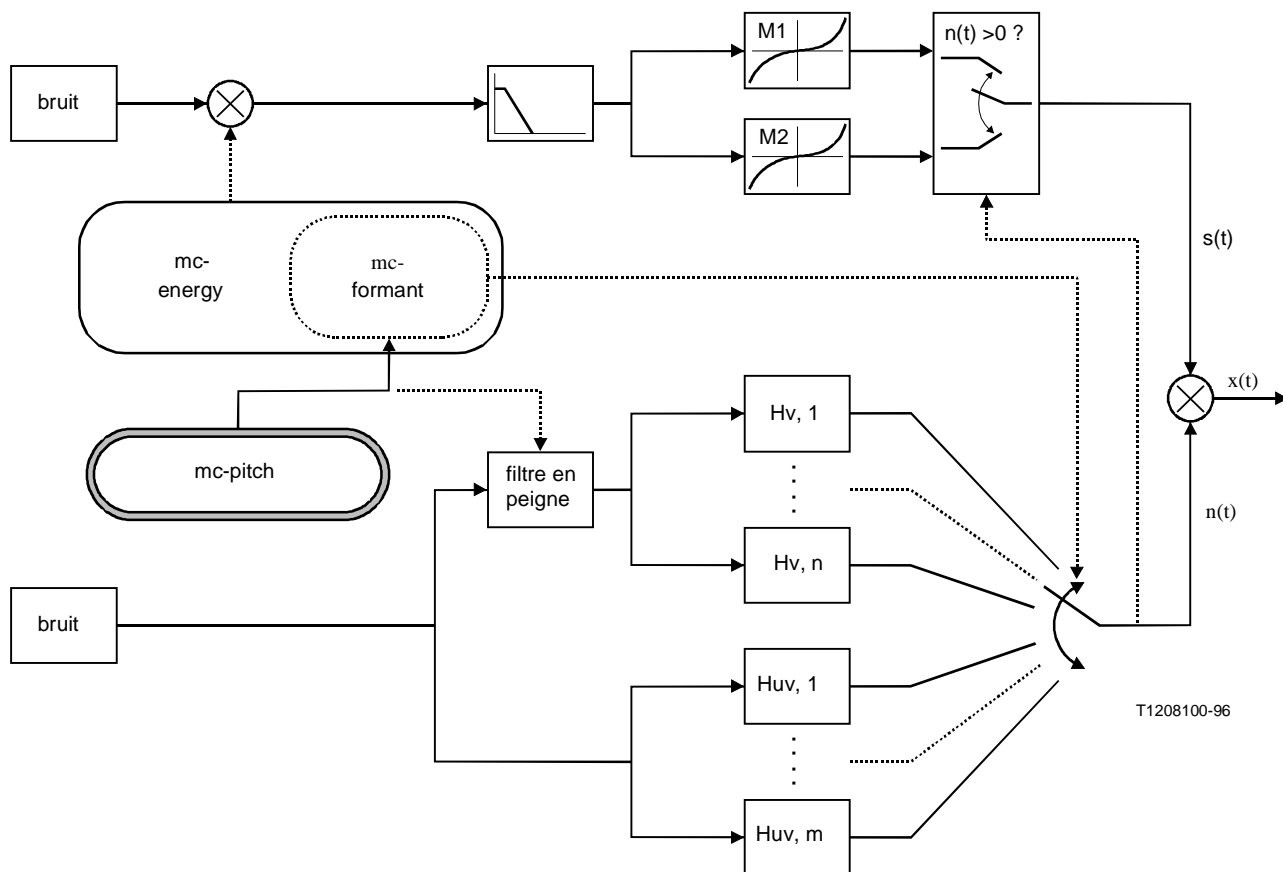


FIGURE 14/P.501
Schéma fonctionnel du générateur MSMP

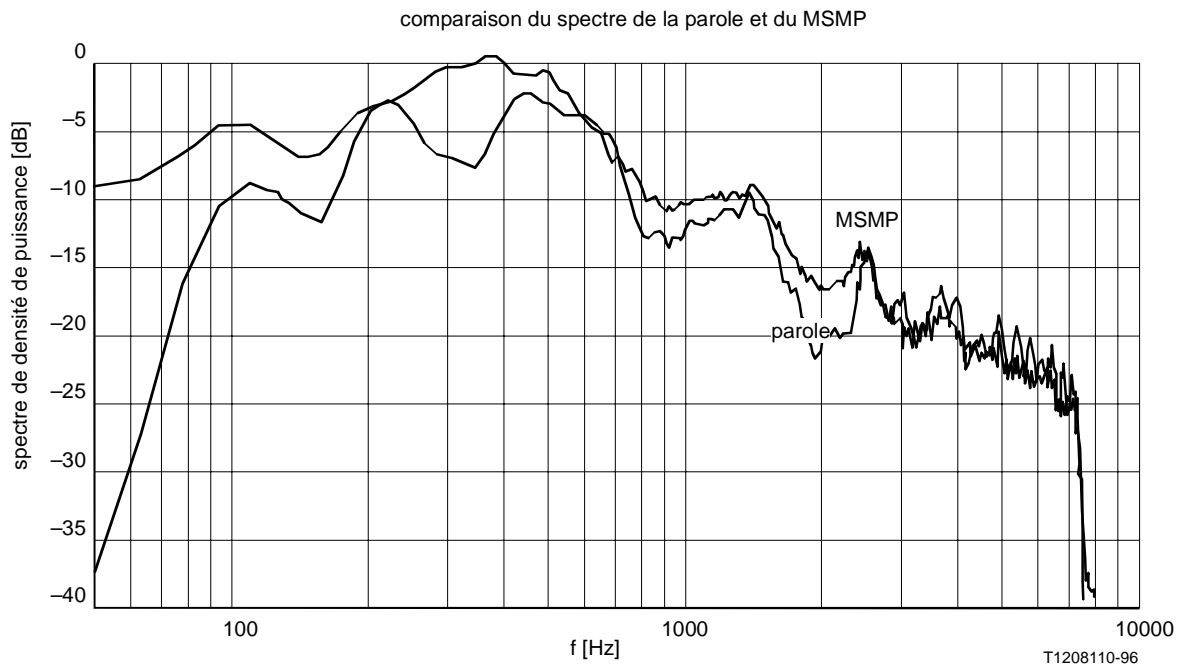


FIGURE 15/P.501

Comparaison du spectre de la parole et du MSMP

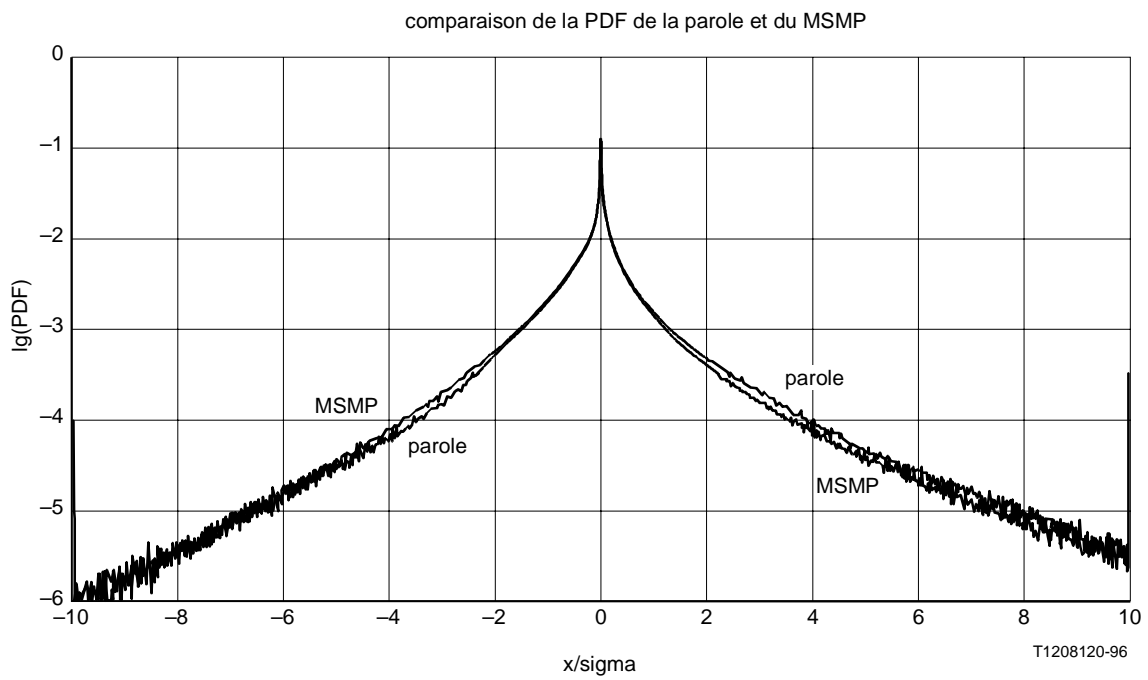


FIGURE 16/P.501

Comparaison de la PDF de la parole et du MSMP

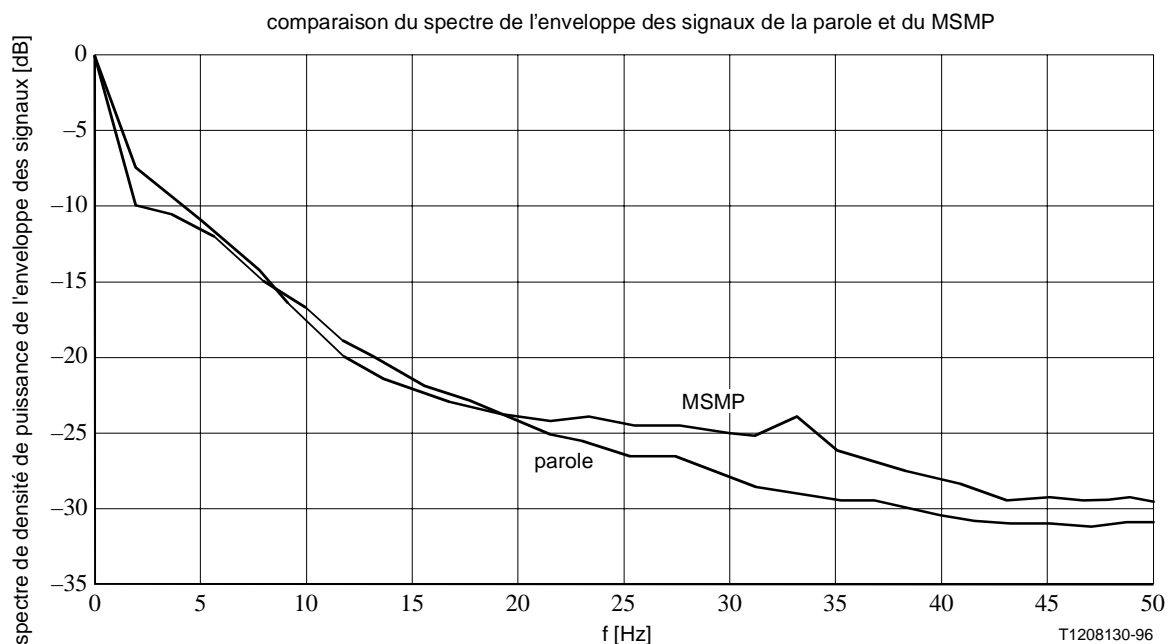


FIGURE 17/P.501

Comparaison du spectre de l'enveloppe des signaux de la parole et du MSMP

5.2.4.4.2 Application

Des signaux MSMP peuvent être utilisés si pour la mesure on a besoin d'une séquence de conversation "type". Les dispositions décrites ci-dessus peuvent être appliquées pour produire un signal d'essai, qui est adapté soit à une langue donnée (par exemple, anglais ou allemand), soit à une combinaison de plusieurs langues. A cet égard, il suffit d'utiliser un autre ensemble de données de signaux vocaux pour initialiser les chaînes de Markov. Le MSMP permet non seulement de prescrire le spectre et la PDF des signaux sur longue durée, mais également d'inclure les caractéristiques importantes sur brève durée des changements de formants et de hauteur des sons "naturels". D'un côté, ce modèle permet de s'affranchir de la dépendance vis-à-vis des données vocales dans son application, mais, d'un autre côté, il offre la possibilité d'inclure des caractéristiques particulières nécessaires de certains signaux vocaux (comme par exemple la différence homme/femme, la variation faible/forte de caractéristiques, etc.).

En règle générale, il faut disposer de temps d'intégration longs (>10 s) si à partir des mesures on doit calculer des paramètres tels que les réponses en fréquence ou les équivalents en sonie.

5.3 Signaux vocaux

Pour certaines applications, il est possible d'utiliser des signaux de voix réelle (unidirectionnelle ou conversationnelle). Si le dispositif soumis à l'essai présente un comportement marqué non linéaire et/ou variable par rapport au temps, les notions de réponse en fréquence, de distorsion et de rapport signal/bruit ne sont plus directement applicables ou peuvent ne plus être pertinentes. On peut toutefois étudier le comportement du dispositif soumis à l'essai en utilisant des fragments de voix réelle et juger la qualité globale suivant un modèle de perception chez un auditoire humain. Grâce à un modèle de ce type il est en effet possible de définir une "fonction de réponse en fréquence perceptuelle" ou bien des mesures spéciales, par exemple fondées sur l'utilisation de paramètres psychoacoustiques (mesures de distorsion perceptuelle).

Bibliographie

- [1] GIERLICH (H.W.): A measurement technique to determine the transfer characteristics of hands-free telephones, *Signal Processing*, Vol. 27, n° 3, p.281-300, 1992.
- [2] STEENEKEN (H.J.M.), HOUTGAST (T.): A physical method for measuring speech transmission quality, *J. Acoust. Soc. Am.*, Vol. 67, n° 1, p.318-326, janvier 1980.
- [3] ITU-T COM 12-14, *Speech and Speech-Model Signals: A Comparison*, Commission d'études 12 de l'UIT-T, Genève, mai 1993.
- [4] HALKA (U.), HEUTE (U.): Speech-Model Processes Controlled by Discrete Markov-Chains, *Proc. Asilomar Conf. Sig. Syst. Comp.*, p.1196-1200, Pacific Grove, Etats-Unis, 1993.
- [5] LINDE (Y.), BUZO (A.), GREY (R.M.): An Algorithm for Vector Quantizer Design, *IEEE Transactions on Communications*, Vol. COM-28, n° 1, janvier 1980.

SÉRIES DES RECOMMANDATIONS UIT-T

- Série A Organisation du travail de l'UIT-T
- Série B Moyens d'expression
- Série C Statistiques générales des télécommunications
- Série D Principes généraux de tarification
- Série E Réseau téléphonique et RNIS
- Série F Services de télécommunication non téléphoniques
- Série G Systèmes et supports de transmission
- Série H Transmission des signaux autres que téléphoniques
- Série I Réseau numérique à intégration de services
- Série J Transmission des signaux radiophoniques et télévisuels
- Série K Protection contre les perturbations
- Série L Construction, installation et protection des câbles et autres éléments des installations extérieures
- Série M Maintenance: systèmes de transmission, de télégraphie, de télécopie, circuits téléphoniques, et circuits loués internationaux
- Série N Maintenance: circuits internationaux de transmission radiophoniques et télévisuels
- Série O Spécifications des appareils de mesure
- Série P Qualité de transmission téléphonique**
- Série Q Commutation et signalisation
- Série R Transmission télégraphique
- Série S Equipements terminaux de télégraphie
- Série T Equipements terminaux et protocoles 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 Z Langages de programmation