



UNION INTERNATIONALE DES TÉLÉCOMMUNICATIONS

UIT-T

P.80

SECTEUR DE LA NORMALISATION
DES TÉLÉCOMMUNICATIONS
DE L'UIT

(03/93)

**QUALITÉ DE LA TRANSMISSION TÉLÉPHONIQUE
ESSAIS SUBJECTIFS D'OPINION**

**MÉTHODES D'ÉVALUATION SUBJECTIVE
DE LA QUALITÉ DE TRANSMISSION**

Recommandation UIT-T P.80

(Antérieurement «Recommandation du CCITT»)

AVANT-PROPOS

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

La Conférence mondiale de normalisation des télécommunications (CMNT), qui se réunit tous les quatre ans, détermine les thèmes que les Commissions d'études de l'UIT-T doivent examiner et à propos desquels elles doivent émettre des Recommandations.

La Recommandation UIT-T P.80, élaborée par la Commission d'études XII (1988-1993) de l'UIT-T, a été approuvée par la CMNT (Helsinki, 1-12 mars 1993).

NOTES

1 Suite au processus de réforme entrepris au sein de l'Union internationale des télécommunications (UIT), le CCITT n'existe plus depuis le 28 février 1993. Il est remplacé par le Secteur de la normalisation des télécommunications de l'UIT (UIT-T) créé le 1^{er} mars 1993. De même, le CCIR et l'IFRB ont été remplacés par le Secteur des radiocommunications.

Afin de ne pas retarder la publication de la présente Recommandation, aucun changement n'a été apporté aux mentions contenant les sigles CCITT, CCIR et IFRB ou aux entités qui leur sont associées, comme «Assemblée plénière», «Secrétariat», etc. Les futures éditions de la présente Recommandation adopteront la terminologie appropriée reflétant la nouvelle structure de l'UIT.

2 Dans la présente Recommandation, le terme «Administration» désigne indifféremment une administration de télécommunication ou une exploitation reconnue.

© UIT 1994

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

TABLE DES MATIÈRES

	<i>Page</i>
1 Introduction	1
2 Méthodes recommandées	1
2.1 Essais d'opinion de conversation	1
2.2 Essais d'opinion d'écoute	1
2.3 Enquêtes et essais	2
2.4 Autres essais	3
Annexe A – Essais d'opinion sur les conversations.....	3
A.1 Montage des essais	3
A.2 Montage de l'expérience	7
A.3 Tâche de conversation	7
A.4 Procédure de mesure.....	8
Annexe B – Essais d'écoute – Notation par catégorie absolue (ACR)	10
B.1 Enregistrements de la source	10
B.2 Choix des conditions du circuit	12
B.3 Montage de l'expérience	13
B.4 Procédure des essais d'écoute	13
Annexe C – Essais de détection de la réponse discontinue alternative	16
Annexe D – Méthode de notation par catégorie de dégradation (DCR).....	17
D.1 Introduction	17
D.2 Procédure de la notation par catégorie de dégradation (DCR)	17
D.3 Analyse statistique	18
Annexe E – Méthode du seuil pour la comparaison de systèmes de transmission avec un système de référence	18
E.1 Introduction	18
E.2 Procédure à suivre pour les essais.....	19
E.3 Présentation des signaux.....	19
E.4 Sources de paroles	20
E.5 Environnement d'écoute.....	20
E.6 Personnes qui écoutent	20
E.7 Fiabilité	20
Références	20
Bibliographie.....	21

MÉTHODES D'ÉVALUATION SUBJECTIVE DE LA QUALITÉ DE TRANSMISSION

(modifiée à Helsinki, 1993)

1 Introduction

La présente Recommandation fournit aux Administrations des indications pour effectuer des essais subjectifs d'évaluation de la qualité de transmission dans leurs laboratoires. Mais elles ne concernent pas les essais des types décrits en détail dans d'autres Recommandations, à savoir:

- a) mesure des équivalents de référence et des équivalents relatifs, voir la Recommandation P.72 (*Livre rouge*, Tome V, Genève, 1985 et le *Manuel de téléphonométrie*, Genève, 1987);
- b) détermination des équivalents pour la sonie – voir la Recommandation P.78;
- c) mesure de l'affaiblissement équivalent pour la netteté (AEN) – voir la Recommandation P.45 (résumée dans le *Livre jaune*, Tome V, 1981, en détail dans le Tome V du *Livre rouge* de 1962).

Elle ne traite pas non plus des diverses sortes d'essais spécialisés servant à mettre au point des éléments d'équipements téléphoniques et permettant de diagnostiquer des pannes et des défauts, comme les Diagnostic Rhyme Tests [1] et d'autres essais propres à l'étude de certains aspects de l'émission de parole.

Cette Recommandation présente les méthodes approuvées qui permettent d'évaluer dans quelle mesure on peut s'attendre à un comportement satisfaisant des connexions téléphoniques.

Les méthodes décrites ci-après ont une application universelle quelle que soit la forme des facteurs de dégradation présents. Exemples de facteurs de dégradation: affaiblissement (qui dépend souvent de la fréquence), bruit de circuit, bruit ambiant, effet local, écho pour la personne qui parle, distorsions de non-linéarité de différentes sortes comprenant le codage à faible débit, le temps de propagation, les effets préjudiciables des dispositifs actionnés par la voix, les distorsions temporelles dues à la commutation par paquets et les dégradations en fonction du temps de la propagation, y compris celles qui ont pour origine les appareils à haut-parleur. Il convient aussi de tenir compte de combinaisons possibles de plusieurs de ces facteurs.

Les références sont communes à toutes les parties et sont fournies à la fin. Toutes les références au *Manuel de téléphonométrie* s'appliquent à la 2^e édition (1993).

2 Méthodes recommandées

2.1 Essais d'opinion de conversation

Les essais de conversation en laboratoire visent à reproduire dans la mesure du possible les conditions réelles de service perçues par les usagers du téléphone. Pour cela, il est nécessaire de choisir avec soin les conditions du circuit et les sujets et d'exécuter les essais de façon appropriée.

Il importe que les conditions simulées lors de l'essai soient correctement spécifiées et réalisées, et qu'elles soient mesurées exactement avant et après chaque expérience; en outre des dispositifs auxiliaires, comme la numérotation et la sonnerie, doivent être fournis et des enregistrements fidèles des résultats de chaque essai doivent être conservés. L'Annexe A fournit une description détaillée de cette méthode, expose des considérations et recommande des précautions.

2.2 Essais d'opinion d'écoute

Les essais d'opinion d'écoute ne parviendront sans doute pas au même niveau de réalisme que les essais de conversation, de sorte que les restrictions sont moins sévères à certains égards; mais les caractéristiques artificielles qu'il faut accepter impliquent un contrôle strict de bien des choses qui sont autorisées, dans des essais de conversation, à trouver leur propre équilibre.

La méthode d'essai recommandée pour des essais d'écoute est celle de la notation par catégories absolues (ACR) (*absolute category rating*) décrite à l'Annexe B, qui est conforme à la méthode des jugements par catégories recommandée pour les essais de conversation (voir l'Annexe A) et adoptée en partie pour les mêmes raisons. Les notations de catégorie s'appliquent à de brefs groupes de phrases indépendantes, qui passent par plusieurs processus standards et ceux à l'essai. Cette méthode est bien établie et a été appliquée à des connexions analogiques et numériques et à des codecs numériques. Lors des travaux qui ont abouti à l'établissement des Recommandations G.726 MICDA à 32 kbit/s et G.722, par exemple, des laboratoires de différents pays ont utilisé pour les essais subjectifs la même méthode dans les mêmes conditions physiques et avec des systèmes de transmission identiques et les résultats se sont avérés très homogènes.

Les autres méthodes habituellement employées sont la méthode de détection de la réponse discontinue alternative, la notation par catégories de dégradation (DCR) (*degradation category rating*) et la méthode de seuil.

L'Annexe C décrit les essais de détection de la réponse discontinue alternative qui permettent d'évaluer les valeurs de seuil de certaines quantités et les probabilités correspondantes. Par exemple, le niveau au-dessus duquel le brouillage sur une seule fréquence sera probablement gênant ou décelable, ou la probabilité qu'une diaphonie dans une gamme de niveaux soit intelligible peuvent être le mieux déterminés par cette méthode.

La méthode de notation par catégories de dégradation (DCR), décrite à l'Annexe D peut se substituer à celle de la notation par catégories absolues. Elle compare le système à mesurer avec une référence fixe de qualité élevée et la dégradation (d'«inaudible» à «très gênant») est notée sur une échelle à cinq notes. On peut utiliser cette méthode quand la dégradation (notamment numérique) est faible. Elle peut se révéler particulièrement utile pour évaluer par exemple des algorithmes destinés à des processus de parole numérique très semblables et servir ainsi de moyen d'optimisation du système, une fois que les méthodes de l'Annexe A et de l'Annexe B ont montré que la connexion qui fait l'objet de la dégradation en question dans le cas le plus défavorable ne dépasse pas les limites acceptables.

La méthode de seuil, qui convient également pour l'optimisation d'un système, est décrite dans l'Annexe E. Une comparaison directe du système à mesurer avec un système de référence, comme un appareil de référence à bruit modulé (MNRU) (*modulated noise reference unit*) décrit dans la Recommandation P.81, permet d'égaliser la valeur de la condition de référence (Q pour les processus numériques) qui égale la qualité de fonctionnement du système à mesurer.

On trouvera dans le paragraphe 2.6 du *Manuel de téléphonométrie*, des renseignements concernant d'autres types de méthodes de mesure subjective, qui comprennent les méthodes de notation.

Les essais d'écoute unidirectionnels trouvent des applications directes pour l'évaluation de systèmes de transmission physiques qui sont essentiellement unidirectionnels, qu'il s'agisse par exemple de circuits de diffusion, de systèmes d'annonce en direct ou d'annonces enregistrées qui peuvent faire l'objet d'affaiblissement, de bruit et de distorsion.

Les résultats des essais d'écoute peuvent être appliqués, mais avec certaines réserves, à la prévision de l'évaluation d'une conversation sur un système bidirectionnel, comme une connexion dans un réseau téléphonique public à commutation. Il faut néanmoins veiller à tenir dûment compte des facteurs supplémentaires suivants:

- dégradations de la parole (par exemple, par l'effet local et l'écho);
- dégradations de la conversation (par exemple par le temps de propagation et la mutilation de la parole par l'action de dispositifs actionnés par la voix).

Les paragraphes ci-dessous étudient la préparation du texte à prononcer, le traitement de ce texte, les principes de l'expérience, qui comprennent le choix des conditions du circuit, la procédure des essais d'écoute et le traitement des résultats.

2.3 Enquêtes et essais

Si les importantes dispositions nécessaires peuvent être prises et si l'intérêt de l'étude le justifie, la qualité de transmission peut être déterminée par des «observations de service». La Recommandation P.82 spécifie des moyens pour faire ces observations, y compris les questions à poser aux usagers lors des enquêtes. 100 interviews au moins sont nécessaires par condition pour assurer la précision des résultats.

La méthode des observations de service a, dans bien des applications, pour inconvénient qu'on peut ne pas vérifier vraiment les caractéristiques détaillées des connexions téléphoniques mesurées. Néanmoins, cette méthode permet d'apprécier globalement le comportement de l'«équipement» dans un contexte réel.

Le paragraphe 2.5.8.3 du *Manuel de téléphonométrie* donne des précisions à ce sujet.

2.4 Autres essais

La référence [2] décrit une méthode qui compense en grande partie les inconvénients de la technique d'interview du paragraphe 2.3 tout en conservant bien des avantages. Cette méthode, appelée SYBIL, permet d'admettre une petite proportion d'appels ordinaires en vertu d'arrangements spéciaux qui modifient la qualité de transmission normale d'après un programme d'essai. Si un appel donné a fait l'objet de ce traitement, un volontaire est prié de voter en composant un ensemble de chiffres pour donner son opinion. De la sorte, tous les résultats sont enregistrés par l'ordinateur central et le secret est absolument assuré.

Annexe A

Essais d'opinion sur les conversations

(Cette annexe fait partie intégrante de la présente Recommandation)

A.1 Montage des essais

A.1.1 Conditions physiques

A.1.1.1 Cabines d'essai

Les deux sujets sont installés dans des cabines insonorisées près du point à partir duquel l'expérience est commandée. Le volume de la salle n'est pas inférieur à 20 m³ et le temps de réverbération inférieur à 500 ms (en général de l'ordre de 200 à 300 ms), pour les appareils tenus à la main comme les combinés téléphoniques ou pour les casques; et pas inférieur à 30 m³ pour les systèmes à haut-parleur (une attention supplémentaire est nécessaire si le temps de réverbération est une variable expérimentale).

Les dimensions intérieures de la cabine sont conçues de manière à réduire à un minimum les effets du schéma d'ondes stationnaires. Le rapport type est 5:4:3.

La construction matérielle des salles doit permettre d'obtenir une atténuation suffisante des sons de l'ambiance de bruit extérieure et de satisfaire ainsi aux spécifications du A.1.1.2.1.

On décorera les cabines de façon à recréer un environnement naturel.

A.1.1.2 Bruit

A.1.1.2.1 Bruit minimal

Le niveau de bruit ambiant (quand on n'établit pas volontairement un bruit d'origine externe) est réduit le plus possible. Pour des raisons pratiques, comme le renouvellement régulier de l'aire dans la cabine, l'objectif est une limite supérieure de NC25 [3] ou de NR25 [4]. Ces valeurs correspondent à peu près au niveau de bruit dans les maisons (chambres à coucher), les hôpitaux et les bibliothèques.

A.1.1.2.2 Bruit d'origine externe

Le bruit d'origine externe est injecté selon le spectre nécessaire (par exemple spectre de Hoth pour représenter le bruit de salle type – voir A.1.1.2.2.1) au niveau voulu (par exemple 50 dBA) mesuré avec un sonomètre de précision conforme à la Publication CEI 651:1979, utilisé avec la «pondération A» et la caractéristique de mesure «rapide». Si des conversations différentes dans la même expérience exigent des niveaux différents de bruit de salle, on veille à empêcher que les transitions n'apparaissent trop clairement aux sujets. Théoriquement, il faudrait modifier le bruit de salle uniquement quand les sujets ont quitté les pièces insonorisées. Si c'est impossible, on modifie progressivement le niveau (sans dépasser 4 dB par seconde), au moment où aucune conversation n'est en cours et quand l'attention des sujets est occupée ailleurs – à communiquer avec l'opérateur, par exemple.

On trouvera en A.1.1.2.2.1 et A.1.1.2.2.2 des spectres définis par leurs caractéristiques à long terme.

Pour certaines applications, il est nécessaire d'utiliser un bruit dont le niveau ou le spectre varie, comme des enregistrements sur bande de bruit réel de bureau ou de circulation. En pareils cas, il faut s'assurer que les caractéristiques statistiques sont stables en moyenne sur une période raisonnablement courte, par exemple une minute.

Il est recommandé de mesurer le niveau et le spectre de bruit au moins deux fois, à savoir au début et à la fin de l'expérience. Toute variation importante entre les deux mesures, comparées l'une à l'autre, doit être évaluée par l'expérimentateur car elle peut jeter un doute sur la validité de l'expérience.

Il est indispensable de faire en sorte que les haut-parleurs et les amplificateurs soient capables de reproduire fidèlement le bruit de salle requis.

A.1.1.2.2.1 Bruit de salle

Pour le bruit de salle, le spectre de la densité de puissance doit correspondre à celui publié par Hoth [5]. Le Tableau A.1 donne la densité spectrale dont le niveau a été réglé pour donner 50 dBA sur un sonomètre conforme à la Recommandation de la CEI (publication 179) [6]. La Figure A.1 illustre ce procédé. Le spectre est indépendant du niveau, c'est-à-dire que pour 40 dBA, le niveau dans chaque bande sera inférieur de 10 dB à celui indiqué dans le Tableau A.1. Des informations complémentaires concernant la puissance dans chaque bande d'un tiers d'octave sont également données dans le Tableau A.1.

TABLEAU A.1/P.80

Spectre de bruit de salle

Fréquence (Hz)	Densité spectrale (dB SPL/Hz)	Largeur de la bande $10 \log_{10} \Delta f$ (dB)	Puissance totale dans chaque bande d'un tiers d'octave	Tolérance (dB)
100	32,4	13,5	45,9	±
125	30,9	14,7	45,4	
160	29,1	15,7	44,9	
200	27,6	16,5	44,1	
250	26,0	17,6	43,6	
315	24,4	18,7	43,1	
400	22,7	19,7	42,3	
500	21,1	20,6	41,7	
630	19,5	21,7	41,2	
800	17,8	22,7	40,4	
1000	16,2	23,5	39,7	
1250	14,6	24,7	39,3	
1600	12,9	25,7	38,7	
2000	11,3	26,5	37,8	
2500	9,6	27,6	37,2	
3150	7,8	28,7	36,5	
4000	5,4	29,7	34,8	
5000	2,6	30,6	33,2	
6300	-1,3	31,7	30,4	
8000	-6,6	32,7	26,0	

NOTES

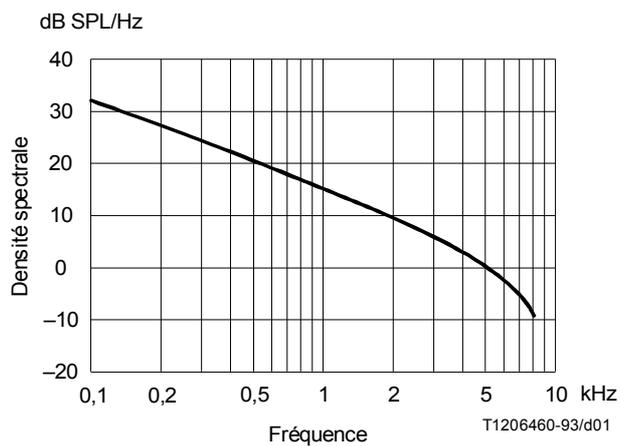
1 Le signal électrique d'entrée, par exemple un bruit blanc, est limité aux bandes d'un tiers d'octave centrées sur les fréquences préférées de l'ISO (ISO 266), de 100 Hz à 8000 Hz, les bords des bandes étant conformes aux filtres décrits dans la publication 225 de la CEI.

2 Il est difficile de limiter le bruit de salle acoustique à basses fréquences, en particulier dans la région non spécifiée au-dessous de 100 Hz, à cause des dimensions des coffrets typiques d'essai, de l'affaiblissement médiocre de ces coffrets et de l'influence des bruits ambiants, par exemple celui d'un appareil de climatisation. Par conséquent on a intérêt à choisir un coffret d'essai où ces niveaux de pression acoustique (SPL) parasite en basses fréquences sont réduits à un minimum.

A.1.1.2.2.2 Bruit à l'intérieur de véhicules

L'utilisation de deux spectres représentant le bruit à l'intérieur de véhicules [7] et [8] a été recommandée. Ces spectres sont représentés de façon adéquate par des courbes simplifiées [9]; l'un d'eux se rapporte à des véhicules en mouvement, l'autre à des véhicules en stationnement. Le Tableau A.2 donne les densités spectrales ainsi que des informations complémentaires sur la puissance de chaque bande d'un tiers d'octave. La densité spectrale pour des véhicules en mouvement est indiquée en a) de la Figure A.2), celle de véhicules en stationnement en b) de la Figure A.2). Ces spectres sont indépendants du niveau.

Le Tableau A.3 donne les valeurs des niveaux de pression acoustique (SPL) non pondérés pour différentes vitesses, calculées pour les bandes d'un tiers d'octave de l'ISO centrées sur 63 Hz à 8000 Hz.



SPL Niveau de pression acoustique

FIGURE A.1/P.80
Densité spectrale de bruit de salle

TABLEAU A.2/P.80

Spectre de bruit à l'intérieur de véhicules

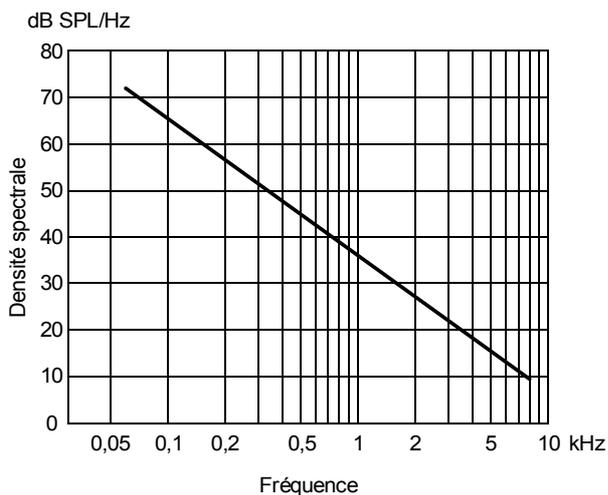
Fréquence (Hz)	Densité spectrale (dB SPL/Hz)		Largeur de la bande $10 \log_{10} \Delta f$ (dB)	Puissance totale dans chaque bande d'un tiers d'octave (dB SPL)		Tolérance (dB)
	en mouvement	en stationnement		en mouvement	en stationnement	
63	72,3	58,3	11,7	84,0	70,0	±
80	69,3	55,0	12,7	82,0	66,7	
100	66,5	49,8	13,5	80,0	63,3	
125	63,3	45,1	14,7	78,0	60,0	
160	60,3	42,0	15,7	76,0	56,7	
200	57,5	36,8	16,5	74,0	53,3	
250	54,4	34,7	17,6	72,0	52,3	
315	51,3	32,6	18,7	70,0	51,3	
400	48,3	30,6	19,7	68,0	50,3	
500	45,4	28,7	20,6	66,0	49,3	
630	42,3	26,6	21,7	64,0	48,3	
800	39,3	24,6	22,7	62,0	47,3	
1000	36,5	22,8	23,5	60,0	46,3	
1250	33,3	20,6	24,7	58,0	45,3	
1600	30,3	18,6	25,7	56,0	44,3	
2000	27,5	16,8	26,5	54,0	43,3	
2500	24,4	14,7	27,6	52,0	42,3	
3150	21,3	12,6	28,7	50,0	41,3	
4000	18,3	10,6	29,7	48,0	40,3	
5000	15,4	8,7	30,6	46,0	39,3	
6300	12,3	6,6	31,7	44,0	38,3	
8000	9,3	4,6	32,7	42,0	37,3	

Niveaux calculés de la pression acoustique des spectres

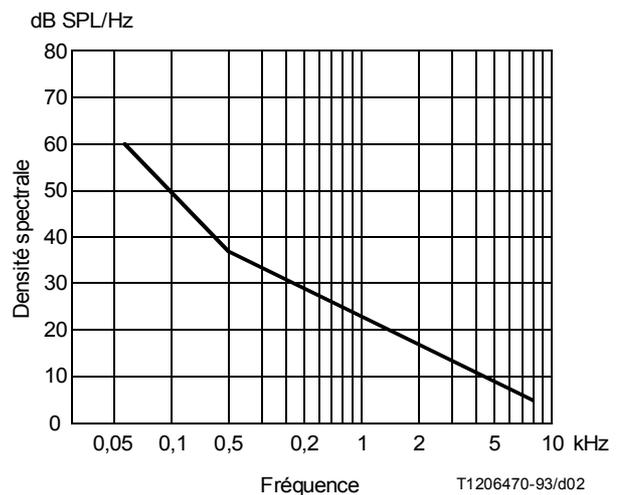
Spectres		Niveau de pression acoustique non pondéré (dB SPL)
En mouvement	30 km/h	80
	80 km/h	85
	110 km/h	90
En stationnement		75

NOTES relatives aux Tableaux A.2 et A.3:

- 1 Ces valeurs s'entendent pour des véhicules types. On peut ajuster les niveaux à discrétion vers le bas pour des véhicules de luxe et vers le haut pour des véhicules plus bruyants.
- 2 A cause de la difficulté pratique de générer de si hauts niveaux de pression acoustique à basses fréquences et parce que la parole normale ne contient pas d'énergie apparente au-dessous d'environ 63 Hz, bande de fréquences à laquelle l'oreille est aussi relativement insensible, il conviendrait probablement de restreindre le spectre de bruit recommandé à des fréquences supérieures à 63 Hz. Toutefois, il ne faut pas oublier que les vibrations à basses et moyennes fréquences ont des effets physiologiques et psychologiques importants qui devraient faire l'objet d'une étude à part.
- 3 Le signal électrique d'entrée, par exemple un bruit blanc, est limité à des bandes d'un tiers d'octave centrées sur les fréquences préférées de l'ISO, de 63 Hz à 8000 Hz (ISO 266), et dont les bords sont conformes aux filtres décrits dans le document CEI 225.
- 4 Il est difficile de limiter le bruit de salle acoustique à basses fréquences, en particulier dans la région non spécifiée au-dessous de 63 Hz, à cause des dimensions des coffrets typiques d'essai, de l'affaiblissement médiocre de ces coffrets et de l'influence des bruits ambiants, par exemple celui d'un appareil de climatisation. Par conséquent, on a intérêt à choisir un coffret d'essai où ces niveaux de pression acoustique parasite en basses fréquences sont réduits à un minimum.



a) Densité spectrale pour des véhicules en mouvement



b) Densité spectrale pour des véhicules en stationnement

FIGURE A.2/P.80

Densité spectrale du bruit des véhicules

A.1.1.3 Position de mesure du bruit

Il est recommandé que la mesure du niveau de pression acoustique (SPL) (*sound pressure level*) dans les cabines d'essai (voir A.1.1.1) soit effectuée comme suit:

- mobilier en place;
- aucun sujet ou personnel d'essai présent;
- le SPL doit être mesuré à une distance verticale de 740 mm au-dessus de l'axe du siège du sujet à l'aide d'un sonomètre conforme à la Recommandation P.54 et utilisant la pondération «A»;
- le spectre du bruit d'origine externe doit être mesuré en 1/3 d'octaves, centré sur les fréquences préférées comme indiqué dans l'ISO R.266 et doit demeurer dans les limites des tolérances spécifiées, par exemple ± 3 dB pour le bruit de Hoth (voir A.1.1.2.2.1);
- dans les salles où plusieurs sujets doivent être soumis aux essais, la différence en dBA, pour toutes les positions des sujets, ne doit pas varier de plus de ± 2 dB.

NOTE – Il est suggéré que la distance minimale entre chaque haut-parleur et la position de mesure soit de 1,5 m.

A.1.2 Etablissement de la connexion

Pour établir la connexion en laboratoire, il faut tenir compte:

- des combinés téléphoniques;
- de l'établissement initial de la communication;
- de la représentation en laboratoire des connexions téléphoniques.

Il est recommandé que la caractéristique d'efficacité en fonction de la fréquence soit mesurée au moins deux fois, à savoir au début et à la fin de l'expérience. Toute variation importante entre les deux mesures comparées l'une à l'autre, doit être évaluée par l'expérimentateur, car elle peut jeter un doute sur la validité de l'expérience.

Le paragraphe 2.5.8.2 du *Manuel de téléphonométrie* donne des précisions à ce sujet.

A.1.3 Surveillance

La surveillance peut prendre bien des formes mais les trois plus couramment employées sont les suivantes:

- *système d'intercommunication* – Indispensable pour permettre au sujet et à l'opérateur de communiquer entre eux;
- *surveillance visuelle* – Deux buts: la sécurité et l'observation des particularités du sujet (c'est-à-dire sa façon de tenir le combiné);
- *enregistrements sur bande et système d'enregistrement* – Cela facilite le recueil d'autres renseignements utiles, comme la durée de la communication, la tension de la parole et l'activité de la parole.

Le paragraphe 2.5.8.2 du *Manuel de téléphonométrie* donne des précisions sur cet aspect.

A.2 Montage de l'expérience

On peut appliquer plusieurs méthodes pour concevoir des expériences, des carrés latins, des carrés de Youden, des blocs incomplets équilibrés et la randomisation avec reproduction, pour n'en citer que quelques-unes. L'opérateur doit décider la méthode à utiliser en fonction du nombre de conditions de mesure, de la précision des résultats et de l'aptitude à formuler des conclusions valables à partir des résultats.

Les montages qui conviennent les mieux sont du type gréco-latin $n \times n$, dont le *Manuel de téléphonométrie* donne une description détaillée en 2.5.8.2.

A.3 Tâche de conversation

On s'efforcera de faire en sorte que les conversations aient un sens et que les sujets aient toute latitude pour tirer parti des possibilités de transmission du circuit d'essai.

En règle générale, chaque conversation doit avoir un début et une fin naturelles. Elle ne doit jamais s'achever au milieu de la tâche, sauf en cas de nécessité absolue (à l'exception des essais de conversation simplifiés qui sont décrits dans le *Manuel de téléphonométrie* en 2.5.8.2).

Le paragraphe 2.5.8.2 d) du *Manuel de téléphonométrie* donne un exemple de tâche de conversation.

A.4 Procédure de mesure

A.4.1 Choix des sujets

Les sujets participant aux essais de conversation sont choisis au hasard parmi la population normale utilisant le téléphone, les conditions étant:

- a) qu'ils n'aient pas participé directement à des travaux d'évaluation de la qualité des circuits téléphoniques ou à des travaux connexes comme le codage de la voix;
- b) qu'ils n'aient pas participé à des essais subjectifs au cours des six mois précédents au moins, ni à des essais de conversation depuis un an au moins.

Si la population disponible est trop limitée on en tiendra compte pour tirer des conclusions des résultats.

On n'équilibrera pas le nombre des sujets masculins et féminins, à moins que le montage de l'expérience ne l'exige. Les sujets sont arbitrairement disposés par paires dans le montage expérimental avant l'essai et restent ainsi pendant toute sa durée.

A.4.2 Echelles d'appréciation subjective

Les échelles d'appréciation subjective ci-dessous sont celles que recommande le CCITT.

A.4.2.1 Echelle d'appréciation subjective de conversation

Diverses échelles d'appréciation de catégorie à cinq notes peuvent être utilisées à différentes fins. La présentation et le libellé des échelles d'appréciation subjective, telles qu'elles sont perçues par les sujets au cours des essais, sont très importants et doivent se conformer à une norme établie après des années d'expérience. L'échelle d'appréciation subjective suivante est celle qui est le plus souvent employée pour les applications du CCITT et un libellé équivalent devrait être utilisé selon la langue, ce qui pourrait se traduire par de légères variations par rapport au texte anglais original.

Voici la notation par catégorie et par sujet à la fin de chaque conversation:

Opinion au sujet de la connexion que vous venez d'utiliser

Excellente
Bonne
Passable
Médiocre
Mauvaise

L'opérateur affecte les valeurs suivantes aux notes:

Excellente = 5; Bonne = 4; Passable = 3; Médiocre = 2; Mauvaise = 1,

et la suite du traitement statistique appliquera ces valeurs. La moyenne arithmétique d'un ensemble quelconque de ces notes d'opinion s'appelle **note moyenne d'opinion sur la conversation**; elle est représentée par le symbole Y_C (ou par le symbole Y_c si la notation avec indice n'est pas disponible).

NOTES

1 Dans le passé, les équivalences Excellente = 4, Bonne = 3, Passable = 2, Médiocre = 1 et Mauvaise = 0 étaient souvent utilisées. Si l'on utilise les résultats d'essais antérieurs, il faut se souvenir que les notes moyennes doivent toutes être relevées d'une unité pour être comparables à celles que l'on obtient à présent, mais il n'y a par ailleurs aucune différence dans le traitement numérique qui puisse être appliquée.

2 Cette échelle a été établie en anglais et l'équivalent doit être inséré en fonction de la langue, ce qui peut se traduire par de légères différences par rapport au texte anglais original.

A.4.2.2 Echelle de difficulté

Il s'agit d'une réponse binaire obtenue de chaque sujet à la fin de chaque conversation.

Avez-vous, ou votre partenaire, éprouvé des difficultés pour parler ou écouter dans cette connexion?

Oui
Non

L'opérateur affecte les valeurs suivantes à la réponse:

Oui = 1, Non = 0.

La quantité évaluée (pourcentage de réponses «oui») s'appelle pourcentage de difficulté; elle est désignée par le symbole % D. La proportion simple correspondante est désignée par le symbole d; autrement dit, % D = 100 d.

NOTE – Il arrive fréquemment que la nature de la difficulté soit exigée, auquel cas l'opérateur demande en général au sujet de décrire comment il a perçu cette difficulté.

La présentation et le libellé de l'échelle d'appréciation subjective, telle qu'elle est perçue par les sujets lors des expériences sont très importants et doivent respecter les normes obtenues à l'issue de plusieurs années d'expérience: voir A.4.3.

A.4.2.3 Autres échelles d'appréciation subjective

Il s'agit de variantes des méthodes «estimation de magnitude» et «adaptation de modalités croisée» [10], qui permettent d'obtenir l'une des réactions suivantes:

- a) une série numérique de catégories 1, 2, 3, 4, 5 (et désignées comme telles au sujet), mais avec les descriptions jointes seulement à la première et à la dernière pour identifier la dimension subjective;
- b) une marque numérique sur une échelle de 1 à un nombre beaucoup plus grand que cinq, par exemple 10 ou 100; ou
- c) une longueur proportionnelle à une caractéristique (par exemple, la qualité) marquée manuellement sur une ligne droite donnée.

Le paragraphe 2.6.2 du *Manuel de téléphonométrie* passe en revue les méthodes expérimentales.

A.4.3 Instructions aux sujets

Des instructions sont données aux sujets quand ils arrivent pour leur première visite. Il est normal que les sujets reçoivent avant leur arrivée, une lettre qui contient des informations non techniques sur l'expérience et sur ce que l'on attend d'eux. Le Tableau 3/2.5, dans le paragraphe 2.5.8.2 du *Manuel de téléphonométrie* donne un exemple de cette lettre.

On demande aux sujets s'ils ont lu et compris la lettre. Des précisions leur sont données, le cas échéant, ainsi que la possibilité de poser des questions. Les salles insonorisées et leurs installations leur sont présentées. Ils apprennent combien de communications seront faites lors de cette visite. Lors des visites suivantes, ils sont simplement avisés que la procédure sera la même que précédemment, moyennant peut-être un nombre de communications différent. Le Tableau 4/2.5 dans le paragraphe 2.5.8.2 du *Manuel de téléphonométrie* donne un exemple de quelques détails opérationnels nécessités par une expérience.

A.4.4 Recueil des données

Les niveaux de conversation et les données connexes, comme les durées et les coefficients d'activité peuvent être obtenus des enregistrements sur bande, mais ils sont en général mesurés en direct par des appareils de mesure pilotés par ordinateur et stockés directement dans des fichiers d'ordinateur pour être analysés ultérieurement.

L'expérimentateur recueille deux réponses subjectives par conversation et par sujet. Les données essentielles se composent des notes d'opinion sur la conversation et de la décision concernant les difficultés.

Ces réponses peuvent être recueillies à l'aide de la méthode primitive où chaque sujet consigne son résultat par écrit sur un formulaire. Le Tableau 5/2.5 dans le paragraphe 2.5.8.2 du *Manuel de téléphonométrie* donne un exemple de ce formulaire. Etant donné que les sujets, surtout les plus jeunes, utilisent très fréquemment des claviers, le recueil des notes au moyen de cette méthode est acceptable.

A.4.5 Traitement des résultats

On ne trouvera ci-après qu'une brève description de ce sujet très vaste.

Chaque conversation donne lieu à deux opinions sur la conversation au moyen des appréciations: Excellent, Bon, Passable, Médiocre, Mauvais (qui correspondent respectivement aux notes 5, 4, 3, 2 et 1), de deux votes sur l'échelle de difficultés (Oui = 1, Non = 0), à deux niveaux de parole active mesurés et à une valeur de durée. Dans certains cas, une information peut être recueillie au sujet d'autres variables également; par exemple, des enregistrements vidéo peuvent être faits pour observer comment les sujets tiennent le combiné, ou d'autres données peuvent être fournies par les formules d'opinion ou des enregistrements audio.

Il n'est pas recommandé de calculer des écarts types distincts pour chaque condition. Les limites de confiance doivent être évaluées et des essais de signification doivent être faits au moyen de techniques classiques d'analyse de variance.

Les hypothèses usuelles qui sous-tendent l'analyse de variance sont satisfaites d'assez près dans le cas des notes d'opinion, du niveau de parole active et de la plupart des autres variables en cause; mais elles ne le sont pas – notamment l'hypothèse d'une variance résiduelle constante – s'agissant d'une variable aléatoire binaire comme la note de difficulté. Néanmoins, l'expérience confirme l'observation formulée dans d'autres domaines [11] selon laquelle la technique de l'analyse de variance est assez robuste pour donner des résultats raisonnables, même dans les cas d'écart extrême par rapport aux conditions statistiquement idéales. Les résultats de la première phase de l'analyse de variance des notes sur la difficulté doivent être considérés avec quelque prudence; mais une fois établi qu'il n'y a pas d'anomalies inexplicables dans les résultats, ni d'incompatibilités inexplicables avec le résultat de l'analyse correspondante des résultats Y_C , la seconde phase (analyse détaillée des moyennes des combinaisons de condition d'extrémité) peut être entreprise en toute confiance à l'aide d'une transformation mathématique.

Le paragraphe 2.5.9 du *Manuel de téléphonométrie* décrit en détail cette analyse.

Pour faciliter encore l'examen des données, il faut tracer, s'il y a lieu, des graphiques montrant la note moyenne d'opinion en fonction du paramètre à mesurer, par exemple Y_C en fonction de l'affaiblissement du circuit. Sur ces graphiques l'axe vertical doit toujours être Y_C .

Annexe B

Essais d'écoute – Notation par catégorie absolue (ACR)

(Cette annexe fait partie intégrante de la présente Recommandation)

B.1 Enregistrements de la source

Afin d'éliminer une variabilité intempestive de la source de parole, il convient d'abord de préparer des échantillons de parole ayant les caractéristiques normalisées désirées sous forme d'enregistrement ou de stockage, comme suit:

B.1.1 Environnement d'enregistrement

Le locuteur doit être assis dans une salle calme d'un volume de 30 à 120 m³ et donnant lieu à un temps de réverbération inférieur à 500 ms (de préférence de l'ordre de 200 à 300 ms). Le niveau de bruit de salle doit être inférieur à 30 dBA sans crête dominante dans le spectre.

La caractéristique de bruit de salle doit être indiquée aussi complètement que possible, par exemple en dBA, spectre à long terme et distribution de l'amplitude en fonction du temps. Il est souhaitable d'enregistrer un échantillon de 30 secondes du bruit de salle pour un examen détaillé si cela se révèle nécessaire.

B.1.2 Système d'émission

Quel que soit le système d'émission choisi, par exemple un système téléphonique local ou un système de référence intermédiaire (IRS) (*intermediate reference system*) spécifié dans la Recommandation P.48, le système doit être étalonné selon la Recommandation pertinente (par exemple la Recommandation P.64) et la caractéristique efficacité-fréquence à l'émission doit être communiquée en détail.

Il est recommandé que la caractéristique d'efficacité à l'émission de la connexion soit mesurée au moins deux fois, à savoir au début et à la fin de l'expérience. Toute variation importante entre les deux mesures, comparées l'une à l'autre, doit être évaluée par l'expérimentateur, car elle peut jeter un doute sur la validité de l'expérience.

B.1.3 Système d'enregistrement

Le système d'enregistrement doit être de qualité élevée (studio), c'est-à-dire l'un des suivants:

- a) un magnétophone à bande classique à deux pistes. Le type d'égalisation doit être mentionné, mais CEI est recommandé. Une bande de grande qualité (sans effet d'empreinte, faible bruit) doit toujours être utilisée;
- b) un processeur audio numérique à deux voies avec un magnétoscope à cassette de qualité élevée ou un appareil d'enregistrement sonore numérique sur bande (DAT) (*digital audio tape*);
- c) un système de stockage numérique piloté par ordinateur.

Le troisième système est le meilleur et le plus polyvalent, mais le choix d'un des autres systèmes est souvent dicté par des considérations d'ordre pratique. Dans ces derniers, une des deux pistes doit servir à enregistrer la parole, l'autre à insérer des signaux de commande à un niveau et à une fréquence choisis pour éviter les problèmes de diaphonie.

B.1.4 Texte à prononcer

Le texte à prononcer doit se composer de phrases courtes, simples et claires choisies au hasard parmi celles faciles à comprendre (dans la littérature actuelle non technique ou les journaux, par exemple) et constituées en listes dans un ordre aléatoire, de sorte qu'il n'y ait pas de relation évidente de sens entre une phrase et la phrase suivante). Les phrases très courtes ou très longues doivent être éliminées, le but étant que lors de son énoncé, chaque phrase soit comprise dans un intervalle de temps d'environ 4 secondes, occupant environ la moitié de cette durée pour la parole proprement dite. Le Tableau B.1 donne des exemples de phrases.

L'opérateur doit décider combien de phrases sont nécessaires dans chaque groupe pour constituer un échantillon de parole. Un minimum de deux et un maximum de cinq sont recommandés. L'intervalle de temps entre les phrases, pendant lequel le bruit de circuit peut être entendu et les processus adaptatifs peuvent adopter de nouveaux états est également important. Il est recommandé d'enregistrer les plus longs groupes qui peuvent être nécessaires, du fait qu'il est toujours possible d'obtenir des groupes qui peuvent être nécessaires, du fait qu'il est toujours possible d'obtenir des groupes plus courts en copiant ou en reproduisant des parties de groupes plus longs.

Les groupes sont combinés en listes comprenant 5 ou 10 groupes chacune, de sorte qu'une liste complète puisse être utilisée comme une série d'échantillons soumis au même traitement mais avec un niveau d'écoute ou un autre paramètre variant quand la liste est reproduite.

TABLEAU B.1/P.80

Exemples de texte à prononcer (pour des sujets anglophones)

You will have to be very quiet.
There was nothing to be seen.
They worshipped wooden idols.
I want a minute with the inspector.
Did he need any money?

B.1.5 Procédure d'enregistrement

La procédure d'enregistrement suivante est largement utilisée et il est recommandé de l'appliquer.

La parole est enregistrée à partir d'un microphone linéaire et d'un amplificateur à faible bruit ayant une caractéristique de fréquence uniforme, spécifiés dans la publication 581-5 de la CEI. Le microphone est installé de 140 à 200 mm devant les lèvres du locuteur. Une boule anti-vent est utilisée si l'on constate des bouffées d'air provenant du locuteur.

Les mêmes paroles sont enregistrées simultanément à partir de la sortie émission d'un système de référence intermédiaire (Recommandation P.48), avec le combiné tenu normalement, ou d'un autre téléphone si l'expérience en cours l'exige.

Deux systèmes d'enregistrement distincts sont utilisés simultanément: un pour enregistrer la parole à large bande dans une voie, l'autre pour enregistrer les sons vocaux téléphoniques dans la voie correspondante. L'autre voie de chaque système d'enregistrement sert à enregistrer les signaux de commande, comme expliqué en 2.5 du *Manuel de téléphonométrie*.

Ce système d'enregistrement en double garantit que les mêmes paroles sont enregistrées sous deux formes (sons vocaux téléphoniques et parole à large bande). Normalement, un seul système suffit pour une expérience quelconque, mais dans certains cas il est indispensable d'en utiliser deux, et il est avantageux de toute façon de pouvoir faire des mesures comparatives sur les deux versions.

Le niveau de parole active défini dans la Recommandation P.56 est observé pendant l'enregistrement. On veille pendant le processus d'enregistrement à ce que le niveau de la parole active dans les deux systèmes d'enregistrement soit compris entre 20 et 30 dB au-dessous du point de saturation du système d'enregistrement pour chaque phrase mesurée séparément. Tout groupe de phrases pour lequel cela n'est pas le cas doit être enregistré à nouveau.

Il est recommandé que le rapport du niveau de parole active au niveau de bruit psophométriquement pondéré (pour la définition, voir 3.2.3/P.83), SNR(p) sur le support d'enregistrement soit supérieur à 40 dB avec un objectif de 50 dB.

Tous les échantillons de parole servant à une expérience doivent être différents: c'est indispensable pour les essais d'efforts d'écoute et souhaitable pour d'autres types d'essais.

B.1.6 Locuteurs

Il doit y avoir autant de locuteurs que l'exige le montage de l'expérience (voir B.3).

Les locuteurs doivent prononcer les phrases couramment mais sans effet théâtral et ne souffrir d'aucune déficience d'élocution (bégaiement, par exemple); ils doivent adopter un niveau d'élocution commode pour eux et qu'ils peuvent conserver assez régulièrement.

B.1.7 Niveaux de parole

Une fois terminés, les enregistrements sont reproduits et on mesure le niveau de parole de chaque phrase. Les listes (annonces, phrases et tonalités de commande) sont ensuite réenregistrées sur un second système, moyennant les réglages de gain nécessaires, de façon à mettre chaque groupe de phrases au niveau de parole active normalisé spécifié ci-dessous, tout en conservant les relations de temps appropriées entre les phrases et les signaux de tonalité dans l'autre voie.

En ce qui concerne la parole à bande étroite, le niveau normalisé est obtenu par la mesure et le réglage du signal enregistré dans la bande étroite directement; la cible recommandée est de -26 dB ($\pm 0,5$ dB) par rapport au niveau de saturation du système d'enregistrement. La tonalité d'étalonnage à un niveau efficace égal au niveau actif moyen de la parole réenregistrée.

Pour la parole à large bande, il faut tenir compte de l'usage prévu des enregistrements. Il y a parfois lieu d'adopter les mêmes niveaux que pour les sons vocaux téléphoniques, mais si l'enregistrement doit passer, pour être reproduit, par un haut-parleur ou une bouche artificielle, les niveaux de parole cibles doivent être de nature à assurer l'égalité à la sortie de toute la chaîne de reproduction électroacoustique.

B.1.8 Signal d'étalonnage

Au début de chaque enregistrement, 20 secondes ou plus de tonalité sont insérées au stade de réenregistrement (pour les besoins de l'étalonnage) à un niveau qui se trouve dans une relation connue avec le niveau moyen de la parole active (le plus commode est qu'il soit égal à ce niveau). Cette tonalité d'étalonnage est en général de 1000 Hz mais elle peut avoir une autre fréquence si les enregistrements doivent être reproduits par l'intermédiaire de systèmes (comme certains types de codeur de sous-bande) qui réagissent de manière typique à 1000 Hz.

Cette tonalité peut être ensuite utilisée plus tard pour régler les niveaux moyens de parole d'entrée (voir B.4.3).

B.2 Choix des conditions du circuit

B.2.1 Niveaux d'entrée et d'écoute de la parole

Pour choisir les conditions du circuit, on portera une attention particulière:

- à la gamme des niveaux d'entrée;
- à la gamme des niveaux d'écoute:
 - il n'y a pas de niveau d'écoute optimum universel;
 - divers niveaux d'écoute interviennent dans la pratique;
 - considérations de possibilités de comparaison;
 - des interactions peuvent se produire.

Le *Manuel de téléphonométrie* explique en détail ces aspects en 2.5.8.1.

B.2.2 Locuteurs

Des processus complexes affectant souvent différemment les voix masculines et féminines, le montage expérimental doit prévoir deux types de voix comme facteur d'équilibrage; les notations de la parole masculine et féminine devront être évaluées séparément, et on n'en calculera la valeur moyenne que dans le cas où elles produisent des effets et des interactions principaux qui ne sont pas statistiquement différents.

De plus, pour diminuer le risque que les résultats dépendent trop des particularités des voix choisies, il est indispensable que plus d'une voix masculine et plus d'une voix féminine soient utilisées dans un montage équilibré.

B.2.3 Conditions de référence

Chaque expérience doit comprendre des conditions de référence, de manière que des expériences faites dans différents laboratoires ou à des moments différents dans un même laboratoire puissent être raisonnablement comparées. Ces conditions de référence peuvent comprendre l'appareil de référence à bruit modulé (MNRU) (*modulated noise reference unit*) conforme à la Recommandation P.81; d'autres dégradations commandées conviennent dans d'autre cas (par exemple, rapport signal/bruit; voir 3.2.3/P.83).

B.2.4 Autres conditions

Outre les spécifications de B.2.1 à B.2.3 inclus, d'autres conditions seront insérées en fonction de l'objet de l'essai. Par exemple, le bruit de salle pourra être une variable, ainsi qu'un taux d'erreur binaire pour un système numérique ou l'évanouissement de Rayleigh pour un système radioélectrique.

B.3 Montage de l'expérience

Le montage de l'expérience repose sur les mêmes principes que ceux que donne A.2.

De plus, ce montage doit tenir compte:

- a) des spécifications du B.2;
- b) de l'effet de l'ordre de présentation.

Pour un échantillon donné de sujets, l'essai est limité par la longueur maximale possible pour la séance sans fatigue. Si l'expérience est trop importante pour être réalisée en une seule séance, il est prudent de la subdiviser en deux séances, voire davantage. Théoriquement, aucune séance ne devrait durer plus de 20 minutes et en aucun cas une séance ne devrait dépasser 45 minutes.

B.4 Procédure des essais d'écoute

B.4.1 Conditions ambiantes d'écoute

La salle d'écoute doit satisfaire aux mêmes conditions que la salle d'enregistrement (voir B.1.1) si ce n'est que le bruit ambiant (voir A.1.1.2.2) doit être réglé au niveau approprié. Voir A.1.1.2.2.1 et A.1.1.2.2.2 pour des exemples de spectre de bruit.

Il est recommandé que le niveau et le spectre de bruit soient mesurés au moins deux fois, à savoir au début et à la fin de l'expérience. Toute variation importante entre les deux mesures, comparées l'une à l'autre, doit être évaluée par l'expérimentateur, car elle peut jeter un doute sur la validité de l'expérience.

B.4.2 Système d'écoute

Quel que soit le système d'écoute choisi (par exemple un système téléphonique local, un système de référence intermédiaire spécifié dans la Recommandation P.48 ou un système à haut-parleur), il faut l'étalonner conformément à la Recommandation pertinente (par exemple, la Recommandation P.64) et décrire complètement la caractéristique d'efficacité en fonction de la fréquence.

Il est recommandé que la caractéristique d'efficacité en fonction de la fréquence de la connexion soit mesurée au moins deux fois, à savoir au début et à la fin de l'expérience. Toute variation importante entre les deux mesures, comparées l'une à l'autre, doit être évaluée par l'expérimentateur, car elle peut jeter un doute sur la validité de l'expérience.

B.4.3 Niveau d'écoute

Le gain du système doit être fixé de manière que la tonalité d'étalonnage (voir B.1.8) reproduite à partir des bandes traitées produise le niveau d'écoute voulu.

Les variations du niveau d'écoute qu'exige le montage de l'expérience peuvent être acceptées

- a) si l'on utilise des affaiblisseurs/amplificateurs dans le système d'écoute; ou
- b) être incluses au stade du traitement ou de la reproduction.

Cette seconde méthode n'est pas recommandée, parce qu'il est difficile de conserver un rapport signal/bruit suffisamment élevé à un faible niveau et parce que la souplesse et la variété de la randomisation sont fortement réduites.

Le niveau d'écoute devrait toujours être enregistré; le paragraphe 2.5 du *Manuel de téléphonométrie* donne des précisions à ce sujet.

B.4.4 Personnes qui écoutent

Les sujets qui participent aux essais d'écoute sont choisis au hasard parmi la population normale qui utilise le téléphone, pour autant:

- a) qu'ils n'aient pas directement participé à des travaux d'évaluation de la qualité des circuits téléphoniques ou à des travaux connexes comme le codage de la voix;
- b) qu'ils n'aient participé à aucun essai subjectif au cours des six mois précédents au moins, ni à des essais d'appréciation d'écoute depuis au moins un an; et
- c) qu'ils n'aient jamais entendu les mêmes listes de phrase auparavant.

Si la population disponible est trop restreinte, il faut en tenir compte pour tirer des conclusions des résultats.

Dans certains cas le tri des sujets peut être nécessaire et une méthode fondée sur l'Annexe B/P.78 pourra être applicable.

B.4.5 Echelles d'appréciation subjective recommandées par le CCITT

On peut utiliser diverses échelles d'appréciation de catégories à 5 notes pour différentes applications. La présentation et le libellé des échelles d'appréciation, tels qu'ils sont perçus par les sujets lors des essais sont très importants et doivent satisfaire à la norme atteinte après des années d'expérience. Les échelles d'appréciation suivantes sont les plus fréquemment employées pour les applications du CCITT et un libellé équivalent doit être utilisé en fonction de la langue, ce qui peut se traduire par de petites variations par rapport au texte anglais d'origine:

- a) *Echelle de qualité d'écoute*

<i>Qualité de la parole</i>	<i>Note</i>
Excellente	5
Bonne	4
Passable	3
Médiocre	2
Mauvaise	1

La quantité évaluée d'après les notes (note moyenne d'opinion de qualité d'écoute) est représentée par le symbole Y_{LQ} , mais on utilise le symbole Y_{lq} quand une notation avec indice n'est pas disponible.

- b) *Echelle des efforts d'écoute*

L'en-tête de l'échelle d'appréciation des efforts d'écoute est particulièrement important, car sans lui, les autres descriptions risquent d'être très mal comprises.

<i>Effort nécessaire pour comprendre le sens des phrases</i>	<i>Note</i>
Détente absolue; aucun effort	5
Attention nécessaire, pas d'effort appréciable	4
Effort modéré	3
Effort considérable	2
Incompréhensible en dépit de tous les efforts possibles	1

La quantité évaluée d'après les notes (note moyenne d'opinion d'effort d'écoute) est représentée par le symbole Y_{LE} mais en l'absence de notation avec indice, le symbole Y_{le} est utilisé.

- c) *Echelle de niveau sonore préféré*

<i>Niveau sonore préféré</i>	<i>Note</i>
Beaucoup plus fort que préféré	5
Plus fort que préféré	4
Selon préférence	3
Plus faible que préféré	2
Bien plus faible que préféré	1

La quantité évaluée d'après les notes (note moyenne d'opinion de niveau sonore préféré) est représentée par le symbole Y_{LP} mais en l'absence de notation avec indice, le symbole Y_{lp} est utilisé.

NOTE – On trouvera en 2.6 du *Manuel de téléphonométrie* et dans le Rapport 751 du CCIR, Volume VIII.3, 1986, des exemples d'autres échelles d'appréciation subjective qui ne doivent être utilisées que si les trois échelles d'appréciation ci-dessus ne répondent pas aux besoins de l'opérateur.

B.4.6 Directives pour les sujets

Le Tableau B.2 donne un exemple de directives types. Celles-ci doivent être données (oralement aussi, si nécessaire) avant le début de l'expérience. Quand le sujet a compris les directives, il doit écouter la liste préliminaire et donner son opinion. Il ne faut pas suggérer au sujet que les échantillons préliminaires comprennent le meilleur ou le pire dans la gamme à couvrir, ou qu'ils contiennent toute la gamme des conditions que le sujet est censé entendre. Après la liste préliminaire, il faut prévoir un délai suffisant pour répondre aux questions possibles des sujets. Les questions relatives à la procédure ou au sens des directives doivent donner lieu à des réponses, mais il faut donner comme réponse à toute question d'ordre technique «Nous ne pouvons rien vous dire à ce sujet jusqu'à la fin de l'expérience».

TABLEAU B.2/P.80

Directives pour les sujets

EXPÉRIENCE D'ÉCOUTE N° ...	
Dans la présente expérience, vous écouterez de brefs groupes de phrases par l'intermédiaire du combiné téléphonique et vous donnerez votre opinion des paroles que vous entendrez.	
Une boîte avec 5 boutons-poussoirs lumineux se trouve devant vous sur la table. Quand toutes les lampes s'allument vous entendrez ... phrases. Ecoutez-les et quand les lampes s'éteignent, appuyez sur le bouton approprié pour indiquer votre opinion selon l'échelle suivante.	
EFFORT QU'EXIGE LA COMPRÉHENSION DES PHRASES	
5	Détente absolue, aucun effort
4	Attention nécessaire, pas d'effort appréciable
3	Effort modéré
2	Effort considérable
1	Incompréhensible en dépit de tous les efforts possibles
Le bouton sur lequel vous venez d'appuyer s'allumera pendant un bref moment. La lampe s'éteindra ensuite et il y aura une brève pause avant que toutes les lampes se rallument pour le groupe suivant de ... phrases.	
Il y aura une plus longue pause chaque fois après ... groupes (pour chacun desquels il faut donner une opinion). Il y aura au total ... groupes lors de cette visite, et un nombre analogue lors de votre (vos) prochaine(s) visite(s).	
Nous vous remercions de votre assistance pour cette expérience.	

B.4.7 Analyse statistique et annonce des résultats

La moyenne numérique (entre les sujets) doit être calculée pour chaque condition à chaque niveau d'écoute et ces moyennes seront énumérées pour examen préliminaire (de manière que certains effets, comme ceux dus aux voix masculines et féminines puissent être décelés).

Le calcul d'écart types distincts pour chaque condition n'est pas recommandé. Les limites de confiance doivent être évaluées et des essais de signification seront faits selon les techniques classiques d'analyse de variance.

NOTE – Dans le passé, les équivalences, par exemple Excellent = 4, Bon = 3, Passable = 2, Médiocre = 1 et Mauvais = 0 ont été souvent utilisées. Si l'on utilise donc les résultats d'expériences antérieures, il ne faut pas oublier d'augmenter d'une unité toutes les notes, pour qu'elles soient comparables à celles que l'on vient d'obtenir, le traitement numérique que l'on peut appliquer restant par ailleurs le même.

La méthode d'analyse des échelles d'appréciation subjective du B.4.5.1 repose sur les principes énoncés en A.4.5.

Pour faciliter encore l'examen des données, on tracera, s'il y a lieu, des graphiques montrant la note moyenne d'opinion en fonction du paramètre à mesurer, par exemple Y_{LE} en fonction de l'affaiblissement du circuit. L'axe vertical sur le graphique doit toujours être la note Y_{LE} .

On fera avec soin une moyenne des notes pour les locuteurs masculins et féminins, ce qui n'implique pas que cette opération se justifie pour une étude détaillée et pour l'interprétation des résultats, sauf si cela est justifié par les essais de signification.

Annexe C

Essais de détection de la réponse discontinue alternative

(Cette annexe fait partie intégrante de la présente Recommandation)

Pour obtenir des renseignements sur la possibilité de détection ou sur une autre caractéristique analogue d'un son (comme l'écho) en fonction d'une quantité objective (comme le niveau d'écoute) la meilleure méthode est celle de la réponse discontinue alternative dont le principe est semblable à celui décrit en [12], qui repose sur la «méthode à trois opérateurs sans affaiblissement caché» pour déterminer les équivalents de référence et les équivalents relatifs – cette procédure est décrite en 2.2 du *Manuel de téléphonométrie*. La principale différence est que la réaction des sujets n'est pas une décision sous la forme de «Référence» ou d'«Essai» (désignation du plus fort de deux circuits) mais un vote sur une échelle comme en [13]:

Echelle d'appréciation subjective de possibilité de détection

- A Gênant
- B Décelable
- C Indécelable

dans laquelle on suppose que B signifie «Décelable mais pas gênant».

On peut utiliser des échelles de cette sorte, comportant en général trois notes, pour divers essais de réponse discontinue alternative; par exemple, l'échelle décrite ci-dessus peut être utilisée quand le stimulus est l'écho, la réverbération, l'effet local, la mutilation de commutation de voix ou des tonalités brouilleuses, alors que la diaphonie, voire l'écho dans certains cas, peut être estimée d'après l'échelle Intelligible – Décelable – Indécelable.

Il est parfois admissible de considérer ces votes comme des notes d'opinion en leur donnant les valeurs respectives 2, 1 et 0 et de les traiter de la même manière que des notes d'opinion d'écoute ou de conversation. Souvent, néanmoins, cela n'est pas satisfaisant parce que les décisions sur une échelle de possibilité de détection (voir ci-dessus) ne constituent pas vraiment des équivalents de réactions sur une échelle continue alors qu'on peut légitimement considérer que c'est bien le cas de votes sur des échelles comme la «préférence de sonie» (voir le B.4.5.1) – mais représentent en fait deux dichotomies distinctes (par exemple décelable/indécelable ou gênant/non gênant), qui sans être indépendantes peuvent faire appel à des processus psychologiques différents: en d'autres termes, le caractère gênant ou l'intelligibilité diffèrent par nature et pas seulement par ce niveau du caractère décelable. Pour cette raison, une méthode d'analyse plus profitable consiste à exprimer la probabilité de réaction séparément selon chaque dichotomie, en fonction d'une variable objective, en adaptant les équations aux probits ou aux logis, puis en utilisant les quantiles ou d'autres paramètres comme base de comparaison entre conditions de circuit, d'une manière analogue à celle qui est utilisée pour appliquer les notes de netteté.

La conduite effective des expériences de ce type ressemble à celle des essais d'effort d'écoute (Annexe B), mais il existe quelques différences. En particulier, il est préférable que la première présentation du signal à chaque série ait lieu à un niveau d'écoute élevé, afin que l'auditeur n'ait pas de doute sur le type de signal qu'il doit utiliser pour ses décisions. Quand l'effet local ou l'écho interviennent, le sujet sera tenu de parler et d'écouter.

De simples mesures audiométriques, décrites dans la Recommandation P.78 [14] sont d'ordinaire effectuées sur des sujets qui participent à ces expériences, de sorte que les résultats peuvent être exprimés par rapport à leur seuil d'audibilité.

On trouvera en [15], dans les paragraphes 3.5.3 et 4.5.1 des exemples d'application de ces techniques.

On étudie parfois le bruit, l'évanouissement et d'autres perturbations au moyen de réactions sur une échelle comprenant beaucoup plus de notes, par exemple [16]:

- A *Inaudible* – Bruit absolument indécelable
- B *Juste audible* – Une écoute attentive permet tout juste de déceler le bruit.
- C *Léger* – Bruit décelable, mais pas gênant.

- D *Modéré* – Bruit légèrement gênant.
- E *Plutôt fort* – Le bruit cause une perturbation appréciable.
- F *Fort* – Le bruit est très gênant, mais la communication continue.
- G *Intolérable* – Le bruit est si fort que la communication est abandonnée, ou que l'opérateur est prié de changer la ligne.

Ces échelles sont plus proches du type continuum quantifié, comme l'échelle de préférence de sonie, et on peut les traiter de la même façon.

Annexe D

Méthode de notation par catégorie de dégradation (DCR)

(Cette annexe fait partie intégrante de la présente Recommandation)

D.1 Introduction

La notation par catégorie absolue (ACR) (*absolute category rating*) a tendance à rendre moins efficace la distinction entre circuits de bonne qualité. Une version modifiée de cette méthode (voir l'Annexe B), appelée notation par catégorie de dégradation (DCR) (*degradation category rating*) [17], assure une plus grande efficacité. Cette procédure est adaptée de la Recommandation du CCIR [18] pour l'évaluation des circuits de bonne qualité. La procédure DCR, qui repose en particulier sur une échelle de perturbation et sur une référence de qualité élevée avant chaque configuration à évaluer, semble convenir pour évaluer la parole de bonne qualité.

D.2 Procédure de la notation par catégorie de dégradation (DCR)

D.2.1 Echantillons de parole

Chaque configuration est évaluée au moyen de jugements concernant quatre locuteurs lisant deux échantillons différents. Chaque échantillon doit comporter deux phrases séparées par une seconde au moins. Ces deux échantillons (S1 et S2), soit quatre phrases différentes, doivent être choisis dans un ensemble plus large composé de phrases équilibrées phonétiquement, afin que la note moyenne obtenue de l'évaluation des circuits de référence (par exemple MNRU pour les processus numériques) pour ces phrases soit à peu près la même que celle obtenue pour l'ensemble plus large. L'ensemble (corpus) se compose donc de huit échantillons définis ainsi:

- échantillons S1, S2 à lire par le locuteur T1;
- échantillons S1, S2 à lire par le locuteur T2;
- échantillons S1, S2 à lire par le locuteur T3;
- échantillons S1, S2 à lire par le locuteur T4.

Il en résulte une répétition des deux échantillons pendant l'essai. Mais nous pensons que cela importe peu pour la procédure, quand une dégradation est évaluée par rapport à la référence, spécialement pour la bonne qualité téléphonique quand l'intelligibilité de la parole est presque parfaite. Le recours à des échantillons différents pour chaque configuration, comme cela se fait pour les expériences ACR, dans lesquelles les effets du locuteur et de la phrase se mélangent, pourrait expliquer le manque d'efficacité de cette procédure. Quelques variations de ce plan de base sont permises: augmenter le nombre de locuteurs, combiner les effets de phrases et les locuteurs, mais chaque configuration doit être évaluée sur le même corpus.

D.2.2 Conditions de référence

Des conditions de référence seront incluses; par exemple pour les processus numériques, le bruit multiplicatif avec des valeurs Q dans la gamme 10 30 dB avec un maximum de quatre échelons est souhaitable.

On choisira une référence de qualité pour l'insérer avant chaque jugement. On utilise habituellement des signaux originaux, c'est-à-dire des échantillons avec une dégradation qui n'est pas supérieure à celle qu'introduisent les systèmes d'émission et les limitations de largeur de bande de fréquences. Le choix de la référence de qualité dépend donc de l'application, c'est-à-dire que, pour la téléphonie normale, le signal de la source est limité à la largeur de bande de 3,4 kHz, que pour la téléphonie à large bande, il est limité à la bande de 7 kHz et que, pour le son de haute qualité, le signal est limité à la bande de 15 ou 20 kHz.

D.2.3 Présentation des stimuli

Les stimuli sont présentés aux auditeurs par paires (A-B) ou par paires répétées (A-B-A-B), dans lesquelles A est l'échantillon de référence de qualité élevée et B le même échantillon traité par un codec. L'échantillon de référence a pour but d'ancrer chaque jugement des auditeurs. L'utilisation d'une référence et de jugements subjectifs par rapport à cette référence est une procédure très courante en psychoacoustique. Elle tend à se traduire par une bonne efficacité pour l'évaluation globale fournie par les auditeurs. Les échantillons A et B doivent être séparés par 0,5 s et dans la procédure de répétition de paire (A-B-A-B) la séparation entre les deux paires doit être de 1 s.

Il semble que l'effet d'ordre classique observé dans un essai d'écoute d'un échantillon (ACR par exemple) n'est pas constaté avec la procédure DCR. Ainsi, un seul ordre aléatoire de présentation peut être utilisé. Par conséquent, les conditions d'essai et de référence de base correspondent à huit fois (quatre locuteurs × deux échantillons) le nombre des conditions nominales. La temporisation des réactions des auditeurs est la même que pour l'essai ACR, à savoir 5 s entre deux présentations (paires ou paires répétées).

D.2.4 Directives pour les essais

Les sujets doivent avoir pour instructions de noter les conditions selon l'échelle de catégorie de dégradation à cinq notes suivante:

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

La quantité évaluée d'après les notes (notes d'évaluation moyenne par catégorie de dégradation) est représentée par le symbole Y_{DCR} mais, en l'absence de notation avec indice, le symbole Y_{dcr} est utilisé.

D.3 Analyse statistique

Les efficacités peuvent être quantifiées au moyen d'un essai de comparaison statistique multiple. Quand une comparaison de circuits *a posteriori* est nécessaire, un essai de «différences honnêtement significatives» (HSD) (*honestly significant difference*) de Tuckey [19] peut être appliqué efficacement. Le test HSD permet de faire toutes les comparaisons par paires parmi les moyennes et de déterminer si les différences entre les moyennes sont significatives.

Annexe E

Méthode du seuil pour la comparaison de systèmes de transmission avec un système de référence

(Cette annexe fait partie intégrante de la présente Recommandation)

E.1 Introduction

La comparaison directe d'un système de transmission avec un système de référence permet d'évaluer la qualité du système à mesurer en ce qui concerne la caractéristique de dégradation du système de référence qui peut être modifiée et réglée selon des valeurs définies. Le rapport signal/bruit (pour la définition, voir 3.2.3/P.83) SNR(p) donne un exemple d'une telle caractéristique. La méthode décrite ici conduit à un seuil de qualité défini comme un niveau de préférence de 50% entre le MNRU et le système numérique.

E.2 Procédure à suivre pour les essais

Une procédure d'essai d'écoute seulement est présentée. Une paire de signaux comprenant un signal de référence et un signal d'essai est présentée aux auditeurs qui sont priés d'indiquer lequel des deux ils estiment présenter la meilleure qualité (évaluation de préférence). L'équivalent subjectif est défini comme la valeur de référence correspondant au point d'intersection de la courbe de régression des notes de préférence au niveau de préférence de 50%. Un exemple de rapport signal/bruit obtenu avec des notes de préférence fictives est représenté à la Figure E.1.

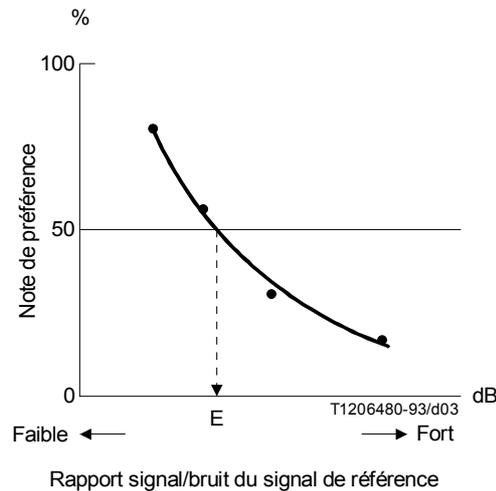


FIGURE E.1/P.80

Exemple de seuil d'équivalence E avec note de préférence fictive

E.3 Présentation des signaux

Un signal de référence A et un signal d'essai B sont présentés en un nombre égal de paires A-B et de paires B-A selon un ordre aléatoire. Différents niveaux de dégradation séparés, par exemple, par des intervalles de 2 dB, sont introduits dans le trajet de référence en sorte que la gamme des notes de préférence s'étage de 20% à 80%, le niveau de préférence de 50% se trouvant au milieu de la gamme des dégradations. Un diagramme temporel de la présentation est représenté à la Figure E.2.



T1206490-93/d04

FIGURE E.2/P.80

Diagramme temporel de la présentation

Le sujet doit formuler un jugement et répondre en disant «A est meilleur» ou «B est meilleur» (choix forcé). La réponse «A égale B» ou «aucune différence» est interdite. La durée de la présentation devrait être limitée à 6 minutes environ afin de ne pas fatiguer la personne qui écoute. Un plus grand nombre d'échantillons d'écoute sont présentés après une période de repos appropriée. Au moins 2, mais de préférence 4 ou 5 répétitions (reprises de présentations identiques) sont recommandées.

NOTE – Si le système de référence est disponible dans le matériel et si la caractéristique de dégradation peut être facilement changée entre les présentations, on peut utiliser une procédure simplifiée. Dans ce cas, l'équilibre entre des perceptions de qualité égale est effectué par le sujet. L'ajustement se fait pendant la pause entre les paires. Le signal de référence est toujours présenté en premier. La présentation se poursuit jusqu'au moment où le sujet indique que le seuil d'égalité a été atteint.

E.4 Sources de paroles

Il est nécessaire d'utiliser des phrases courtes prononcées au moins par 2 hommes et par 2 femmes (de préférence 4 ou 6). Chaque personne qui parle doit prononcer des phrases différentes. La durée doit être de 2,5 à 5 secondes pour la parole et de moins de 10 à 15 secondes pour des signaux de musique. Il faut éviter que se produisent des clics au début et à la fin des échantillons. Un microphone linéaire ayant une largeur de bande suffisante doit être utilisé pour enregistrer les signaux source dans une salle anéchoïque où le bruit ambiant est inférieur à 20 dBA avec un temps de réverbération inférieur à 0,3 seconde dans la bande de 125 à 8000 Hz. Si l'on utilise un équipement d'enregistrement numérique, le niveau du bruit de quantification devrait être inférieur au niveau de bruit calculé sur 14 bits linéaires MIC.

E.5 Environnement d'écoute

Pour les essais d'écoute, il convient d'utiliser un système de lecture du son haute fidélité. Si l'écoute se fait par haut-parleurs, l'équipement de reproduction devra être de qualité studio et la salle d'écoute devra être conforme aux dispositions du Rapport 797 du CCIR ou à la publication 268-13 de la CEI. Si l'écoute se fait par casque, il faut donner la préférence à une écoute diotique (binaurale). La largeur de bande devra être au moins égale à celle du système numérique soumis aux essais.

E.6 Personnes qui écoutent

S'il est préférable que les participants qui écoutent soient choisis conformément à la description faite pour la méthode ACR (voir l'Annexe B), il ne s'agit cependant pas là d'une condition stricte dans l'essai de comparaison par paires. Si les essais d'écoute ont pour but de recueillir les opinions de personnes non averties, il faut utiliser des sujets non avertis. Mais, si tel n'est pas le cas, il faut utiliser des sujets informés et la fiabilité de l'essai d'écoute peut être améliorée en augmentant le nombre de répétitions de l'essai par sujet. Le nombre minimal de participants est de 6; de préférence, on en utilisera 12 ou plus. Plusieurs sujets peuvent écouter simultanément le même essai, mais il faut s'assurer qu'ils répondent bien indépendamment les uns des autres.

E.7 Fiabilité

Etant donné que l'on admet par hypothèse que les variations des notes de préférence dans les essais subjectifs obéissent à une distribution t, la largeur de la variation de la note avec r donne une fiabilité de 95% à la note u ($0 \leq u \leq 1$) sur l'ensemble des essais (nombre de répétitions pour chaque paire de présentations multiplié par le nombre de sujets et par le nombre de signaux source) est présentée sur la formule (E-1):

$$r = \pm t(n - 1, 0,05) \cdot \sqrt{u(1 - u)/(n - 1)} \quad (\text{E-1})$$

NOTE – On s'attend que la méthode de seuil donnera des résultats stables et fiables, même pour des systèmes de qualité élevée avec peu de dégradations.

Des dégradations peuvent être introduites dans le système de référence, par exemple en ajoutant un bruit blanc. Pour les systèmes numériques le bruit multiplicatif défini dans la Recommandation P.81 (MNRU) est recommandé. Pour les codeurs numériques à large bande, l'emploi d'un MNRU à large bande décrit dans l'Annexe A/P.81 est recommandé. Pour certains besoins, il conviendra d'utiliser un bruit spécialement étudié au lieu du bruit blanc.

Références

- [1] VOIERS (W.D.): Evaluating processed speech using the Diagnostic Rhyme Test, Speech Technology, Volume 1, n° 4, pages 30 à 39, janvier-février 1983.
- [2] CCITT *The SIBYL method of subjective testing, Livre rouge*, Volume V, Supplément n° 5 de la Recommandation P.74.

- [3] BERANEK (L.L.): Noise and Vibration Control, pages 564 à 566, *McGrawHill*, 1971.
- [4] ISO *Assessment of noise with respect to community response*, Norme ISO 1996, 1971.
- [5] HOTH (D.F.): Room noise spectra at subscribers' telephone locations, *J.A.S.A.*, Volume 12, pages 99 à 504, avril 1941.
- [6] Publication de la Recommandation de la CEI 179, *Precision sound level meters*, 1965.
- [7] Question 24/XII du CCITT, contribution COM XII-120, *Noise inside light motor vehicles*, Période d'études 1981-1984.
- [8] Question 24/XII du CCITT, contribution COM XII-134, *Internal vehicle noise spectra*, Période d'études 1981-1984.
- [9] Contribution COM XII-208 du CCITT, *Comparison of the results of vehicle noise submitted by France and BT*, Période d'études 1981-1984.
- [10] STEVENS (S.S.): Psychophysics: Introduction to its perceptual, neural, and social prospects, *John Wiley and Sons*, 1975.
- [11] CLARINGBOLD (P.J.): The within-animal bioassay with quantal responses, *Journal of the Royal Statistical Society*, Series B, Volume 18, n° 1, pages 133 à 137, 1956.
- [12] CCITT *Mesure des équivalents de Référence et des équivalents relatifs*, Recommandation P.72, § 2.1.2, *Livre jaune*, Volume V, 1981.
- [13] RICHARDS (D.L.): Telecommunication by speech, § 3.5.2, *Butterworths*, Londres, 1973.
- [14] CCITT: *Méthodes d'essais subjectifs pour déterminer les équivalents pour sonie conformément à la Recommandation P.76*, Recommandation P.78, Annexe B, *Livre Bleu*, Volume V, 1989.
- [15] RICHARDS (D.L.): Telecommunication by speech, § 3.5.3 et § 4.5.1, *Butterworths*, Londres, 1973.
- [16] Ibid, § 4.2.1.6.
- [17] COMBESURE (P.) et autres: Quality evaluation of speech coded at 32 kbit/s by means of degradation category ratings, *Proc. ICASSP 82* (International Conference on Acoustics, Speech and Signal Processing), Volume 2, Paris, mai 1982.
- [18] CCIR Subjective assessment of the quality of television pictures (EBU), Document 11/17, Période d'études 1978-1982.
- [19] TUCKEY (J.W.): The problem of multiple comparisons, *Ditton*, Princetown University, Ed. 1953.

Bibliographie

GABRIELSSON (A.): Statistical treatment of data from listening tests on sound-reproducing systems, Rapport TA n° 92, novembre 1979, *KTH Karolinska Institutet*, Department of Technical Audiology, S 10044 Stockholm, Suède.

Publication CEI 268-13, Annexe 3, § 3.3 (version condensée du rapport ci-dessus).

Imprimé en Suisse

Genève, 1994