



UNION INTERNATIONALE DES TÉLÉCOMMUNICATIONS

UIT-T

SECTEUR DE LA NORMALISATION=
DES TÉLÉCOMMUNICATIONS
DE L'UIT

P.502

(05/2000)

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

Appareils de mesures objectives

**Méthodes d'évaluation objective des systèmes
de communication vocale utilisant des signaux
de test complexes**

Recommandation UIT-T P.502

(Antérieurement Recommandation du CCITT)

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

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

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

Recommandation UIT-T P.502

Méthodes d'évaluation objective des systèmes de communication vocale utilisant des signaux de test complexes

Résumé

La présente Recommandation décrit des méthodes et des procédures d'évaluation de terminaux, de composants de réseaux et de systèmes de transmission complexes. Les méthodes de test utilisent, pour la plupart, les signaux de test décrits dans les Recommandations UIT-T P.50, P.59 et P.501. Pour diverses implémentations techniques et situations de conversation, on donne les incidences possibles sur la qualité vocale telle qu'elle est perçue subjectivement et on décrit les procédures de mesure correspondantes.

Source

La Recommandation P.502 de l'UIT-T, élaborée par la Commission d'études 12 (1997-2000) de l'UIT-T, a été approuvée le 18 mai 2000 selon la procédure définie dans la Résolution 1 de la CMNT.

Mots clés

Conversation simple, méthodes d'analyse, parole simultanée, qualité vocale.

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 1 de la CMNT.

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

NOTE

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

DROITS DE PROPRIÉTÉ INTELLECTUELLE

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

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

© UIT 2001

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

TABLE DES MATIÈRES

		Page
1	Domaine d'application	1
2	Références.....	1
3	Définitions et abréviations	2
4	Caractéristiques de convergence des annuleurs d'écho.....	3
4.1	Dégradation de la qualité vocale telle qu'elle est perçue subjectivement	3
4.2	Paramètres objectifs associés pour le mode conversation simple.....	4
4.3	Méthodes d'analyse	4
4.3.1	Méthode d'évaluation du temps de convergence (T_c).....	4
4.3.2	Affaiblissement d'adaptation pour l'écho, avec pondération temporelle (ERL_{tst}) – Conversation simple	5
5	Evaluations de la qualité vocale en période de parole simultanée.....	6
5.1	Dégradation de la qualité vocale telle qu'elle est perçue subjectivement	6
5.2	Paramètres objectifs associés.....	6
5.3	Méthodes d'analyse	7
5.3.1	Méthode utilisant des signaux CSS en période de parole simultanée	7
5.3.2	Test en période de parole simultanée, avec séquences combinées parallèles.....	11
6	Caractéristiques de compression-extension et de commande AGC	20
6.1	Dégradation de la qualité vocale telle qu'elle est perçue subjectivement	21
6.2	Paramètres objectifs associés.....	21
6.3	Méthodes d'analyse	21
7	Qualité de transmission du bruit de fond.....	23
7.1	Dégradation de la qualité telle qu'elle est perçue subjectivement.....	23
7.2	Paramètres objectifs associés.....	23
7.3	Méthodes d'analyse	24
8	Caractéristiques de commutation.....	25
8.1	Dégradation de la qualité vocale telle qu'elle est perçue subjectivement	26
8.2	Paramètres objectifs associés.....	26
8.3	Méthodes d'analyse	27
8.3.1	Plage d'atténuation et caractéristiques de commutation	27
8.3.2	Niveau de seuil et temps d'établissement (pour le niveau minimal d'activation)	28
8.3.3	Temps de maintien.....	29
8.3.4	Niveau seuil et temps de commutation pour passer de RCV à SND (de SND à RCV).....	29

	Page
8.3.5 Caractéristiques de commutation en présence de bruit de fond.....	31
Annexe A – Méthode de mesure détaillée de l'affaiblissement ERL_t avec pondération temporelle	31
A.1 Algorithme relatif à l'affaiblissement d'adaptation pour l'écho.....	31
A.1.1 Affaiblissement d'adaptation pour l'écho, avec pondération temporelle (ERL_t)	32
A.1.2 Modélisation de l'audibilité de l'écho	32
A.1.3 Expression des résultats de mesure de ERL_t	34
A.1.4 Algorithme de mesure de ERL_t	35
Annexe B – Filtres de mesure en période de parole simultanée pour la méthode A.....	39
Annexe C – Description de la séquence de conditionnement	40
C.1 Conditionnement de l'annuleur avant la parole simultanée	40
C.1.1 Gabarits d'activité pour le conditionnement en période de parole simultanée.....	40
C.1.2 Synchronisation des gabarits d'activité pour le conditionnement en période de parole simultanée	42
C.1.3 Compensation associée aux filtres de mesure	42
Appendice I – Références bibliographiques.....	43
Appendice II – Exemples d'évaluation.....	44
II.1 Quelques exemples d'évaluation conformément au paragraphe 5	44
II.1.1 Réponses en fréquence en période de parole simultanée.....	44
II.1.2 Variations de niveau en période de parole simultanée	49
II.1.3 Commutation en période de parole simultanée	51

Introduction

La présente Recommandation décrit des méthodes et des procédures d'évaluation de terminaux, de composants de réseaux et de systèmes de transmission complexes. Les méthodes de test sont décrites en fonction des divers paramètres et systèmes à évaluer. Elles utilisent, pour la plupart, les signaux de test décrits dans les Recommandations UIT-T P.50, P.59 et P.501. Pour diverses implémentations techniques et situations de conversation, on donne les incidences possibles sur la qualité vocale telle qu'elle est perçue subjectivement et on décrit les procédures de mesure correspondantes.

Recommandation UIT-T P.502

Méthodes d'évaluation objective des systèmes de communication vocale utilisant des signaux de test complexes

1 Domaine d'application

Le but de la présente Recommandation est de définir des méthodes de test qui peuvent être utilisées pour évaluer certains artefacts ayant une incidence sur la qualité vocale de terminaux et de systèmes de transmission de la parole. Ces méthodes sont fondées sur les signaux de test définis dans les Recommandations P.50, P.59 et P.501.

La Recommandation décrit un ensemble de méthodes de test qui permettent d'examiner divers paramètres jugés importants pour l'évaluation des systèmes de communication vocale. Chaque paramètre de qualité de fonctionnement est caractérisé par la dégradation vocale telle qu'elle est perçue subjectivement et par les paramètres objectifs associés. Pour chaque paramètre, des méthodes d'analyse sont décrites.

2 Références

La présente Recommandation se réfère à certaines dispositions des Recommandations UIT-T et textes suivants qui, de ce fait, en sont partie intégrante. Les versions indiquées étaient en vigueur au moment de la publication de la présente Recommandation. Toute Recommandation ou tout texte étant sujet à révision, les utilisateurs de la présente Recommandation sont invités à se reporter, si possible, aux versions les plus récentes des références normatives suivantes. La liste des Recommandations de l'UIT-T en vigueur est régulièrement publiée.

- Recommandation UIT-T G.122 (1993), *Influence des systèmes nationaux sur la stabilité et l'écho pour la personne qui parle dans les connexions internationales.*
- Recommandation UIT-T G.168 (2000), *Annuleurs d'écho pour les réseaux numériques.*
- Recommandation UIT-T P.10 (1998), *Terminologie relative à la qualité de transmission téléphonique et aux appareils téléphoniques.*
- Recommandation UIT-T P.50 (1999), *Voix artificielle.*
- Recommandation UIT-T P.51 (1996), *Bouche artificielle.*
- Recommandation UIT-T P.56 (1993), *Mesure objective du niveau vocal actif.*
- Recommandation UIT-T P.57 (1996), *Oreilles artificielles.*
- Recommandation UIT-T P.58 (1996), *Simulateur de tête et de torse pour la téléphonométrie.*
- Recommandation UIT-T P.59 (1993), *Voix conversationnelle artificielle.*
- Recommandation UIT-T P.340 (2000), *Caractéristiques de transmission des postes téléphoniques mains-libres.*
- Recommandation UIT-T P.501 (2000), *Signaux de test à utiliser en téléphonométrie.*
- Recommandation UIT-T P.581 (2000), *Utilisation de simulateurs de tête et de torse pour les essais des terminaux mains-libres.*
- Recommandation UIT-T P.800 (1996), *Méthodes d'évaluation subjective de la qualité de transmission.*

- Recommandation UIT-T P.810 (1996), *Appareil de référence à bruit modulé (MNRU)*.
- Recommandation UIT-T P.830 (1996), *Evaluation subjective de la qualité des codecs numériques à bande téléphonique et à large bande*.
- *Manuel de téléphonométrie* de l'UIT-T, 2^e édition; Genève, 1992.
- Publication de la CEI 60651 (1979), *Sonomètres*.

3 Définitions et abréviations

La présente Recommandation définit les termes suivants:

caractéristiques AGC: caractéristiques de commande automatique du gain.

plage d'atténuation (a_H , *attenuation range*): plage d'atténuation, en dB, insérée dans le sens émission ou réception d'un terminal ou d'un système.

atténuation vocale dans le sens émission en conversation simultanée (A_{sdt} , *send speech attenuation during double talk*)

atténuation vocale dans le sens réception en conversation simultanée (A_{rdt} , *received speech attenuation during double talk*)

temps d'attaque: temps nécessaire pour activer complètement un trajet de transmission (par un compresseur-extenseur).

facteur de crête: rapport puissance de crête/puissance efficace d'un signal.

compression-extension: atténuation/amplification d'un signal en fonction du niveau.

signal de source composite (CSS, *composite source signal*): signal composé temporellement de différents éléments.

renforcement de l'affaiblissement d'adaptation pour l'écho (ERLE, *echo return loss enhancement*): mesure servant à déterminer l'amélioration perçue en ce qui concerne la perturbation due aux signaux d'écho.

affaiblissement d'adaptation pour l'écho (ERL, *echo return loss*): mesure servant à déterminer la perturbation perçue due aux signaux d'écho.

affaiblissement d'adaptation pour l'écho, conversation simultanée (ERL_{dt} , *echo return loss, double talk*): mesure servant à déterminer la perturbation perçue due aux signaux d'écho en conversation simultanée.

affaiblissement d'adaptation pour l'écho, avec pondération temporelle, conversation simple (ERL_{tst} , *echo return loss, temporally weighted, single talk*): mesure servant à déterminer la perturbation perçue due aux signaux d'écho en période de conversation simple et compte tenu de certains effets psychoacoustiques.

affaiblissement d'adaptation pour l'écho, avec pondération temporelle, conversation simultanée (ERL_{tdt} , *echo return loss, temporally weighted, double talk*): mesure servant à déterminer la perturbation perçue due aux signaux d'écho en conversation simultanée compte tenu de certains effets psychoacoustiques.

transformée de Fourier rapide (FFT, *fast fourier transformation*)

processus vocal markovien (MSMP, *Markov speech model process*)

Voir Recommandation UIT-T P.501.

processeur non linéaire (NLP, *non-linear processor*): processeur généralement utilisé dans les annuleurs d'écho pour couper l'écho résiduel.

séquence de bruit pseudo-aléatoire (séquence PN) (PN-sequence, *pseudo noise sequence*): bruit pseudo-aléatoire au contenu fréquentiel défini, dérivé par transformation de Fourier inverse d'un spectre fréquentiel prédéfini.

sens réception (RCV, *receiving direction*)

temps de libération: temps nécessaire pour désactiver complètement un trajet de transmission (par un compresseur-extenseur).

entrée réception (R_{in} , *receive input*): point d'accès de réception (électrique) d'un dispositif testé.

sens émission (SND, *sending direction*)

sortie émission (S_{out} , *send output*): point d'accès d'émission (électrique) d'un dispositif testé.

équivalent de couplage du terminal (TCL, *terminal coupling loss*): affaiblissement de l'écho dans un terminal, mesuré du point R_{in} au point S_{out} , y compris les équivalents SLR et RLR.

T_c (temps de convergence, *convergence time*)

Voir 4.3.1.

T_H (temps de maintien, *hang-over time*)

Voir Recommandation UIT-T P.340.

Tic (temps de convergence initial, *initial convergence time*)

Voir temps de convergence.

T_R (temps d'établissement, *build-up time*)

Voir Recommandation UIT-T P.340.

T_S (temps de commutation, *switching time*)

Voir Recommandation UIT-T P.340.

V_{TH} (niveau seuil, *threshold level*)

Voir Recommandation UIT-T P.340.

4 Caractéristiques de convergence des annuleurs d'écho

Le présent paragraphe décrit les caractéristiques de convergence des annuleurs d'écho. Il décrit des méthodes permettant d'évaluer les effets subjectifs de divers paramètres des annuleurs d'écho et il propose en outre des méthodes objectives permettant de décrire ces paramètres.

4.1 Dégradation de la qualité vocale telle qu'elle est perçue subjectivement

La profondeur de convergence, ou le renforcement de l'affaiblissement d'adaptation pour l'écho (ERLE, *echo return loss enhancement*), décrit la capacité d'un annuleur d'écho à annuler les signaux renvoyés dans le sens de transmission opposé via un trajet d'écho. Il peut s'agir d'un écho acoustique dans le cas d'un téléphone mains-libres, ou d'un écho hybride dans le cas d'une conversion deux à quatre fils. Un faible renforcement ERLE signifie que les signaux d'écho résiduel seront plus audibles.

Le temps de convergence décrit la rapidité avec laquelle l'annuleur d'écho atteint un état stable pour lequel le signal d'écho résiduel renvoyé est suffisamment atténué sans insertion d'affaiblissement dans l'un ou l'autre trajet de transmission vocale. C'est le temps qu'il faut pour atteindre

l'affaiblissement ERL à 3 dB près, ou un affaiblissement de [25] dB. Il faut qu'un annuleur d'écho ait une convergence rapide pour éviter que l'écho parvienne à un locuteur au début d'un appel.

Une rafale d'écho peut être générée lorsqu'un annuleur d'écho a des difficultés à converger sur un trajet d'écho donné. La dégradation subjective est fonction de la longueur/du niveau des rafales d'écho, de leur rapprochement, de leur nombre par minute ainsi que du retard du trajet d'écho.

4.2 Paramètres objectifs associés pour le mode conversation simple

Les caractéristiques de réduction de l'écho sont déterminées par les paramètres suivants:

- affaiblissement d'adaptation pour l'écho en fonction du temps, défini comme étant l'affaiblissement d'adaptation pour l'écho (ERL);
- affaiblissement d'adaptation pour l'écho avec pondération temporelle (ERL_t);
- temps que l'annuleur AEC met pour converger, défini comme étant le temps de convergence (T_c).

En outre, les paramètres suivants s'appliquent en situation de parole simultanée:

- qualité de fonctionnement duplex en fonction du temps;
- réponse en fonctionnement duplex pour les paramètres ci-dessus;
- réponse en termes d'atténuation en présence de dégradations environnementales ou du réseau.

Les paramètres relatifs à la situation de parole simultanée sont traités au paragraphe 5.

4.3 Méthodes d'analyse

Le présent sous-paragraphe propose quelques techniques de mesure objective pouvant servir à évaluer les paramètres ci-dessus. Si des annuleurs d'écho vocal du réseau sont testés, on trouvera un montage approprié dans la Recommandation UIT-T G.168. Les conditions de test des téléphones mains-libres figurent dans la Recommandation UIT-T P.340.

4.3.1 Méthode d'évaluation du temps de convergence (T_c)

La description de la méthode d'évaluation du temps de convergence des annuleurs d'écho figure dans la Recommandation UIT-T G.168. Le signal d'écho est donc mesuré au moyen d'un hypsomètre conforme à la CEI 60651. Un filtre de pondération exponentiel avec une constante de temps de 35 ms (CEI 60651, "Impulsion") est appliqué lors de l'intégration de la sortie de l'hypsomètre. Le signal de sortie mesuré est affiché sous la forme d'un diagramme représentant le niveau en fonction du temps. On donne généralement une limite en fonction du temps, qui ne doit pas être dépassée. Un exemple type d'un tel diagramme est donné sur la Figure 1.

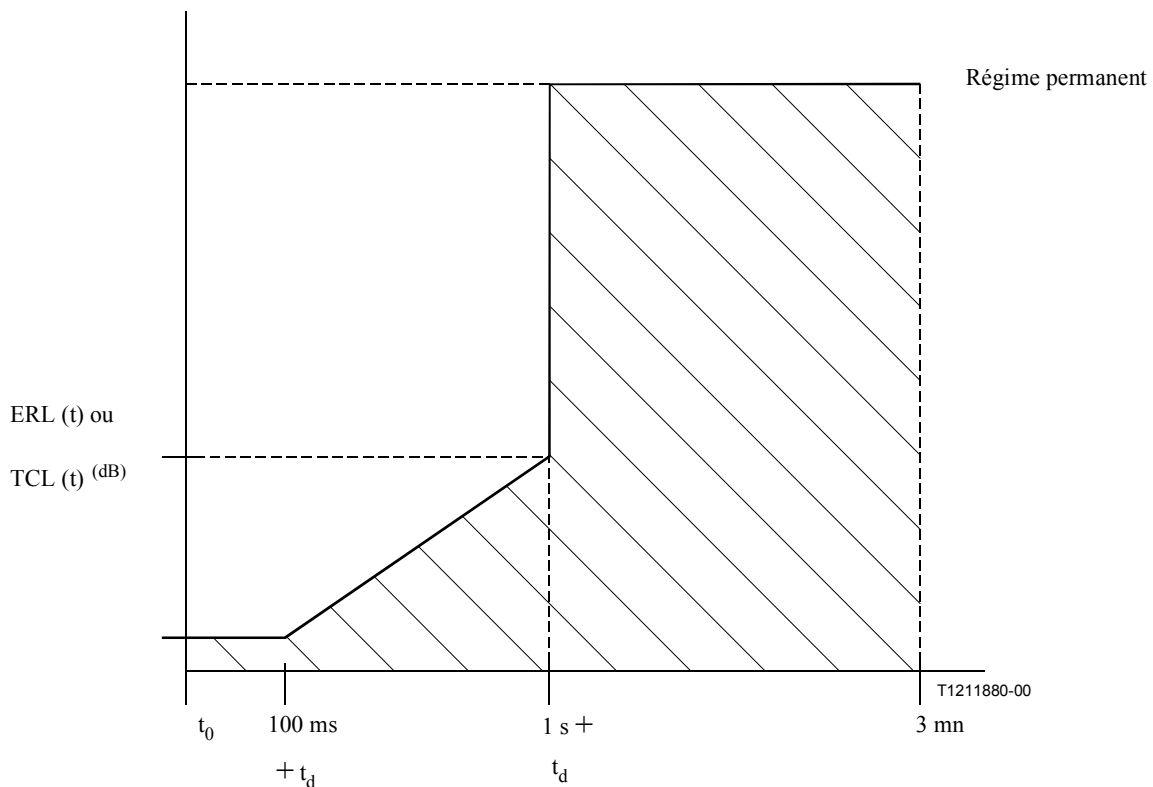


Figure 1/P.502 – Exemple de limite type de l'affaiblissement ERL ou de l'équivalent TCL en fonction du temps

L'affaiblissement ERL est généralement déterminé lors de la mesure des anneaux d'écho de réseau. L'équivalent TCL est défini comme étant l'équivalent de couplage d'un terminal, y compris les interfaces acoustiques.

NOTE – On ne donne pas de définition exacte de T_c . On pourrait par exemple le définir comme étant le temps mis pour atteindre un affaiblissement ERL de 3 dB au-dessus du régime permanent.

4.3.2 Affaiblissement d'adaptation pour l'écho, avec pondération temporelle (ERL_{tst}) – Conversation simple

Les méthodes de mesure de l'affaiblissement d'adaptation pour l'écho décrites dans les Recommandations UIT-T G.122 et G.168 sont les méthodes traditionnelles qui sont actuellement à la base de tous les calculs et de la planification. La méthode relative à l'affaiblissement ERL_t , plus proche des impressions subjectives compte tenu des effets de l'écho en fonction du temps, est décrite ci-dessous.

On mesure l'affaiblissement d'adaptation pour l'écho avec pondération temporelle à partir de l'interface réseau. Cette méthode permet de mesurer les rafales d'écho, mais peut aussi remplacer toute mesure de l'affaiblissement d'adaptation pour l'écho à long terme.

L'Annexe A donne un pseudo-code permettant d'implémenter cette méthode. Le signal de test à utiliser doit être aussi semblable que possible à la voix humaine. On peut utiliser d'autres signaux de test, mais ceux-ci risquent de donner des résultats optimistes.

Le signal de test est appliqué au point R_{in} (voir par exemple la Figure 10) pendant 30 secondes de sorte que les différentes unités fonctionnelles (en particulier l'anneau d'écho acoustique) puissent atteindre leur régime permanent. En cas de mesure d'anneaux d'écho acoustique, aucun signal autre que le retour acoustique du ou des haut-parleurs n'est appliqué au ou aux microphones.

Les signaux électriques sont enregistrés aux points R_{in} et S_{out} pendant la minute qui suit. On aligne ensuite dans le temps les enregistrements aux points R_{in} et S_{out} en ajoutant au signal R_{in} le retard causé dans le système entre R_{in} et S_{out} . La valeur ERL_{tst} , qui est fonction du temps, est la différence (en dB) entre le niveau du signal au point R_{in} et celui au point S_{out} calculée au moyen de l'algorithme donné dans l'Annexe A.

NOTE – Les trajets d'écho peuvent changer pendant la mesure, ils dépendent de l'environnement et de l'utilisation des équipements.

5 Evaluations de la qualité vocale en période de parole simultanée

La situation la plus critique dans toute conversation est celle de la parole simultanée. Les équipements faisant intervenir n'importe quel type de traitement du signal non linéaire ou variable dans le temps sont susceptibles d'entraîner une dégradation relativement forte de la qualité vocale (telle qu'elle est perçue subjectivement), notamment d'un paramètre du type "capacité de parole simultanée".

5.1 Dégradation de la qualité vocale telle qu'elle est perçue subjectivement

Les effets les plus gênants en période de parole simultanée sont les suivants:

- phrases, mots, syllabes interrompus ou transmis partiellement en période de parole simultanée ou peu après/avant;
- transmission vocale ou de bruit de fond avec un niveau variable dans le temps, entraînant des "variations de niveau en période de parole simultanée" qui sont gênantes;
- écho en période de parole simultanée.

En ce qui concerne la parole simultanée, les intervalles de temps situés peu avant et peu après la parole simultanée sont les plus critiques. Pendant ces intervalles, l'auto-masquage de la propre voix n'est plus efficace (voir Zwicker [5]).

En cas d'utilisation d'un annuleur d'écho, il faut que sa convergence soit rapide pour pouvoir faciliter rapidement la parole simultanée dès le début d'un appel. Cela signifie que le processeur non linéaire peut être supprimé plus tôt, permettant d'avoir une véritable situation de parole simultanée. La profondeur de convergence détermine l'audibilité d'un écho résiduel en période de parole simultanée. En cas de recours à un certain affaiblissement commuté pour renforcer l'affaiblissement d'adaptation pour l'écho, cela peut se traduire par une atténuation vocale audible en période de parole simultanée. Des rafales d'écho sont possibles en période de parole simultanée si des erreurs de détection de parole simultanée font diverger l'annuleur d'écho. Il en résulte des rafales d'écho de volume sonore relativement élevé.

5.2 Paramètres objectifs associés

Les paramètres objectifs associés sont les suivants:

- temps d'établissement (en période de parole simultanée);
- temps de maintien [temps d'extinction] (en période de parole simultanée);
- temps de commutation (en période de parole simultanée);
- plage d'atténuation (en période de parole simultanée);
- distribution de l'atténuation (en période de parole simultanée);
- réponses en fréquence;

- équivalents pour la sonie;
- variation du niveau en période de parole simultanée (caractéristiques de compression-extension).

Il faut en outre tenir compte des paramètres suivants, qui sont principalement associés à des implémentations d'annuleur d'écho:

- temps de convergence en période de parole simultanée;
- affaiblissement d'adaptation pour l'écho (mode parole simultanée): déterminé à partir du niveau de l'écho résiduel en période de parole simultanée;
- atténuation vocale émise (voir aussi plage d'atténuation) en période de parole simultanée: déterminée à partir de la valeur de l'atténuation vocale due à l'insertion d'un affaiblissement;
- affaiblissement d'adaptation pour l'écho avec pondération temporelle en période de parole simultanée: déterminé à partir du niveau pondéré des rafales d'écho.

NOTE – Lors de mesures objectives, il faut dans tous les cas noter si les réponses en fréquence, les équivalents pour la sonie et les niveaux/variations de niveau sont mesurés en période de parole simultanée, peu avant ou peu après.

5.3 Méthodes d'analyse

On peut choisir diverses méthodes pour évaluer la performance d'un système en période de parole simultanée. La description ci-dessous donne un aperçu des différentes techniques. La méthode décrite au 5.3.1 est une méthode générale qui ne suppose aucune implémentation technique spécifique du dispositif testé. Les méthodes décrites aux 5.3.2 à 5.3.4 supposent qu'un annuleur d'écho est implémenté.

5.3.1 Méthode utilisant des signaux CSS en période de parole simultanée

5.3.1.1 Construction du signal

On peut implémenter une méthode de mesure en situation de parole simultanée en utilisant le signal de test représenté sur la Figure 2. Ce signal de test est constitué d'une série de signaux de source composite non corrélés (Recommandation UIT-T P.501) qui sont injectés simultanément dans le sens émission et dans le sens réception. La séquence de test est construite de la façon suivante: les signaux envoyés dans le sens émission commencent avec un niveau élevé, ceux envoyés dans le sens réception commencent avec un bas niveau. On diminue le niveau de chaque séquence de source composite de 0,5 dB dans le sens émission et on l'augmente de 0,5 dB dans le sens réception. La différence totale entre le niveau maximal et le niveau minimal des signaux de source composite dans chaque sens est de 20 dB (30 dB pour les applications de réseau). Pour les terminaux mains-libres, on peut choisir les intervalles de niveau comme suit:

sens réception: -38 dBm à -18 dBm;

sens émission: $-4,7$ dBPa à $-24,7$ dBPa.

Pour des mesures en période de parole simultanée dans le réseau, on peut par exemple choisir les variations de niveau suivantes:

Sens réception: -40 dBm à -10 dBm;

Sens émission: -10 dBm à -40 dBm.

Des variations de niveau différentes sont bien évidemment possibles. Toutes les plages de niveau doivent dépendre de la plage dynamique qu'on souhaite évaluer.

La séquence est généralement construite de façon symétrique; autrement dit, lorsque le niveau minimal est atteint dans le sens émission, le niveau augmente à nouveau tandis que le niveau du signal dans le sens réception diminue à nouveau. La construction symétrique du signal permet en outre d'évaluer le comportement du dispositif testé dans le cas d'une symétrie.

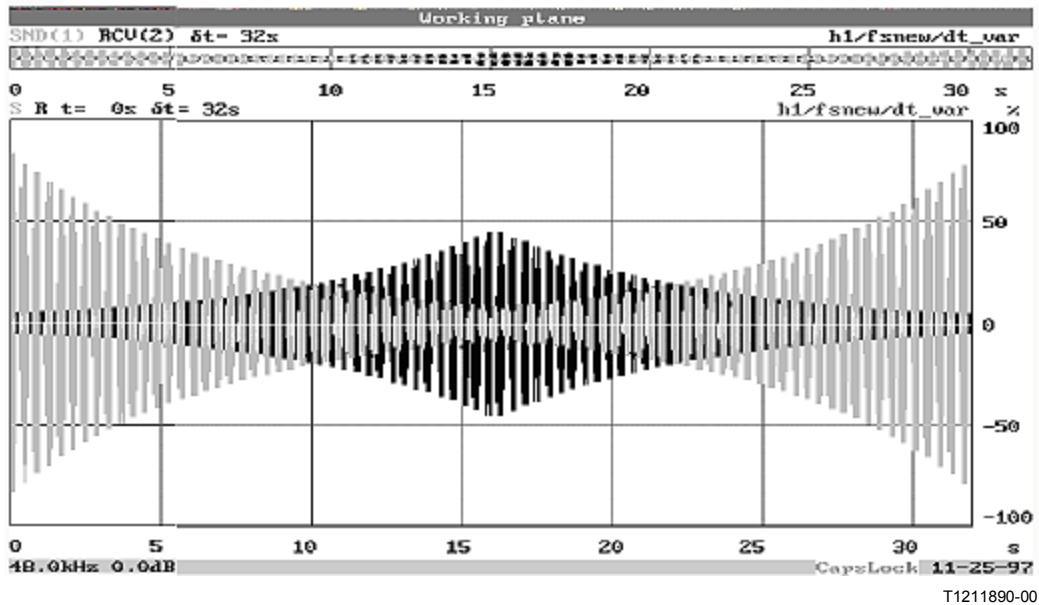
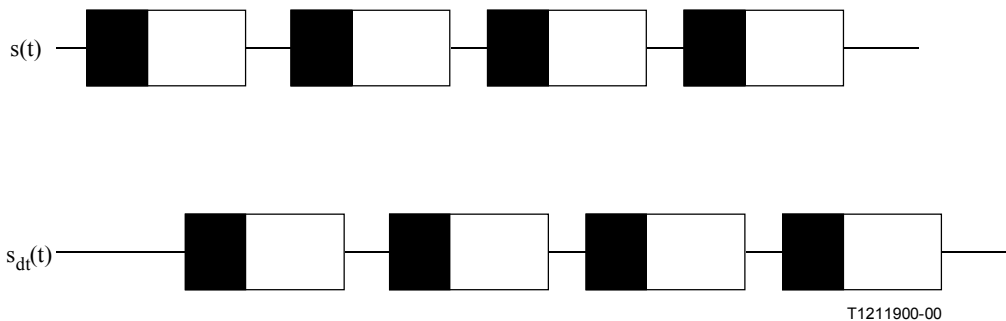


Figure 2/P.502 – Aperçu du signal de test en période de parole simultanée



$s(t)$ Signal dans un sens donné
 $s_{dt}(t)$ Signal de parole simultanée

Figure 3/P.502 – Découpage de la séquence de mesure complète avec vue détaillée du chevauchement des signaux dans le sens émission et dans le sens réception – arrangement de principe

La Figure 3 montre la construction du signal de façon plus détaillée. On constate que le chevauchement des séquences n'est que partiel. Le son voisé (noir) et la fin de la séquence de bruit pseudo-aléatoire (blanc) du canal opposé se chevauchent toujours. La séquence est construite de sorte que, pendant les pauses dans le sens réception, on puisse procéder à des mesures dans le sens émission et pendant les pauses dans le sens émission, on puisse procéder à des évaluations dans le sens réception. Cela est utile, par exemple, dans le cas de dispositifs analogiques pour lesquels il

n'est pas possible d'avoir un découplage suffisant entre le sens émission et le sens réception en raison de capacités d'effet local limitées.

De la même façon, on peut construire une séquence qui commence par un signal d'excitation de niveau élevé dans le sens réception et un signal d'excitation de bas niveau dans le sens émission, au cas où on doit évaluer différents points de départ de niveau.

De manière générale, il est à noter que la construction de signal telle qu'elle est illustrée sur les Figures 2 et 3 correspond à un certain exemple de relation temporelle entre le sens émission et le sens réception. Bien entendu, on peut utiliser d'autres intervalles temporels (par exemple des pauses plus longues, des séquences de bruit pseudo-aléatoires plus longues, différents types de signaux CSS), suivant les spécifications de la mesure à réaliser.

5.3.1.2 Procédures d'évaluation

En ce qui concerne les évaluations en période de parole simultanée, la séquence offre beaucoup de capacités. Avant d'évaluer un dispositif, il faut compenser le retard entre le signal d'excitation et le signal mesuré. Dans une deuxième étape, on extrait les signaux mesurés dans les deux sens et on les rapporte au signal d'excitation. La Figure 4 indique le principe de cette méthode lors de l'évaluation des paramètres dans le domaine fréquentiel (sur la base, par exemple, de la transformation de Fourier).

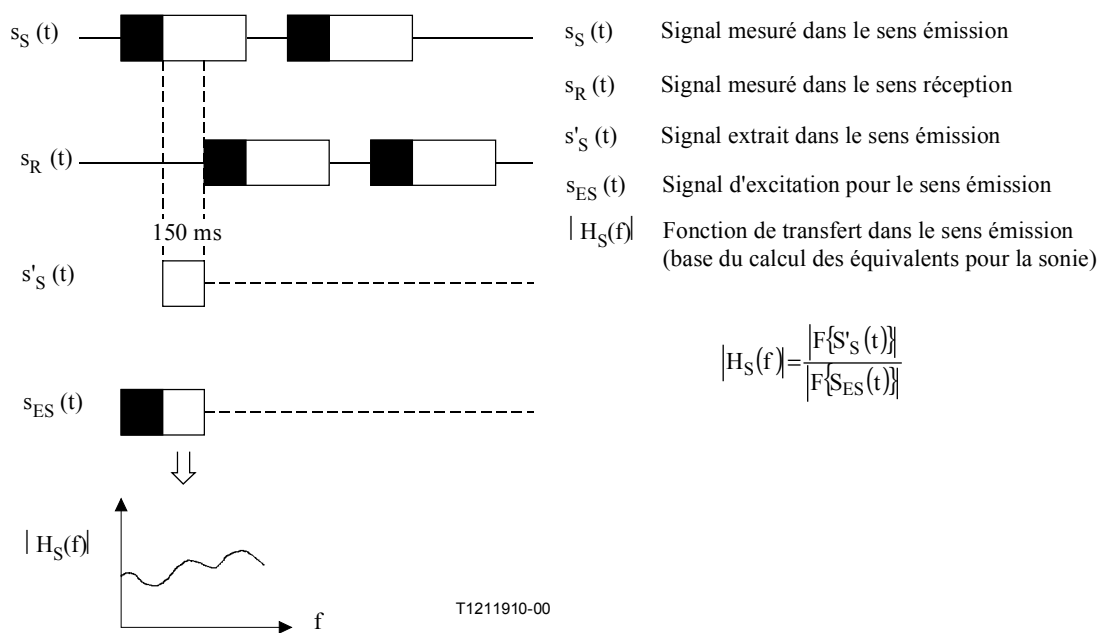


Figure 4/P.502 – Principe de l'extraction de signal et de la détermination des caractéristiques de transfert

NOTE – Il faut toujours s'assurer que l'analyse conduit à une évaluation valide des réponses en fréquence, des équivalents pour la sonie, etc. Les algorithmes de codage utilisés peuvent conduire à une mauvaise évaluation des fonctions de transfert, etc., et peuvent nécessiter des signaux de mesure ou des procédures d'analyse particuliers.

Comme le signal de mesure est un bruit pseudo-aléatoire de la séquence CSS, il est possible, à partir de ce signal de mesure de 150 ms, de calculer les paramètres suivants:

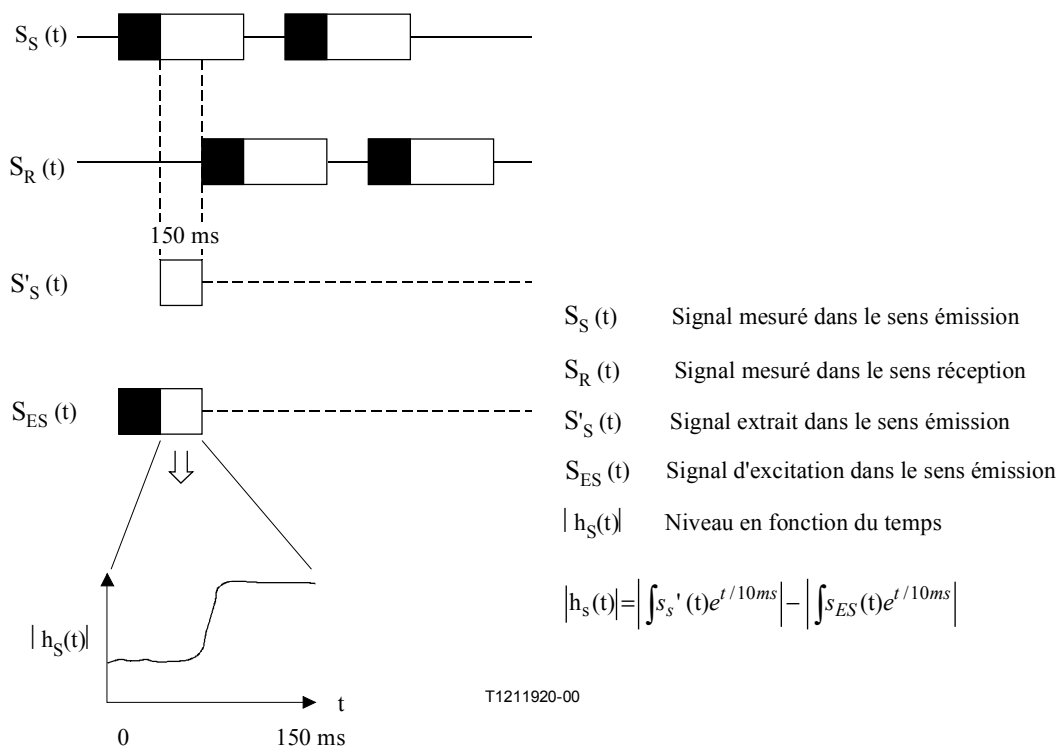
- réponses en fréquence;
- équivalent pour la sonie;

- atténuation à court terme (dans le cas de dispositifs de commutation de niveau);
- atténuation à long terme (lors de mesures à différents instants de la séquence).

Les temps de commutation peuvent être évalués directement dans le domaine temporel en raison des caractéristiques du signal définies de façon exacte:

- temps d'établissement (en raison du chevauchement des signaux uniquement si ce temps est supérieur à 50 ms);
- temps de commutation.

Pour évaluer les temps de commutation, on détermine le niveau en fonction du temps avec une constante de temps brève adéquate, généralement au moins 10 fois plus brève que le temps de commutation du système testé à évaluer. En surveillant le signal de sortie pendant les périodes où un seul signal est présent, on peut évaluer avec une grande précision la commutation ou les variations de niveau. Un exemple général de la procédure est donné sur la Figure 5. Dans cet exemple, le signal de sortie est rapporté au signal d'entrée, mais cela n'est pas toujours nécessaire car on connaît le niveau attendu pendant les périodes d'évaluation d'après le niveau de la partie pseudo-aléatoire du signal CSS.



NOTE – La constante de temps de 10 ms n'est qu'un exemple, on peut utiliser une constante de temps différente suivant les spécifications de l'analyse.

Figure 5/P.502 – Principe de l'extraction de signal et de la détermination de constantes de temps (l'exemple montre le temps de commutation en période de parole simultanée)

Si une évaluation du niveau en fonction de la fréquence est nécessaire, on peut utiliser la transformée de Fourier ou la transformée en ondelette du signal de sortie mesuré. On peut alors évaluer la commutation ou la variation de niveau en fonction de la fréquence. Il faut veiller à ce que le fenêtrage et la longueur de la fenêtre d'analyse conviennent. La longueur de la fenêtre d'analyse doit

être plus courte que l'intervalle de temps disponible pour l'analyse (en raison de la construction du signal et du chevauchement).

5.3.2 Test en période de parole simultanée, avec séquences combinées parallèles

Les tests en période de parole simultanée imposent souvent des contraintes conflictuelles concernant le type de signal de test à utiliser. Contrairement aux séquences combinées séquentielles décrites au 5.3.1, les séquences combinées parallèles permettent de procéder à des évaluations en période de parole simultanée. En règle générale, ces signaux doivent être orthogonaux ou doivent pouvoir être extraits au moyen de filtres à partir du signal devant être transmis au départ. La méthode de test est décrite ci-dessous.

En ce qui concerne les évaluations de l'affaiblissement d'adaptation pour l'écho, le signal de parole simultanée présenté à l'annuleur au point R_{in} (ou au simulateur de bouche) doit être aussi semblable que possible au signal de conditionnement. Les annuleurs verrouillent généralement l'adaptation en période de parole simultanée. Par exemple, si le signal de parole simultanée au point R_{in} était différent du signal de conditionnement, l'écho résiduel serait généralement anormalement élevé. Cette contrainte indique que le signal de parole simultanée au point R_{in} doit être identique au signal de conditionnement au point R_{in} pour les mesures de l'affaiblissement d'adaptation pour l'écho.

Malheureusement, l'utilisation de fichiers de signaux vocaux seuls n'est pas acceptable en période de parole simultanée. Les parties corrélées entre les deux "locuteurs" rendraient non valides certains résultats de test: certaines parties des signaux vocaux de l'un des locuteurs risquent de ressembler à l'écho de l'autre locuteur si les parties sont corrélées. Un autre problème est qu'il faut pouvoir détecter avec une grande précision le début de la parole simultanée pour les évaluations de l'atténuation et de l'écrêtage. Cela serait très difficile à définir sur plusieurs tests utilisant différents fichiers de signaux vocaux, mais est très facile avec des tonalités.

5.3.2.1 Construction du signal – Méthode A

Pour résoudre ces problèmes, les deux types de signaux sont utilisés: des signaux vocaux conformes au signal de conditionnement et des tonalités pour définir avec précision le début de la parole simultanée. La façon dont ils sont utilisés dépend du test. Les signaux vocaux utilisés pendant le conditionnement sont maintenus en période de parole simultanée, selon les besoins. Une tonalité sinusoïdale est mélangée aux signaux vocaux ou injectée telle quelle afin de fournir une référence facilement mesurable pour les évaluations de l'atténuation ou de constituer un début de parole simultanée facilement définissable pour les évaluations de l'écrêtage. Si on utilise des filtres coupe-bande ou passe-bande au point S_{out} (ou à la sortie réception) à la fréquence de la tonalité, on peut surveiller uniquement la tonalité ou uniquement les signaux vocaux.

Lorsque la tonalité est mélangée avec les signaux vocaux, la puissance de la tonalité doit être représentative de la puissance moyenne à long terme des signaux vocaux à sa fréquence, de manière à ne pas imposer à l'annuleur des déviations brutes d'énergie spectrale par rapport à celles du signal de conditionnement. La Recommandation UIT-T P.50 spécifie une relation spectrale moyenne (valeurs d'un tiers d'octave utilisées). Les tonalités données dans le Tableau 1 sont recommandées. Leur puissance est définie comme étant le nombre de dB au-dessous de l'énergie vocale active moyenne dans le fichier de signaux vocaux, lorsqu'elle est mesurée conformément à la Recommandation UIT-T P.56.

Tableau 1/P.502

Fréquence de la tonalité	Niveau relatif de la tonalité (dB) au-dessous du niveau vocal nominal
500	9
1 000	14
1 750	18
2 500	22

5.3.2.2 Evaluation de l'atténuation en période de parole simultanée au moyen de la méthode A

5.3.2.2.1 Atténuation des signaux vocaux dans le sens émission en période de parole simultanée (A_{sdt})

L'exemple montré sur la Figure 6 permet de déterminer l'atténuation dans le sens émission en période de parole simultanée. On peut facilement étendre le concept au sens réception en inversant les signaux et en surveillant la sortie réception.

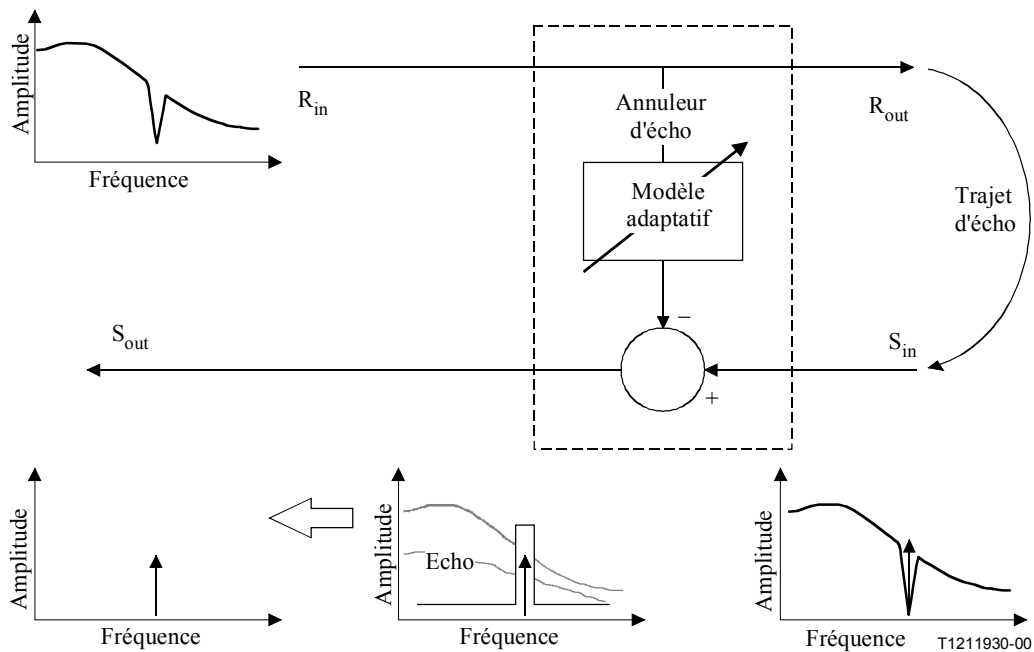


Figure 6/P.502 – Principe du test d'atténuation en période de parole simultanée au moyen de la méthode A

La méthode est expliquée ci-dessous.

Dans le cas où des annuleurs d'écho sont présents, l'objet testé est réinitialisé (si possible) et conditionné comme décrit dans l'Annexe C. Le "locuteur actif avant la parole simultanée" correspond au simulateur de bouche dans le cas de téléphones mains-libres au point S_{in} . Le "locuteur déclenchant la parole simultanée" correspond au point R_{in} .

Il est à noter que les signaux au niveau des simulateurs de bouche S_{in} et R_{in} sont représentés comme ayant subi un filtrage coupe-bande à la fréquence de la tonalité. Ce filtre coupe-bande est absent pendant toute la période de conditionnement mais il est présent juste avant la parole simultanée et

pendant le reste de la mesure. L'idée est de mélanger une tonalité au niveau du simulateur de bouche juste avant la parole simultanée (encore en mode conversation simple), de surveiller sa puissance efficace, de faire passer le point R_{in} en mode parole simultanée et de continuer à surveiller le niveau de la tonalité.

L'atténuation en période de parole simultanée est la différence entre le niveau de tonalité avant la situation de parole simultanée et celui au cours de cette situation. On distingue la tonalité en appliquant un filtre passe-bande à la fréquence de la tonalité au point S_{out} . En poursuivant la mesure en période de parole simultanée, on peut déterminer les caractéristiques de commutation, y compris la vitesse d'insertion et la profondeur. On peut aussi déterminer la vitesse de suppression de l'atténuation en diminuant à nouveau le gabarit d'activité du "locuteur déclenchant la parole simultanée" une fois que la profondeur d'atténuation s'est stabilisée.

On va maintenant décrire les caractéristiques du filtre coupe-bande. Ce filtre est nécessaire au point R_{in} pour faire en sorte que l'écho à la fréquence de la tonalité n'ait pas d'incidence sur la mesure de la tonalité. Il doit présenter une atténuation suffisante pour garantir que les signaux vocaux à la fréquence de la tonalité sont convenablement réprimés de manière à ne pas avoir d'incidence sur le niveau de la tonalité au niveau du simulateur de bouche. Le filtre coupe-bande doit être suffisamment étroit pour minimiser l'incidence sur les fréquences environnantes de sorte que les signaux ne soient pas trop différents des signaux de conditionnement. Des exemples de types de filtre figurent à l'Annexe B.

Le filtre passe-bande a des contraintes semblables. Pris avec le filtre coupe-bande, il doit avoir une atténuation hors bande suffisante pour faire en sorte que les signaux vocaux ou l'écho n'aient pas d'incidence sur le niveau de la tonalité. Il doit aussi avoir une réponse impulsionnelle suffisamment brève pour que l'incidence sur le domaine temporel soit minimale, car la vitesse d'insertion de l'atténuation est aussi mesurée.

La réponse impulsionnelle du filtre coupe-bande n'a pas d'incidence sur la mesure car la tonalité mélangée au niveau du simulateur de bouche aura un niveau suffisant pour couvrir toute oscillation résiduelle du filtre coupe-bande.

La puissance efficace de la tonalité doit être mesurée avec une fenêtre glissante de 8 ms aux fins du lissage. La fenêtre est déplacée par pas de 4 ms pour avoir un chevauchement de 4 ms entre points adjacents afin de lisser les résultats.

La programmation de la mesure doit être très précise afin de tenir compte de toute oscillation du filtre passe-bande. La tonalité doit être injectée au bout de 60 secondes, moins le temps d'oscillation du filtre passe-bande (supposé être <20 ms), moins 8 ms. La mesure de référence de la tonalité en situation de conversation simple commence à 60 secondes moins 8 ms, après la fin de l'oscillation du filtre. Le retard fixé dans le sens de la mesure conduit à une mesure partielle pendant la fin de l'oscillation du filtre passe-bande. Tant que le retard fixé est bref (provisoirement <5 ms), les effets d'oscillation rencontrés seront légers et ne devraient pas avoir d'incidence sur la mesure.

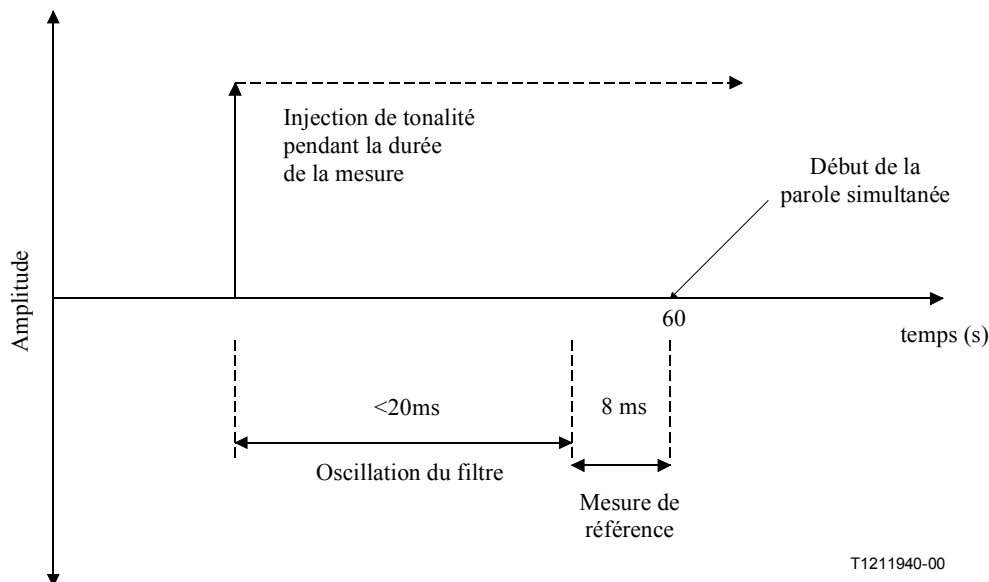


Figure 7/P.502

Lorsque l'annuleur AEC utilise des techniques de subdivision en sous-bandes ou lorsque la technique est inconnue, il est recommandé de répéter le test pour chaque fréquence indiquée ci-dessus. Dans de nombreux cas, les résultats d'évaluation de l'atténuation à n'importe quelle fréquence peuvent ne pas être représentatifs de la qualité subjective. Comme la voix a une énergie spectrale dominante dans la partie basses fréquences, l'affaiblissement commuté devrait être plus audible dans cette partie. Si la profondeur d'atténuation est indépendante de la fréquence, il est recommandé d'utiliser des fréquences de test élevées car les filtres nécessaires auront moins d'incidence sur les niveaux vocaux d'ensemble.

La méthode donnée permet de mesurer l'atténuation en fonction du temps après le passage au mode parole simultanée pour une fréquence donnée. Le résultat de cette méthode peut dépendre grandement de la nature exacte du signal vocal utilisé, notamment lorsque la parole simultanée commence. Il peut aussi dépendre de la fréquence de la tonalité de mesure, qui est une sinusoïde imbriquée dans les signaux vocaux réels créant la parole simultanée.

Les étapes à suivre sont décrites de façon détaillée ci-dessous.

L'annuleur est conditionné comme décrit dans l'Annexe C, avec le gabarit d'activité du "locuteur actif juste avant le début de la parole simultanée" appliqué au point S_{in} (ou au point MRP lorsque la mesure porte sur des terminaux). Le gabarit du "locuteur déclenchant la parole simultanée" est appliqué au point R_{in} . On procède au test décrit avant.

Si on utilise la fenêtre glissante de 8 ms pour moyenniser le signal sinusoïdal mesuré au point S_{out} , la valeur A_{SDT} , qui dépend du temps, est la différence (en dB) entre la première moyenne sur 8 ms avant la parole simultanée et chaque moyenne sur 8 ms après la parole simultanée.

5.3.2.2.2 Atténuation des signaux vocaux dans le sens réception en période de parole simultanée (A_{rdt})

NOTE – Dans le cas où la mesure du signal de sortie dans le sens réception s'applique à des téléphones mains-libres, le microphone de mesure doit être placé le plus près possible du haut-parleur pour assurer la discrimination dans le domaine acoustique.

On réalise le test comme décrit avant pour A_{sdt} , en remplaçant réception par émission et inversement. Les signaux de réception et d'émission sont donc échangés et les résultats sont surveillés à la sortie réception.

5.3.2.3 Evaluation de l'affaiblissement d'adaptation pour l'écho en période de parole simultanée au moyen de la méthode A

L'exemple montré sur la Figure 8 permet de déterminer l'affaiblissement d'adaptation pour l'écho vu du réseau en direction du terminal. On peut facilement étendre le concept à l'affaiblissement sur le trajet d'écho du locuteur en inversant les signaux et en surveillant la sortie réception.

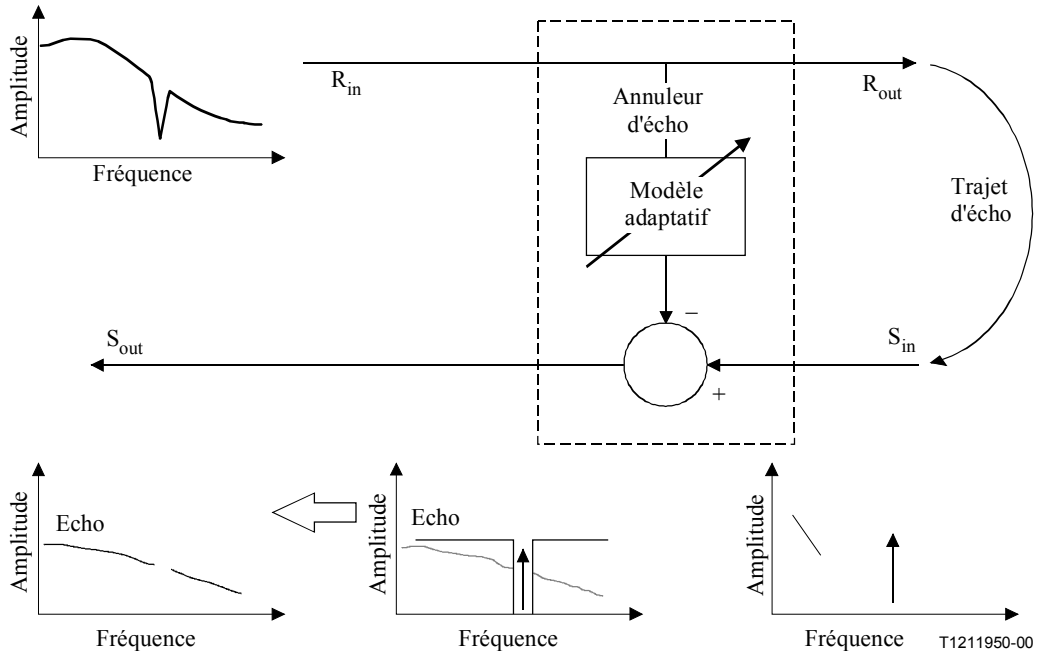


Figure 8/P.502 – Principe de l'évaluation de l'affaiblissement d'adaptation pour l'écho au moyen de la méthode A

Le poste est réinitialisé et conditionné comme décrit dans l'Annexe C. Le "locuteur actif avant la parole simultanée" correspond au point R_{in} . Le "locuteur déclenchant la parole simultanée" correspond au simulateur de bouche ou au point S_{in} .

Le filtre coupe-bande est appliqué au point R_{in} comme décrit au 5.3.2.2. On passe au mode parole simultanée en appliquant une tonalité (uniquement) au niveau du simulateur de bouche. La tonalité est injectée pendant 20 secondes. On obtient l'affaiblissement d'adaptation pour l'écho d'abord en supprimant la tonalité grâce au filtre coupe-bande puis en mesurant l'écho résiduel au moyen de calculs de l'affaiblissement ERL ou de la méthode relative à l'affaiblissement ERL_t décrite dans l'Annexe A. L'utilisation d'une tonalité permet de connaître l'instant précis où on passe au mode parole simultanée.

A la fin de la parole simultanée, on peut poursuivre la mesure de l'affaiblissement d'adaptation pour l'écho pendant 10 secondes pour mesurer la récupération après la parole simultanée. Après cela, il faut introduire une seconde de silence. De cette façon, on peut mesurer le bruit dans le trajet d'écho. Si on peut supposer que le bruit et l'écho ne sont pas corrélés et que le bruit est stationnaire, le bruit mesuré dans la dernière seconde peut être soustrait de l'écho plus le bruit en période de parole simultanée pour aboutir à l'écho en période de parole simultanée.

La programmation de la mesure doit être très précise, compte tenu du retard du trajet d'écho. Ce retard permet d'aligner correctement la source et l'écho.

Dans les détails, la procédure d'évaluation est la suivante:

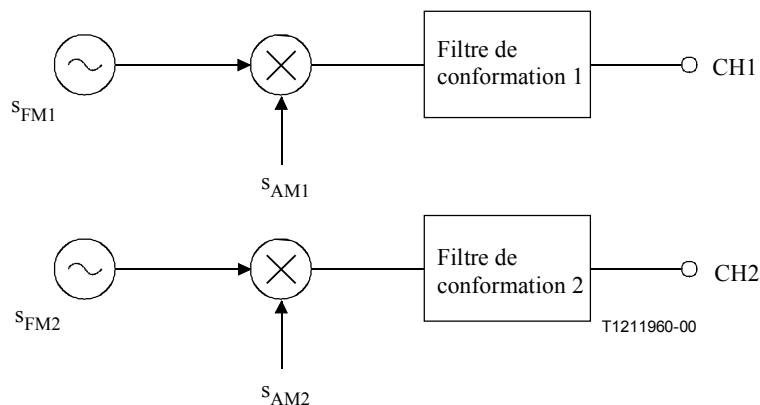
Affaiblissement d'adaptation pour l'écho – parole simultanée (ERL_{dt}), avec pondération temporelle – parole simultanée (ERL_{tdt})

Si des annuleurs d'écho sont présents, toutes leurs unités fonctionnelles sont d'abord réinitialisées puis activées. Pour l'essai de type boîte noire, le système est alimenté sans qu'aucun stimulus acoustique ne soit présent à aucune interface. L'annuleur est conditionné et le gabarit d'activité du "locuteur actif juste avant le début de la parole simultanée" est appliqué au point R_{in}. Le gabarit du "locuteur déclenchant la parole simultanée" est appliqué au point S_{in} (par exemple le simulateur de bouche). Les essais sont réalisés comme décrit dans l'Annexe A.

On enregistre les signaux électriques aux points R_{in} et S_{out} pendant les 20 secondes d'application de la tonalité. On aligne ensuite les enregistrements aux points R_{in} et S_{out} dans le temps en tenant compte du retard dans le système entre R_{in} et S_{out}. On calcule l'affaiblissement ERL_{dt} au moyen du calcul traditionnel de l'affaiblissement d'écho conformément à la Recommandation UIT-T G.122 ou l'affaiblissement ERL_{tdt}. La valeur ERL_{tdt}, qui dépend du temps, est la différence (en dB) entre le niveau du signal au point R_{in} et celui au point S_{out} calculée comme indiqué dans l'Annexe A, pendant les 20 secondes après le passage au mode parole simultanée.

5.3.2.4 Construction du signal – Méthode B

La méthode B utilise des séquences orthogonales générées par un ensemble d'ondes sinusoïdales modulées semblables à la voix humaine, conformées spectralement. La Figure 9 représente le principe général de construction, qui est détaillé dans la Recommandation UIT-T P.501.



$$s_{FM1,2}(t) = \sum A_{FM1,2} * \cos(2\pi t n * F_{01,2}); \quad n = 1, 2, \dots$$

$$s_{AM1,2}(t) = \sum A_{AM1,2} * \cos(2\pi t F_{AM1,2});$$

Figure 9/P.502 – Génération de signal de test avec deux canaux pour des évaluations en période de parole simultanée fondées sur des signaux AM-FM

Les valeurs types sont données dans le Tableau 2.

Tableau 2/P.502

	f_{AM}	f_{FM}	F_0	Filtre de conformation
Canal 1 (CH 1)	$f_{AM1} = 3 \text{ Hz}$	$f_{FM1} = 5 \text{ Hz}$	$F_{01} = 270 \text{ Hz}$	LP, 5 dB/oct.
Canal 2 (CH 2)	$f_{AM2} = 3 \text{ Hz}$	$f_{FM2} = 5 \text{ Hz}$	$F_{02} = 290 \text{ Hz}$	LP, 5 dB/oct.

Pour de plus amples détails, voir la Recommandation UIT-T P.501.

5.3.2.5 Evaluation de l'atténuation en période de parole simultanée au moyen de la méthode B

5.3.2.5.1 Atténuation des signaux vocaux dans le sens émission en période de parole simultanée (A_{sdt})

L'exemple montré sur la Figure 10 permet de déterminer l'atténuation dans le sens émission en période de parole simultanée. Comme pour la méthode A, on peut facilement étendre le concept au sens réception en inversant les signaux et en surveillant la sortie réception.

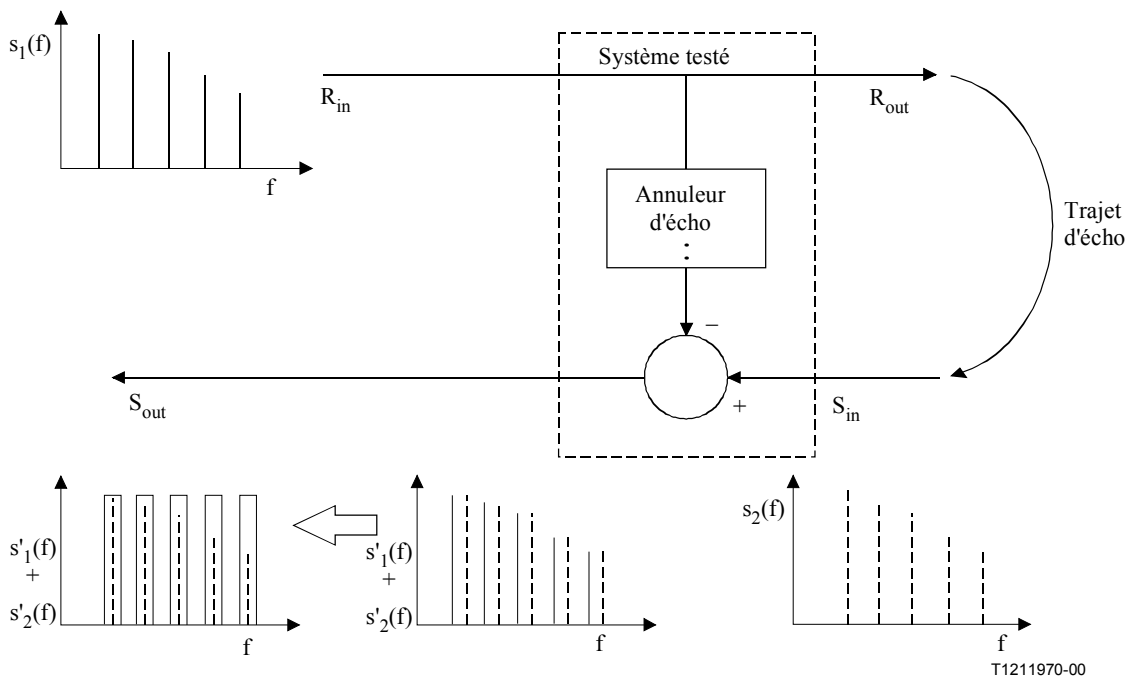


Figure 10/P.502 – Principe de l'évaluation de l'atténuation en période de parole simultanée au moyen de la méthode B

La méthode est expliquée ci-dessous.

Dans le cas où des annuleurs d'écho sont présents, l'objet testé est réinitialisé (si possible) et conditionné comme décrit dans l'Annexe A. Le "locuteur actif avant la parole simultanée" est le simulateur de bouche dans le cas de téléphones mains-libres ou le point S_{in} . Le "locuteur déclenchant la parole simultanée" est le point R_{in} .

Le signal de parole simultanée est analysé pendant la période de parole simultanée selon le principe d'analyse suivant.

Pour que le signal d'écho puisse être extrait de la parole simultanée, il faut soit un filtre spécifique soit un post-traitement spécifique de l'analyse découlant d'une FFT, car le spectre du signal ainsi que du signal de parole simultanée est une sorte de spectre de filtre en peigne auquel une modulation particulière est appliquée. Il faut tenir compte de la fréquence milieu f_{mid} de n'importe quelle composante de fréquence, de la fréquence de modulation correspondante f_{mod} ainsi que des formes de filtre ou de la fonction de fenêtrage de la transformation de Fourier. Dans la méthode avec filtre, la bande de chaque filtre doit être construite de la façon suivante:

$$f_u = f_{mid} - f_{mod} \text{ (fm)}$$

$$f_o = f_{mid} + f_{mod} \text{ (fm)}$$

L'atténuation dans la bande coupée doit être supérieure d'au moins 10 dB au niveau minimal à mesurer dans la bande passante. Il en est de même pour les analyses découlant de transformations de Fourier du signal d'écho mesuré. Dans ce cas, il faut tenir compte de l'effet de "bavure" de fréquence de la fonction de fenêtrage. Pour pouvoir obtenir une séparation suffisante entre le signal d'écho et le signal de parole simultanée dans le domaine des basses fréquences, il convient de choisir une longueur d'au moins 8 k pour la FFT (fréquence d'échantillonnage de 44,1 ou 48 kHz), ce qui correspond à une fenêtre temporelle d'environ 170 ms.

La Recommandation UIT-T P.501 donne un ensemble de fréquences d'excitation types pour le signal de conversation simple et pour le signal de parole simultanée.

Le principe de l'analyse est illustré sur la Figure 11.

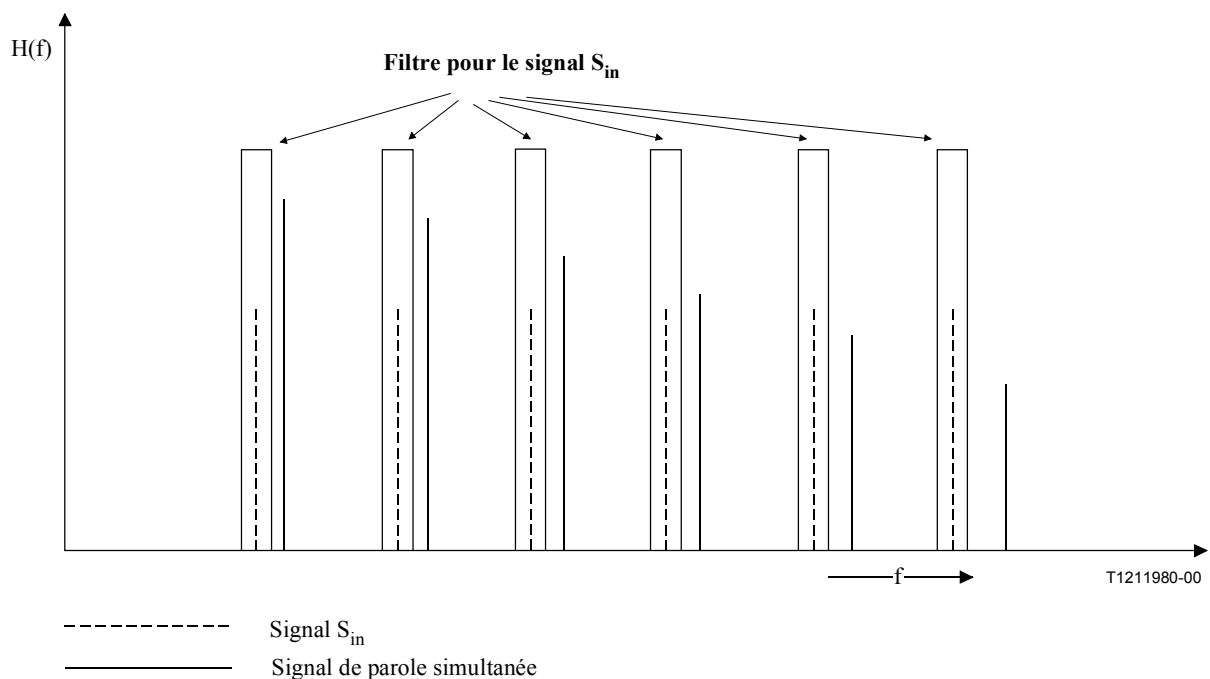


Figure 11/P.502 – Extraction des composantes du signal S_{in} (présentation schématique)

L'atténuation en période de parole simultanée est la différence entre le niveau de tonalité avant la situation de parole simultanée et celui au cours de cette situation. On distingue la séquence de sons voisés en appliquant la procédure avec filtre ou la procédure FFT comme décrit plus haut. En poursuivant la mesure en période de parole simultanée, on peut déterminer les caractéristiques de commutation, y compris la vitesse d'insertion et la profondeur, dans le domaine temporel ou fréquentiel suivant la méthode d'analyse choisie. On peut aussi déterminer la vitesse de suppression

de l'atténuation en diminuant à nouveau le gabarit d'activité du "locuteur déclenchant la parole simultanée" une fois que la profondeur d'atténuation s'est stabilisée.

La programmation de la mesure doit être très précise, compte tenu du retard du trajet d'écho. Ce retard permet d'aligner correctement la source et l'écho.

5.3.2.5.2 Atténuation des signaux vocaux dans le sens réception en période de parole simultanée (A_{rdt})

NOTE – Dans le cas où la mesure du signal de sortie dans le sens réception s'applique à des téléphones mains-libres, le microphone de mesure doit être placé le plus près possible du haut-parleur pour assurer la discrimination dans le domaine acoustique.

On réalise le test comme décrit avant pour A_{sdt} , en remplaçant réception par émission et inversement. Les signaux de réception et d'émission sont donc échangés et les résultats sont surveillés à la sortie réception.

Pour une description détaillée des différentes étapes (1 à 4) de l'évaluation de "l'atténuation des signaux vocaux dans le sens émission en période de parole simultanée (A_{sdt})" et de "l'atténuation des signaux vocaux dans le sens réception en période de parole simultanée (A_{rdt})" (voir 5.3.2.2).

5.3.2.6 Evaluation de l'affaiblissement d'adaptation pour l'écho en période de parole simultanée au moyen de la méthode B

L'exemple montré sur la Figure 12 permet de déterminer l'affaiblissement d'adaptation pour l'écho vu du réseau en direction du terminal. On peut facilement étendre le concept à l'affaiblissement sur le trajet d'écho du locuteur en inversant les signaux et en surveillant la sortie réception.

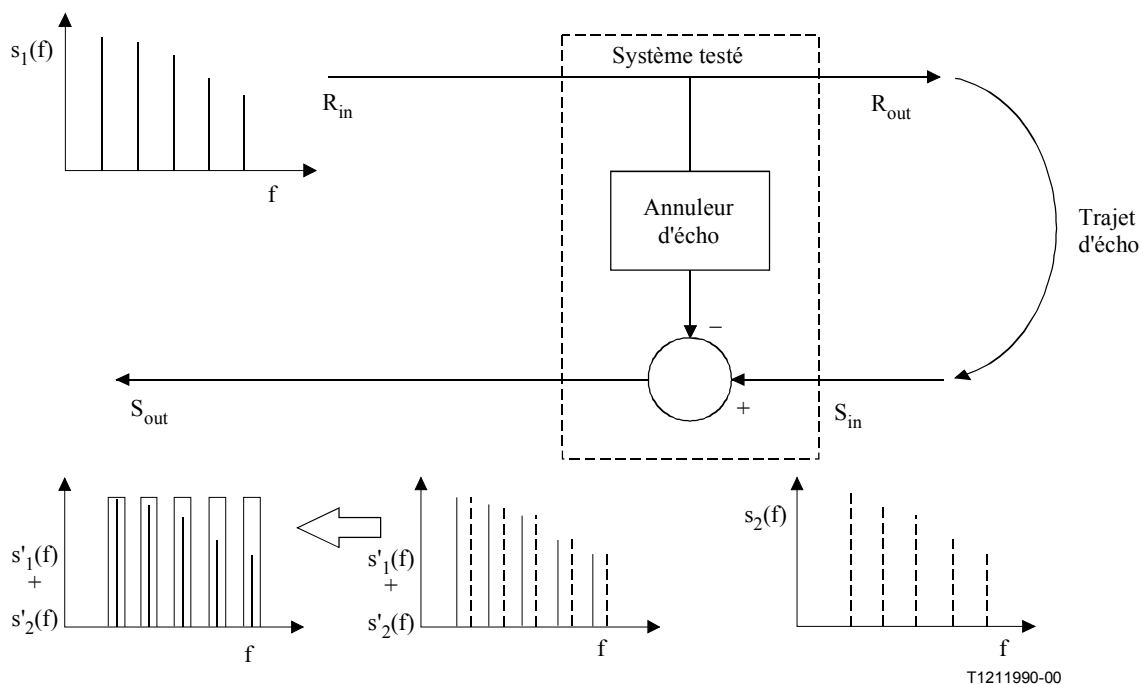


Figure 12/P.502 – Principe de l'évaluation de l'affaiblissement d'adaptation pour l'écho en période de parole simultanée au moyen de la méthode B

La méthode est expliquée ci-dessous.

Dans le cas où des annuleurs d'écho sont présents, le poste est réinitialisé et conditionné comme décrit dans l'Annexe C. Le "locuteur actif avant la parole simultanée" correspond au point R_{in} . Le "locuteur déclenchant la parole simultanée" correspond au simulateur de bouche ou au point S_{in} .

Le signal de parole simultanée est analysé pendant la période de parole simultanée selon le principe d'analyse décrit dans le sous-paragraphe précédent. Au lieu du signal de parole simultanée, on analyse le signal d'écho en appliquant le filtre approprié (voir Figure 13).

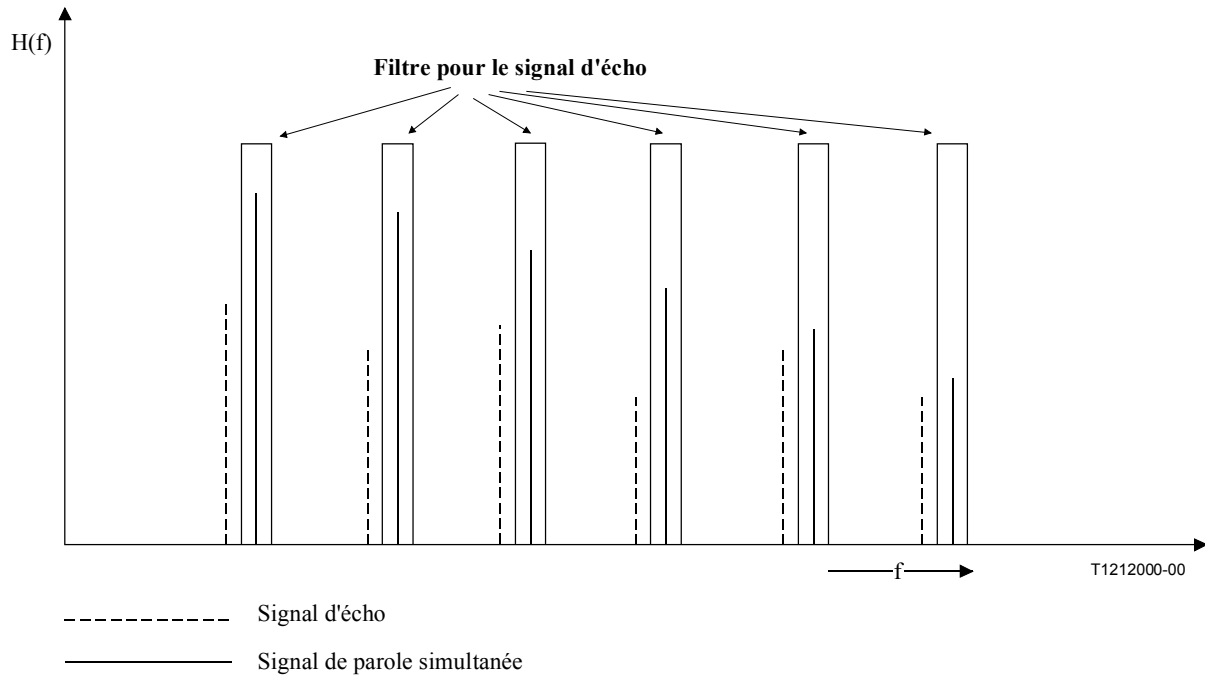


Figure 13/P.502 – Extraction des composantes d'écho du signal de parole simultanée (présentation schématique)

A la fin de la parole simultanée, on peut poursuivre la mesure de l'affaiblissement d'adaptation pour l'écho pendant 10 secondes pour mesurer la récupération après la parole simultanée. Après cela, il faut introduire une seconde de silence. De cette façon, on peut mesurer le bruit dans le trajet d'écho. Si on peut supposer que le bruit et l'écho ne sont pas corrélés et que le bruit est stationnaire, le bruit mesuré dans la dernière seconde peut être soustrait de l'écho plus le bruit en période de parole simultanée pour aboutir à l'écho en période de parole simultanée.

La programmation de la mesure doit être très précise, compte tenu du retard du trajet d'écho. Ce retard permet d'aligner correctement la source et l'écho.

Pour une description détaillée des différentes étapes (1 à 4) de l'évaluation de "l'atténuation des signaux vocaux dans le sens émission en période de parole simultanée (A_{sdt})" et de "l'atténuation des signaux vocaux dans le sens réception en période de parole simultanée (A_{rdt})" (voir 5.3.2.3).

6 Caractéristiques de compression-extension et de commande AGC

On peut utiliser la compression-extension ou la commande AGC pour éviter le débordement des systèmes, pour compenser des niveaux vocaux variables ou pour égaliser les niveaux vocaux dans le réseau. Dans tous les cas, le but est d'améliorer la "qualité" vocale ou les signaux d'entrée de dispositifs opérant sur la base de signaux vocaux (annuleurs d'écho, détecteurs de parole, etc.).

6.1 Dégradation de la qualité vocale telle qu'elle est perçue subjectivement

Les dispositifs de compression-extension ou de commande AGC et, d'une manière générale, tout dispositif qui introduit une amplification de signal vocal variable dans le temps conduisent à des fluctuations du niveau vocal, qui peuvent provoquer une dégradation de la qualité vocale. L'influence telle qu'elle est perçue subjectivement dépend généralement de la plage d'atténuation, des constantes de temps et des caractéristiques de commande de ces dispositifs.

6.2 Paramètres objectifs associés

Les paramètres objectifs associés sont les suivants:

- plage de commande du compresseur-extenseur;
- plage d'atténuation (plage des ajustements de niveau);
- constantes de temps.

6.3 Méthodes d'analyse

En général, il faut utiliser différents types de signaux de test pour déterminer la qualité de fonctionnement de dispositifs de compression-extension ou de commande AGC. Outre des signaux de test artificiels, il convient d'utiliser des signaux vocaux ou des signaux presque semblables à la voix humaine tels que la voix artificielle (Recommandation UIT-T P.50) ou des signaux semblables à la voix humaine (signal MSMP) tels que décrits dans la Recommandation UIT-T P.501.

Les Figures 14 et 15 représentent des signaux de test, générés par une répétition périodique de sons voisés. On peut utiliser ces signaux pour mesurer des ajustements de niveau pour des systèmes qui ont la même réaction dans le cas de la répétition périodique d'un son voisé et dans le cas de signaux vocaux réels. On peut en outre utiliser une voix artificielle.

Le signal représenté sur la Figure 14 est un signal d'entrée dont le niveau croît (décroît) de façon monotone, tandis que le niveau du signal représenté sur la Figure 15 est ajusté par intervalles.

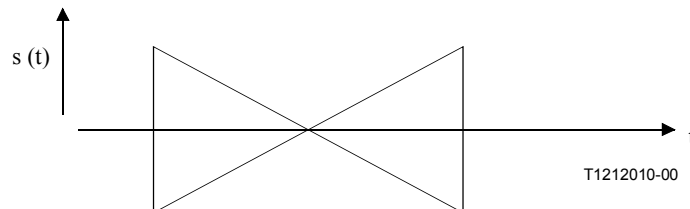


Figure 14/P.502 – Structure d'un signal de test pour déterminer des ajustements de niveau (variation constante du niveau d'entrée)

Les paramètres suggérés pour le signal représenté sur la Figure 14 sont donnés dans le Tableau 3.

Tableau 3/P.502

	Génération du signal	Niveau maximal	Niveau minimal	Variation du niveau
Sens SND	Son voisé, répété périodiquement	-16 dBm (-3,0 dBPa pour les terminaux)	En dessous du bruit de fond	Linéaire
Sens RCV	Son voisé, répété périodiquement	-16 dBm	En dessous du bruit de fond	Linéaire

La durée totale du signal peut être choisie égale à 10 s.

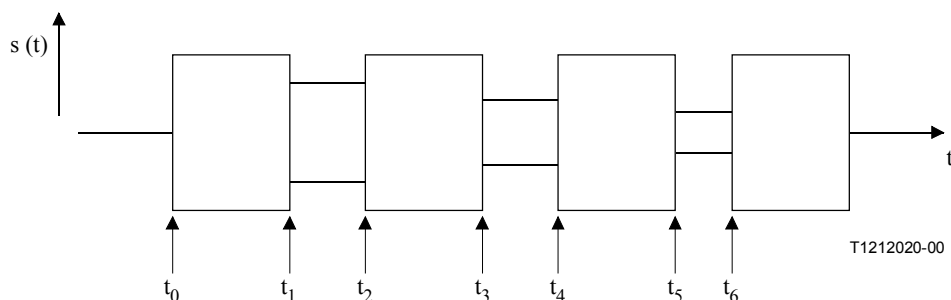


Figure 15/P.502 – Structure d'un signal de test pour déterminer des ajustements de niveau

Les paramètres suggérés pour le signal représenté sur la Figure 15 sont donnés dans le Tableau 4.

Tableau 4/P.502

	Génération du signal	Niveau du signal pendant ($t_1 - t_0$) ($t_3 - t_2$) ($t_5 - t_4$)	Niveau du signal pendant ($t_2 - t_1$)	Niveau du signal pendant ($t_4 - t_3$)	Niveau du signal pendant ($t_6 - t_5$)
Sens SND	Son voisé, répété périodiquement	-16 dBm (-3,0 dBPa pour les terminaux)	-21 dBm (-8,0 dBPa pour les terminaux)	-26 dBm (-13,0 dBPa pour les terminaux)	-31 dBm (-18,0 dBPa pour les terminaux)
Sens RCV	Son voisé, répété périodiquement	-16 dBm	-21 dBm	-26 dBm	-31 dBm

La durée de chacune des périodes du signal peut être choisie égale à 2,5 s.

Le signal de la Figure 14 convient pour déterminer la:

- **plage des ajustements de niveau en fonction du niveau du signal d'entrée**

On applique le signal décrit ci-dessus (Figure 14) et on analyse le niveau du signal de sortie en fonction du temps, en rapportant le signal de sortie mesuré au signal d'excitation aligné dans le temps. Il faut choisir les constantes de temps dans l'intervalle 10-125 ms. Il faut veiller à éviter que les résultats de mesure obtenus soient faussés en raison d'un signal de test qui ne serait pas semblable à la voix humaine.

En théorie, le résultat de l'analyse est un graphe plat en fonction du temps. En cas d'utilisation d'une compression-extension ou d'une commande AGC, le résultat ne doit pas dévier de plus de ± 3 dB du résultat mesuré en moyenne. En cas d'utilisation d'une commande AGC, les constantes de temps doivent être plutôt élevées (>100 ms). Si une compression-extension est détectée, les temps d'attaque doivent être plutôt brefs (10-50 ms); les temps de libération doivent, quant à eux, être suffisamment longs. En cas de détection d'une compression-extension ou d'une commande AGC de plus de ± 3 dB, une évaluation subjective est nécessaire.

Le signal de la Figure 15 convient spécialement pour déterminer la:

– **durée des ajustements de niveau**

Pour cette analyse, il faut utiliser le signal de la Figure 15. L'analyse est faite comme décrit ci-dessus. Pour cela, on analyse le niveau en fonction du temps, en rapportant le signal mesuré au signal d'excitation aligné dans le temps, et on affiche le résultat sous la forme d'un graphe en fonction de t. Pour cette analyse, il faut choisir une constante de temps comprise entre 5 et 10 ms afin d'obtenir une bonne résolution temporelle.

En théorie, il ne devrait pas y avoir de différences de niveau sensibles. Si une compression-extension ou une commande AGC est détectée, elle doit être dans la plage ± 3 dB. La durée mesurée doit donc être telle que décrite ci-dessus.

7 **Qualité de transmission du bruit de fond**

Lors de l'évaluation de la qualité de transmission du bruit de fond, l'auditeur considère le bruit de fond comme un signal. En général, des effets semblables à ceux observés lors de l'application de signaux vocaux peuvent influencer la qualité telle qu'elle est perçue subjectivement. Ce paramètre est de plus en plus important compte tenu de l'utilisation croissante de systèmes de télécommunications modernes dans des environnements bruyants.

7.1 **Dégradation de la qualité telle qu'elle est perçue subjectivement**

L'influence la plus typique est observée dans le sens émission pour l'auditeur situé à l'extrémité distante lorsque du bruit de fond est transmis. En général, la qualité perçue est influencée par:

- les fluctuations de niveau du bruit de fond;
- les interruptions de transmission de bruit;
- les modulations semblables à des artefacts produites par le traitement du signal.

Il est possible que l'influence soit différente:

- au repos;
- avec des signaux vocaux à l'extrémité distante;
- avec des signaux vocaux à l'extrémité proche.

Chaque situation doit être examinée séparément.

7.2 **Paramètres objectifs associés**

Pour les descriptions d'analyse qui suivent, le bruit de fond est considéré comme étant le signal de test. Les effets tels qu'ils sont perçus subjectivement peuvent être décrits par les paramètres suivants:

- plage d'atténuation;
- atténuation dans le sens SND;
- caractéristiques de commutation;
- niveau minimal d'activation dans le sens SND;

- réponse en fréquence;
- sensibilité de la détection du bruit de fond (niveau d'activation, niveau absolu, fluctuations de niveau).

En outre, la qualité est influencée par:

- les caractéristiques de processeurs NLP ou d'écrêteurs de centre conjointement avec des annuleurs d'écho;
- les caractéristiques de systèmes de réduction du bruit.

7.3 Méthodes d'analyse

En général, le bruit de fond simulé est un signal de bruit continu (avec spectre conformé) ou un signal plus sophistiqué, pour représenter des conditions réalistes (par exemple un brouhaha de voix dans un bureau). En pareils cas, il faut caractériser le bruit de fond par sa densité spectrale de puissance à long terme et son niveau moyen appliqué pendant la mesure.

Pour les tests suivants, le signal de bruit de fond est considéré comme étant le signal de mesure et non comme étant une composante perturbatrice. Les analyses sont donc fondées sur le signal de bruit. On peut évaluer la qualité de transmission du bruit de fond (à partir de l'extrémité proche dans le sens SND) au repos, avec des signaux vocaux à l'extrémité distante et avec des signaux vocaux à l'extrémité proche.

Dans tous ces cas, les paramètres importants sont les suivants:

- la sensibilité de la détection du bruit de fond en termes de niveau d'activation;
- le niveau absolu du signal de bruit transmis;
- les fluctuations de niveau du signal de bruit transmis.

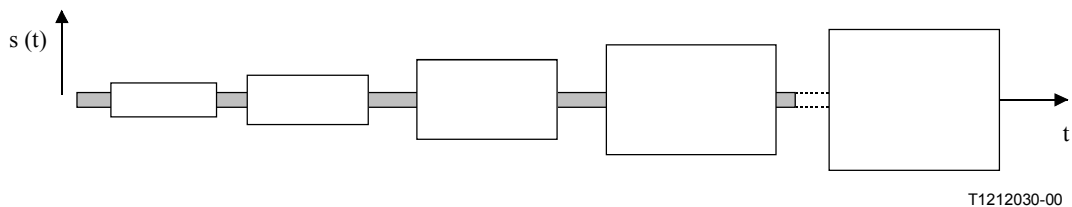
Comme les tests d'écoute pour l'évaluation des propriétés requises pour la transmission du bruit de fond sont en cours, la description détaillée des méthodes d'analyse est incomplète pour le moment.

Dans le **mode repos**, on peut par exemple procéder à des mesures relatives à la transmission du bruit de fond en appliquant un champ de bruit diffus possédant les caractéristiques de H_{oth} ou en utilisant un signal de bruit de fond approprié, par exemple un brouhaha de voix, un bruit de voiture ou d'autres bruits représentant des conditions particulières. On peut appliquer le niveau du signal pendant 20 secondes, en commençant par un niveau d'excitation de 50 dB_{SPL} (A) ou le niveau correspondant mesuré à l'interface électrique. On peut ensuite augmenter le niveau de 3 dB et l'appliquer à nouveau pendant 20 secondes. On construit ainsi une séquence de mesure contenant des blocs de 20 secondes de bruit de H_{oth} avec une augmentation de 3 dB d'un bloc au suivant. Lors de l'application de chaque bloc de 20 s, on ne doit pas détecter de variation du bruit de fond. Les valeurs exactes des constantes de temps ne sont pas encore disponibles. En cas de détection de fluctuations de niveau de plus de ± 3 dB par rapport aux conditions du régime permanent, il faut procéder à une évaluation subjective.

D'une manière générale, le signal de bruit de fond doit être audible tout le temps.

Plus le niveau du bruit de fond transmis est faible, mieux c'est; toutefois, il faut éviter les artefacts liés aux algorithmes de réduction du bruit. Des simulations de bruit de fond plus réalistes sont à l'étude.

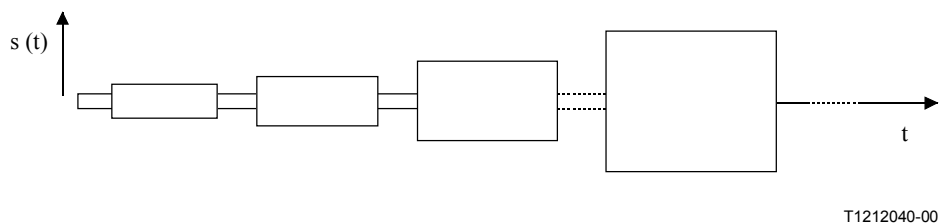
On peut utiliser la structure de signal suivante pour évaluer la qualité de transmission du bruit de fond dans le sens SND **avec signaux vocaux à l'extrémité distante**. La Figure 16 représente un signal de bruit continu appliqué à l'extrémité proche (sens SND, couleur grise) et des signaux vocaux simulés à l'extrémité distante dans le sens RCV (couleur blanche, on peut utiliser des rafales de signaux CSS). La mesure est réalisée dans le sens SND. Sur la Figure 16, le niveau des rafales de signaux CSS varie tandis que le bruit de fond simulé est appliqué avec un niveau constant.



NOTE – La ligne en pointillés indique que le signal de test est répété ou allongé pour obtenir la durée convenable pour la mesure.

Figure 16/P.502 – Exemple de structure de signal de test pour évaluer la qualité de transmission du bruit de fond dans le sens SND (avec signaux vocaux simulés à l'extrémité distante)

On peut utiliser une structure de signal semblable pour évaluer la qualité de transmission du bruit de fond **avec signaux vocaux à l'extrémité proche** (voir la Figure 17). Dans ce cas, le bruit de fond et les signaux vocaux simulés (on peut à nouveau utiliser des signaux CSS) sont appliqués et mesurés dans le même sens (à l'opposé de la Figure 16), par exemple dans le sens SND.



NOTE – La ligne en pointillés indique que le signal de test est répété ou allongé pour obtenir la durée convenable pour la mesure.

Figure 17/P.502 – Exemple de structure de signal de test pour évaluer la qualité de transmission du bruit de fond dans le sens SND (avec signaux vocaux simulés à l'extrémité proche)

Le signal de bruit de fond et les rafales de signaux CSS sont représentés avec la même couleur (blanche) pour indiquer qu'ils sont injectés dans le même sens.

8 Caractéristiques de commutation

La commutation peut influencer sur la qualité de transmission vocale dans diverses situations et conditions: en période de conversation simple, en période de parole simultanée (le locuteur à l'extrémité proche et celui à l'extrémité distante sont tous deux actifs). Quel que soit le cas, des syllabes peuvent être tronquées et des mots ou des phrases complets peuvent même manquer ou être interrompus.

8.1 Dégradation de la qualité vocale telle qu'elle est perçue subjectivement

La dégradation de la qualité vocale telle qu'elle est perçue subjectivement peut être décrite par:

- l'occurrence de coupures dans les signaux vocaux;
- des syllabes manquantes;
- des mots ou des phrases incomplets.

Les personnes désignent généralement la perturbation introduite comme étant des coupures dans les signaux vocaux. On observe souvent une interaction entre écho et commutation.

8.2 Paramètres objectifs associés

Les paramètres objectifs associés sont bien connus, on trouvera une description des paramètres fondamentaux – temps d'établissement, temps de maintien, temps de commutation, etc. – dans la Recommandation UIT-T P.340. Les principaux paramètres objectifs importants sont les suivants:

- plage d'atténuation;
- temps de commutation (T_S), temps de maintien (T_H), temps d'établissement (T_R);
- atténuation dans le sens SND/RCV en période de parole simultanée;
- niveau minimal d'activation pour passer du sens RCV au sens SND et du sens SND au sens RCV;
- atténuation de l'écho.

Les paramètres fondamentaux liés à la commutation vocale sont les suivants: niveau seuil (V_{TH}), temps d'établissement (T_R), temps de maintien (T_H), temps de commutation (T_S) et plage d'atténuation (a_H). Un choix approprié de valeurs pour les paramètres de commutation permet de minimiser la dégradation de la qualité vocale introduite par la commutation vocale. Un choix incorrect de valeurs, notamment de valeurs de temps de commutation, peut conduire à de graves effets de mutilation et à la perte de signaux vocaux au début ou à la fin d'une intervention.

Il faut choisir les niveaux seuils de sorte que la commutation ne soit pas interrompue par des sources de bruit aléatoire (aux alentours) à l'une ou l'autre extrémité de l'appel. En outre, les effets du bruit ambiant de salle/réseau sur le seuil ne doivent pas entraîner une dégradation de la qualité de fonctionnement. On peut utiliser les niveaux de bruit ambiant pour améliorer la qualité en termes de seuil, étant donné que les locuteurs ont tendance à parler plus fort dans un environnement bruyant que dans un environnement calme.

Le temps d'établissement doit être suffisamment court pour ne pas perdre les composantes transitoires initiales des signaux vocaux, mais pas trop court pour que la suppression de l'affaiblissement d'insertion ne soit pas bruyante.

Le temps de maintien doit être suffisamment long pour couvrir les pauses moyennes dans les signaux vocaux de sorte qu'il n'y ait pas de commutation intermittente non souhaitée avant que le locuteur initial n'ait terminé, mais suffisamment court pour permettre une interruption raisonnable de la part du second locuteur.

Le temps de commutation pour passer d'un état actif à l'autre doit être équilibré afin de simuler au mieux le fonctionnement duplex intégral. Le temps de commutation dépend en outre du temps d'établissement et du temps de maintien.

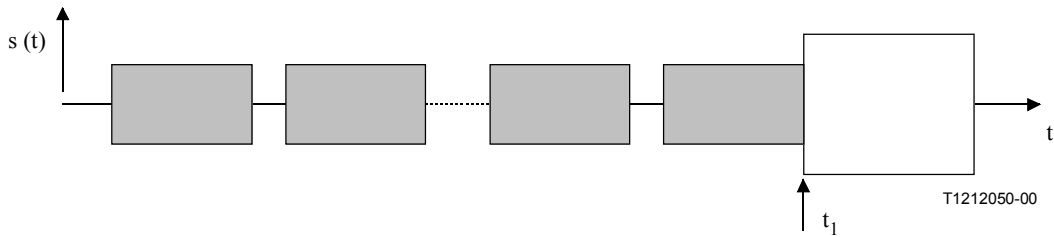
La plage d'atténuation est obtenue à partir de la différence entre le niveau maximal correspondant à l'activation complète et le niveau minimal obtenu immédiatement après l'inversion de transmission.

8.3 Méthodes d'analyse

Tous les niveaux indiqués dans les Tableaux 5 à 7 se rapportent au point MRP dans le cas des niveaux acoustiques ou au point de référence électrique. En cas d'accès électrique au lieu de l'accès acoustique, les niveaux à utiliser sont indiqués entre parenthèses.

8.3.1 Plage d'atténuation et caractéristiques de commutation

L'un des paramètres les plus importants, notamment pour des implémentations avec des dispositifs de commutation de niveau, est la **plage d'atténuation**. On peut évaluer ce paramètre au moyen de la structure de signal de test représentée sur la Figure 18.



NOTE – La ligne en pointillés indique que le signal de test est répété ou allongé pour obtenir la durée convenable pour la mesure.

Figure 18/P.502 – Structure de signal de test pour mesurer la plage d'atténuation

On utilise une répétition périodique de rafales de signaux CSS comme signaux vocaux simulés pour activer l'un des trajets de transmission (couleur grise). A la fin d'une rafale de signaux CSS, indiquée par t_1 sur l'échelle temporelle, le signal de mesure est appliqué sur le trajet opposé (couleur blanche). Ce signal est constitué de la répétition périodique d'un son voisé.

Les valeurs types sont données dans le Tableau 5.

Tableau 5/P.502

	Signal de mesure	Niveau du signal de mesure	Signal d'activation (dans le sens opposé)	Niveau du signal d'activation (dans le sens opposé)
Commutation RCV -> SND	Son voisé dans le sens SND, répétition périodique	-3 dBPa -16,7 dBm	CSS dans le sens RCV	-18,3 dBm (y compris les pauses)
Commutation SND -> RCV	Son voisé dans le sens RCV, répétition périodique	-16,7 dBm	CSS dans le sens SND	-4,7 dBPa (-18,3 dBm) (y compris les pauses)

On peut mesurer les paramètres suivants:

- **Plage d'atténuation**

Pour mesurer la plage d'atténuation, on active d'abord le sens opposé avant de procéder à la mesure dans le sens considéré. La plage d'atténuation est décrite comme étant la différence entre le niveau minimal et le niveau maximal du signal de test transmis par rapport au signal d'excitation. Pour effectuer la mesure, on évalue simplement le niveau en fonction du temps. Les constantes de temps à choisir pour cette mesure sont généralement de l'ordre de 5 ms.

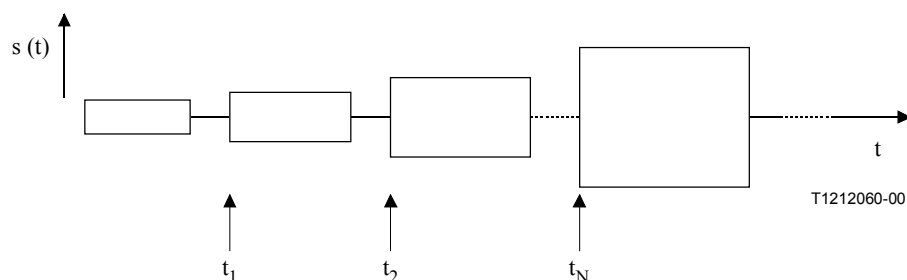
On trouvera les limites de la plage d'atténuation pour les différents types de terminaux mains-libres dans la Recommandation UIT-T P.340. Ces limites peuvent s'appliquer à d'autres dispositifs introduisant une commutation.

- **Caractéristiques de commutation (pour des signaux semblables à la voix humaine), par exemple temps de commutation**

On trouvera les définitions générales (et les limites) relatives aux caractéristiques de commutation dans la Recommandation UIT-T P.340. Les mesures sont grosso modo effectuées comme décrit ci-dessus. Cela signifie que le sens opposé est d'abord activé puis que la mesure est faite dans le sens considéré (voir le signal de test de la Figure 18). On mesure le niveau du signal de test transmis en fonction du temps. Les constantes de temps à appliquer pour les mesures sont à nouveau de l'ordre de 1 ms.

8.3.2 Niveau de seuil et temps d'établissement (pour le niveau minimal d'activation)

La structure de signal représentée sur la Figure 19 est constituée de parties de signal de niveau croissant. On peut utiliser ces séquences pour évaluer le **niveau minimal d'activation** pour passer du mode repos au sens RCV ou au sens SND. Des périodes de signal CSS (comme signal vocal simulé) de niveau croissant conviennent pour ce signal.



NOTE – La ligne en pointillés indique que le signal de test est répété ou allongé pour obtenir la durée convenable pour la mesure.

Figure 19/P.502 – Structure de signal de test pour déterminer le niveau minimal d'activation

On peut généralement choisir les valeurs données dans le Tableau 6.

Tableau 6/P.502

	Durée active/durée de pause	Niveau de la première période	Différence de niveau entre deux périodes
Signal CSS pour le passage au sens SND	248,62 ms/ 451,38 ms	-23 dBPa (Note) (-36,7 dBm)	1 dB
Signal CSS pour le passage au sens RCV	248,62 ms/ 451,38 ms	-36,6 dBm (Note)	1 dB
NOTE – Ces niveaux doivent être suffisamment faibles, pour faire en sorte que la mesure couvre une large plage de niveaux.			

Si on mesure les signaux transmis et qu'on les rapporte au signal de mesure d'origine, on peut déterminer le niveau minimal d'activation. On peut analyser l'activation au début de chaque rafale de signaux (t_1, t_2, \dots, t_N).

Avec ces signaux, on peut évaluer les paramètres suivants:

- **Niveau de seuil (pour des signaux semblables à la voix humaine)**

La séquence de mesure est représentée sur la Figure 19. L'analyse nécessaire pour déterminer le niveau minimal d'activation est une simple analyse de niveau en fonction du temps. Pour la mesure, on choisit une constante de temps comprise entre 1 et 5 ms et on analyse le niveau du signal mesuré en fonction du temps. On affiche le signal d'excitation et on détermine le niveau minimal d'excitation nécessaire pour activer le dispositif testé simplement en évaluant la différence de niveau pendant les parties actives du signal de source composite. Comme on connaît le niveau d'excitation, on peut déterminer le niveau seuil minimal.

- **Temps d'établissement (pour des signaux semblables à la voix humaine, en fonction du niveau)**

L'analyse est grosso modo la même que celle décrite au 8.3.1 sauf que la constante de temps vaut maintenant 1 ms. On détermine alors le temps de commutation par une évaluation du graphe donnant le niveau en fonction du temps.

8.3.3 Temps de maintien

Pour représenter le passage d'un mode actif au mode repos, on peut envoyer un signal d'activation (par exemple un son voisé constitué de signaux CSS) dans un sens, suivi par un second signal dans le même sens mais avec un niveau inférieur, qui n'active pas le téléphone mains-libres (signal de bruit) (voir la Figure 20). La seconde partie du signal mesuré indique donc l'atténuation, à partir de laquelle on peut déterminer le temps de maintien (temps d'extinction).

La durée du son voisé est de 0,5 s afin de parvenir à une condition finale de système stable. Le niveau correspond aux niveaux normalisés. Si une évaluation en fonction du niveau est nécessaire, on peut choisir les niveaux définis au 8.3.2. La seconde partie du signal (signal de bruit) dure 1 s. Il faut choisir un niveau suffisamment faible de manière à ne pas activer l'équipement. Les niveaux qu'il est suggéré d'appliquer sont les suivants: $-34,7$ dBPa pour les terminaux dans le sens émission, -50 dBm pour l'accès électrique dans le sens émission et -50 dBm pour la réception.

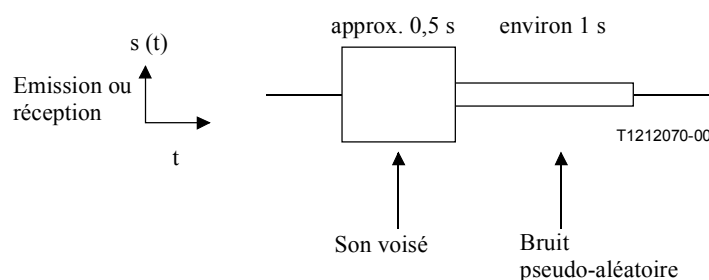
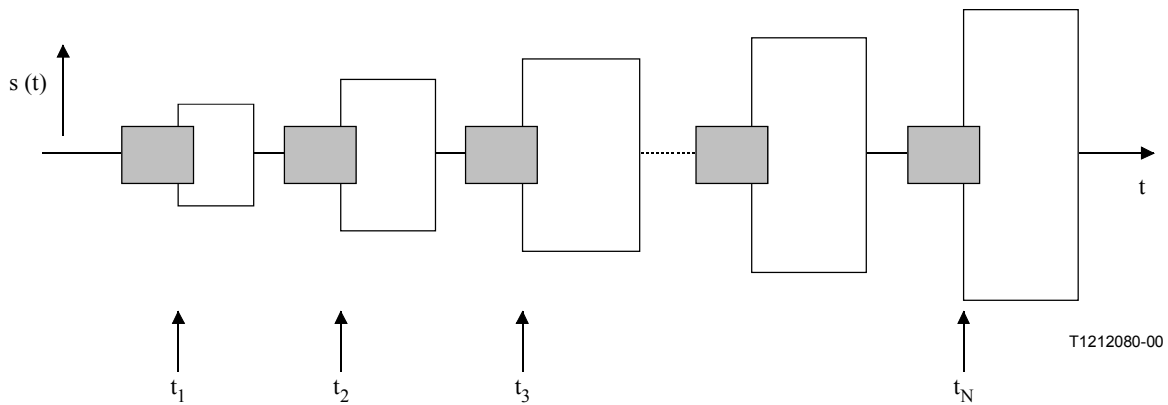


Figure 20/P.502 – Mesure de la réponse correspondant à l'extinction

8.3.4 Niveau seuil et temps de commutation pour passer de RCV à SND (de SND à RCV)

Si les **niveaux seuils pour passer** du sens RCV au sens SND (ou vice versa, c'est-à-dire du sens SND au sens RCV) doivent être mesurés, on peut utiliser les signaux de test donnés avec de légères modifications. Comme indiqué sur la Figure 21, un autre signal est nécessaire dans le sens de transmission opposé (couleur grise). Le niveau du signal de mesure (couleur blanche) augmente à nouveau périodiquement. Des périodes de signal CSS conviennent pour les deux signaux représentés sur la Figure 21, si on doit déterminer les caractéristiques de commutation en appliquant des signaux semblables à la voix humaine. Il faut à nouveau choisir des signaux non corrélés.



NOTE – La ligne en pointillés indique que le signal de test est répété ou allongé pour obtenir la durée convenable pour la mesure.

Figure 21/P.502 – Structure de signal de test pour déterminer le niveau minimal d'activation pour la commutation

Les valeurs à utiliser sont données dans le Tableau 7.

Tableau 7/P.502

	Durée active/durée de pause	Niveau de la première période	Différence de niveau entre deux périodes	Niveau (partie active) dans le sens de transmission opposé
Signal CSS pour passer au sens SND	248,62 ms/ 451,38 ms	-13 dBPa (-26,7 dBm)	1 dB	-16,7 dBm(RCV)
Signal CSS pour passer au sens RCV	248,62 ms/ 451,38 ms	-26,7 dBm	1 dB	-3 dBPa (SND) (-16,7 dBm)

A nouveau, on peut analyser l'activation au début des rafales de signaux (t_1, t_2, \dots, t_N).

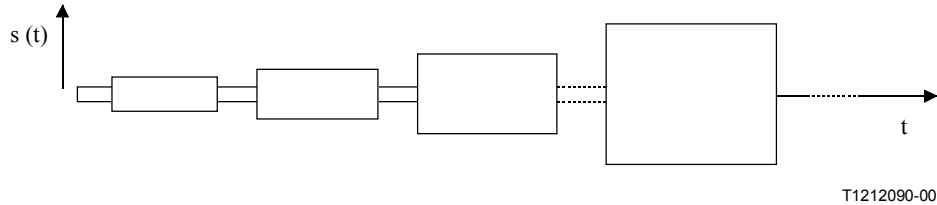
En outre, on peut réaliser les mêmes tests en appliquant un bruit de fond simulé dans le trajet de transmission opposé.

Les paramètres qu'il est possible d'évaluer sont les suivants:

- **Niveau minimal d'activation (pour des signaux semblables à la voix humaine) pour la commutation**
Le niveau minimal d'activation pour la commutation est déterminé comme décrit pour le niveau minimal d'activation, la seule différence est que l'évaluation est faite pendant les pauses de la séquence de parole simultanée. (Voir les temps $t_1, t_2 \dots$ sur la Figure 19).
- **Temps de commutation**
L'analyse est identique à celle décrite pour les temps de commutation nécessaires pour le niveau minimal d'activation (8.3.2).

8.3.5 Caractéristiques de commutation en présence de bruit de fond

On peut utiliser la structure de signal représentée sur la Figure 22 pour déterminer les **caractéristiques de commutation en présence de bruit de fond**. Dans ce cas, on applique simultanément un signal semblable à la voix humaine (CSS) et un bruit de fond simulé sur le même canal. Les paramètres relatifs au signal CSS peuvent être extraits des tableaux donnés plus haut.



NOTE – La ligne en pointillés indique que le signal de test est répété ou allongé pour obtenir la durée convenable pour la mesure.

Figure 22/P.502 – Structure de signal de test pour déterminer le niveau minimal d'activation en présence de bruit de fond

Les paramètres à déterminer sont les suivants:

- **Niveau seuil minimal en présence de bruit de fond**

Pour cette mesure, on applique un bruit de fond (en plus du signal d'excitation, voir la Figure 22). On choisit un bruit de H_{oth} ou un bruit de fond type (de préférence sans grandes fluctuations de niveau) avec un niveau conforme à l'utilisation type du dispositif testé (par exemple téléphone mains-libres). Pour les téléphones de bureau, le niveau type est dans la gamme -54 à -44 dBPa(A). Pour d'autres applications, on peut utiliser d'autres types de bruit de fond – par exemple un bruit de voiture – avec des niveaux différents.

Le niveau minimal d'activation est déterminé comme décrit pour le niveau minimal d'activation sans bruit de fond. Les séquences des signaux de mesure sont choisies de la même façon.

- **Temps d'établissement en présence de bruit de fond**

Ici encore, on utilise la même procédure que celle décrite avant pour l'évaluation des temps de commutation pour le niveau seuil minimal (8.3.2). La seule différence est la présence de bruit de fond appliqué comme décrit ci-dessus.

ANNEXE A

Méthode de mesure détaillée de l'affaiblissement ERL_t avec pondération temporelle

A.1 Algorithme relatif à l'affaiblissement d'adaptation pour l'écho

On décrit la méthode de mesure de l'affaiblissement d'adaptation pour l'écho avec pondération temporelle ERL_t . Pour cette méthode, il faut enregistrer l'écho et le signal de source pendant toute la durée de la mesure et il faut utiliser un post-traitement. Des techniques de mesure en temps réel sont possibles mais elles ne sont pas décrites dans la présente Recommandation.

Il n'est pas recommandé de "geler" l'annuleur pour les mesures de l'affaiblissement ERL. Certains résultats avec des signaux non stationnaires montrent que les temps de convergence et l'affaiblissement ERL qui en découle après le "dégel" dépendent de l'instant où l'annuleur a été "gelé".

A.1.1 Affaiblissement d'adaptation pour l'écho, avec pondération temporelle (ERL_t)

L'affaiblissement ERL avec pondération temporelle (ERL_t) est destiné à:

- fournir une mesure de l'affaiblissement d'adaptation pour l'écho variable dans le temps comportant des valeurs crêtes, avec pondération psychoacoustique (ERL_t);
- fournir une évaluation du nombre de rafales d'écho susceptibles d'être désagréables et l'affaiblissement d'adaptation pour l'écho avec pondération psychoacoustique pendant les rafales.

Le signal d'écho est d'abord filtré pour modéliser la sélectivité en fréquence de l'ouïe humaine pour des niveaux sonores de 30 phones, comme décrit au A.1.2. Cela permet de pondérer la puissance de l'écho de la même façon que le ferait l'ouïe humaine.

On peut alors appliquer une réduction de bruit et synchroniser les fichiers d'écho et de stimulus. La réduction de bruit consiste à mesurer le bruit et à le soustraire de l'écho plus bruit pour arriver à une meilleure estimation de l'écho seul. Cette mesure doit durer au moins deux secondes après l'arrêt de toute activité du stimulus. L'écho et la source sont convertis en trames de 4 ms avec puissance moyennée, pour que la résolution et l'immunité aux erreurs de synchronisation soient adéquates.

Si le stimulus est inactif, l'algorithme ne fait que sauter cette trame et passe aux trames d'écho et de stimulus suivantes. Si le stimulus est déclaré actif, la trame d'écho est comparée à un seuil pour déterminer si un événement d'écho se produit. La période d'activité de l'écho entre des états d'écho inactifs est appelée un "événement" d'écho. Ces événements sont ensuite pondérés au moyen d'une modélisation psychoacoustique.

En utilisant un seuil de -65 (5 dB au-dessus du plancher de bruit de la loi μ ou de la loi A), on peut déterminer ERL_t . De même, pour la loi A, le seuil doit aussi être supérieur de 5 dB au plancher de bruit. L'algorithme de mesure en pseudocode est détaillé au A.1.4.

A.1.2 Modélisation de l'audibilité de l'écho

En ce qui concerne la modélisation de l'audibilité de l'écho, l'algorithme tient compte de 3 aspects fondamentaux du comportement de l'ouïe humaine:

- 1) la sélectivité en fréquence de l'ouïe humaine pour des niveaux sonores de 30 phones (réponse de "Fletcher-Munson" équivalente à 30 dB pour 1 kHz) [1].

On choisit la valeur de trente phones car elle représente les niveaux d'écho pour des terminaux qui sont juste en dehors des spécifications d'équivalent de couplage des terminaux à combiné (déterminées par une analyse de planification d'équivalent). Si on fait varier le niveau sonore entre 20 et 50 phones, on obtient essentiellement la même pondération dans la bande téléphonique. On utilise un filtre pondéré A.

Il est à noter que l'utilisation de cette caractéristique exacte de pondération suppose une écoute de type casque/combiné ou une réponse MAP ("mean audible pressure", pression acoustique moyenne). Pour une écoute en champ libre, par exemple au moyen d'un téléphone mains-libres, il faudrait la pondération MAF ("mean audible field", champ acoustique moyen) de Robinson et Dadson, mais la différence est minime. On utilisera la pondération MAP pour mieux refléter l'utilisation plus courante du combiné.

L'affaiblissement moyen du filtre avec un bruit blanc est de 1,3 dB lorsqu'il est mesuré au moyen de ERL_s ou ERL_t . Avec des signaux non stationnaires, l'affaiblissement dépendra du temps.

- 2) la tendance de l'oreille à combiner le volume sonore de signaux séquentiels même s'ils sont discontinus ("combinaison temporelle"). Cela se produit généralement lorsque les deux signaux sont séparés par une période de silence de durée inférieure à 20 ms [2], [3], [4]. Si deux rafales d'écho sont séparées par une période d'inactivité inférieure à 20 ms, elles sont considérées comme un seul événement d'écho plus long en ce qui concerne le volume sonore. Cet événement se poursuit jusqu'à ce que la coupure entre événements soit d'au moins 20 ms et c'est à ce moment que l'événement d'écho est déclaré terminé. Cette coupure peut être considérée comme un maintien de 20 ms pour l'événement d'écho en cours. Pendant cette période de maintien, les puissances de l'écho et du stimulus ne sont pas incluses dans l'événement. La Figure A.1 donne un exemple de combinaison temporelle.

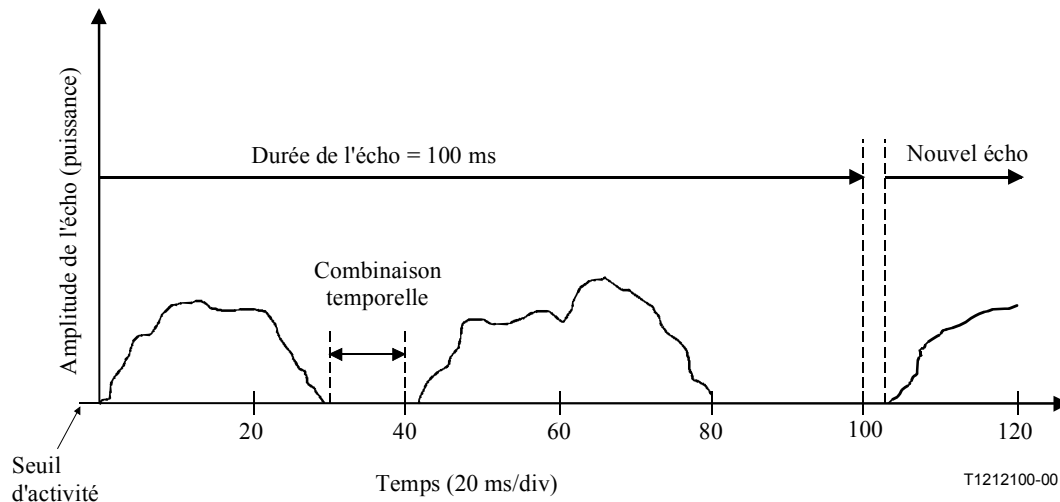


Figure A.1/P.502

- 3) la durée de l'événement d'écho total après combinaison temporelle est mesurée sur la base du comportement d'intégration temporelle naturelle de l'oreille. La durée totale comprend toutes les coupures entre événements qui sont saisies par combinaison temporelle, mais pas le maintien final de 20 ms.

Si la durée est inférieure à 750 ms, on réduit le niveau de l'événement pour tenir compte du comportement d'intégration temporelle de l'ouïe humaine. Une équation décrivant la relation a été déduite d'études d'audition avec du bruit:

$$\text{Pondération liée à l'intégration temporelle} = -23 + 8 \log(t) \text{ en dB}$$

$$\text{où } t = \text{durée totale de l'événement d'écho (ms), } t < 750 \text{ ms}$$

Il est à noter que les tonalités se traduisent par une relation légèrement différente, mais on a estimé que le bruit s'approchait beaucoup plus de la vraie nature de l'écho qu'une sinusoïde.

La Figure A.2 représente graphiquement la pondération temporelle.

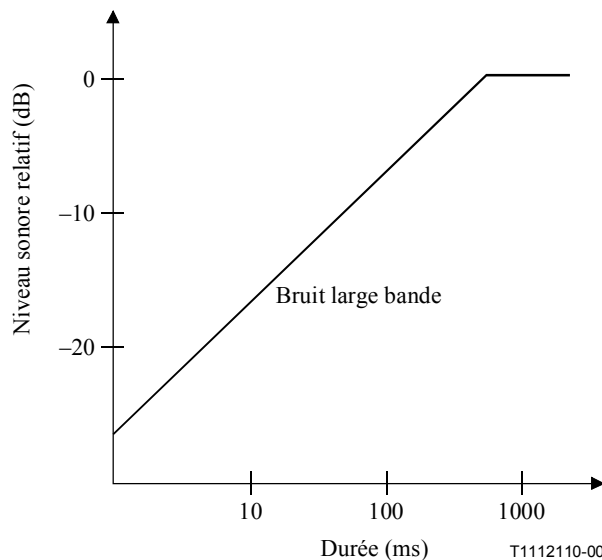


Figure A.2/P.502

Si la durée est supérieure à 750 ms, le niveau de l'événement total n'est pas pondéré. Il est à noter que des résultats de test montrent que des rafales d'écho de durée inférieure à 750 ms sont monnaie courante à partir d'annuleurs.

A.1.3 Expression des résultats de mesure de ERL_t

Les méthodes traditionnelles de mesure de ERL rapportent la puissance de l'écho pendant la durée de la mesure à la puissance de la source pendant la durée de la mesure pour aboutir à l'affaiblissement d'adaptation pour l'écho. Dans la présente méthode, la puissance pondérée finale de l'écho pendant chaque événement est rapportée à la puissance du signal de source pendant le même événement, pour aboutir à l'affaiblissement "ERL_t actif" – AERL_t – de chaque événement. L'écho n'est rapporté au signal de source que pendant l'événement, car c'est la façon dont notre oreille comparerait l'écho.

On obtient une moyenne à long terme de l'affaiblissement d'adaptation pour l'écho actif pondéré en faisant la somme des puissances de tous les échos pondérés pendant les événements actifs et en rapportant cette somme à la puissance de la source telle qu'elle est vue pendant tous les événements uniquement. Le résultat est l'affaiblissement "ERL_t actif à long terme".

Pour comparaison, dans le cas des méthodes traditionnelles de mesure de ERL, on fait la somme des puissances de tous les échos pondérés pendant les événements, puis on la rapporte à la puissance totale de la source telle qu'elle est mesurée pendant toute la durée de la mesure. Le résultat est l'affaiblissement "ERL_t à long terme".

Il est à noter que la terminologie relative aux résultats de mesure de ERL_t a été choisie de sorte qu'elle soit cohérente avec la Recommandation UIT-T P.56.

Les statistiques comprennent en outre les affaiblissements AERL_t minimal et maximal, l'écart-type ("sigma") de AERL_t, la moyenne de AERL_t et le nombre total d'événements d'écho (les événements combinés dus à l'effet "Haas" sont considérés comme un seul événement total). Sont aussi inclus: le nombre d'événements d'écho par minute, le pourcentage de signaux vocaux sans événement d'écho, le nombre d'événements de durée inférieure à 750 ms, la longueur moyenne d'un événement et la durée d'inactivité de la source.

A.1.4 Algorithme de mesure de ERL_t

Cet algorithme permet d'évaluer l'affaiblissement d'adaptation pour l'écho d'un terminal sur la base d'une modélisation psychoacoustique et de prédire les occurrences d'échos susceptibles d'être désagréables. Il incorpore 3 aspects fondamentaux de l'audition humaine:

- sélectivité en fréquence de l'ouïe humaine (réponse de "Fletcher Munson");
- sommation temporelle du niveau des événements séparés de moins de 20 ms (effet "Haas");
- intégration temporelle pour les stimuli de durée inférieure à 750 ms.

L'implémentation de l'algorithme est détaillée ci-après.

On utilise un signal de source tel que décrit dans les paragraphes 4 et 5. Il est recommandé d'utiliser des signaux de stimulus fondés sur la parole car les résultats sont alors les plus représentatifs de l'utilisation réelle. Un signal d'écho ou de bruit traverse toujours le système sans être annulé.

Le stimulus et l'écho doivent être enregistrés et mis à disposition en format numérique. Des données d'utilisateur concernant le type de poste (analogique ou numérique), le retard EPD_n et des tests en période de parole simultanée ou en période de conversation simple doivent être disponibles. Il faut utiliser des paramètres d'étalonnage pour ramener les trames d'écho et de stimulus à des valeurs absolues et il faut que le traitement hybride ait supprimé l'écho hybride pour les postes analogiques 2 fils.

Les fichiers de stimulus et d'écho seront traités sous la forme de valeurs de puissance moyennées sur des trames de 4 ms. Les trames successives du fichier de stimulus seront désignées par x_i , les trames d'écho seront désignées par y_i , où $i = 1, 2, 3, \dots$ est l'indice de trame. Les trames intermédiaires conformes à un "événement d'écho" seront désignées par x_k et y_k , où $k = 1, 2, 3, \dots$ est l'indice de l'événement d'écho et est réinitialisé lorsque l'événement prend fin et qu'un nouveau commence.

Les statistiques calculées pendant la mesure de ERL_t comprennent l'affaiblissement ERL_t actif à long terme ($ALERL_t$), l'affaiblissement ERL_t à long terme ($LERL_t$), les affaiblissements ERL_t actifs minimal et maximal ($MINERL$, $MAXERL$), leur sigma et leur valeur moyenne, le nombre total d'événements d'écho (les événements combinés dus à l'effet "Haas" sont considérés comme un seul événement total) ($NEVENTS$). Sont aussi inclus: le nombre d'événements d'écho par minute ($NEVMIN$), le pourcentage de signaux vocaux sans événement d'écho (PER), le nombre d'événements de durée inférieure à 750 ms ($N750$), la longueur moyenne d'un événement ($AVGEVENT$) et la durée pendant laquelle le stimulus était inactif (DUR). La terminologie relative aux résultats de mesure de ERL_t a été choisie de manière à être cohérente avec la Recommandation UIT-T P.56. La durée d'inactivité de stimulus n'est pas comprise dans les résultats de type temporel.

Algorithme de mesure de ERL_t

- *Etape 1* (facultative mais recommandée)
Calculer la corrélation entre le fichier de stimulus et le fichier d'écho pour déterminer avec précision EPD_n . Utiliser le critère selon lequel le pic de corrélation présent se produit pour EPD_n sauf si un pic de corrélation suivant a une amplitude supérieure d'au moins 10 dB. Ce critère approché est fondé sur des études subjectives sur la détection de retard avec plusieurs impulsions.
- *Etape 2*
Aligner le fichier d'écho et le fichier de stimulus dans le temps en supprimant du fichier d'écho le retard égal à EPD_n .

- *Etape 3*
Chacun des échantillons d'écho est traité par un filtre donnant une valeur approchée du contour de volume sonore égal à la pression acoustique moyenne pour 30 phones. On peut obtenir une valeur précise (à ± 1 dB près de 200 Hz à 2500 Hz) au moyen d'un filtre passe-haut du premier ordre dont le point à -3 dB est à 800 Hz.
- *Etape 4*
Si on peut supposer que le bruit sur le trajet d'écho est stationnaire et non corrélé avec l'écho, on mesure le bruit pendant 2 secondes après l'arrêt de l'activité de la source et de l'écho. On soustrait ensuite le bruit de l'écho plus bruit pour aboutir à une meilleure estimation de l'écho seul.
- *Etape 5*
Les échantillons sont convertis en valeurs absolues au moyen des données d'étalonnage. Les échantillons de stimulus sont combinés en trames de 4 ms avec puissance moyennée, désignées par x_i . Les échantillons d'écho, pour lesquels le bruit a été filtré et qui sont pondérés, sont combinés en trames de 4 ms avec puissance moyennée, désignées par y_i .
- *Etape 6 – début des calculs d'affaiblissement d'adaptation pour l'écho*
Initialiser les variables:
 $i = 0$ (compteur de trames);
 $j = 0$ (compteur de trames pour la durée de signal inactif);
 $n_{k=0} = 0$ (nombre de trames dans l'événement d'écho courant);
NSAMPS = 0 (nombre accumulé de trames pour tous les événements);
HAAS = 0 (compteur jusqu'à 20 ms);
 $e_{i=0} = 0$ (sommation de toutes les puissances d'écho pour tous les événements après pondération, telle qu'elle est vue au compteur de trames i);
 $p_{i=0} = 0$ (sommation de toutes les puissances de stimulus pendant la mesure, telle qu'elle est vue au compteur de trames i);
 $e_{k=0} = 0$ (sommation des puissances d'écho pendant l'événement d'écho particulier après pondération, telle qu'elle est vue au compteur de trames d'événement k);
 $s_{k=0} = 0$ (sommation des puissances de stimulus pendant l'événement d'écho particulier après pondération, telle qu'elle est vue au compteur de trames d'événement k);
WEIGHT = 0 (poids en fonction du temps de l'événement le plus récent);
LEVENT = 0 (niveau d'affaiblissement d'adaptation pour l'écho de l'événement le plus récent, après pondération);
NEVENT = 0 (nombre total d'événements d'écho);
N750 = 0 (nombre total d'événements d'écho de durée inférieure à 750 ms);
MINERL = 75 (niveau minimal d'affaiblissement d'adaptation pour l'écho parmi tous les événements);
MAXERL = 0 (niveau maximal d'affaiblissement d'adaptation pour l'écho pour tous les événements);
EVENT[NEVENT] = 0 (initialiser à zéro la matrice des niveaux d'affaiblissement de tous les événements (en dB); sert dans le calcul de sigma);
TEMPSK = 0 (somme des puissances de stimulus pendant tous les événements);
SUM = 0 (sert dans le calcul de sigma);
SQ = 0 (sert dans le calcul de sigma).

• *Etape 7*

Incrémenter le compteur de trames et lire la puissance d'écho moyennée sur 4 ms y_i ainsi que la puissance de stimulus moyennée sur 4 ms x_i ; s'il n'y a plus de données valides et si on est arrivé à la fin de l'un ou l'autre fichier de mesure, aller à l'étape 8.

1 $i = i + 1$ (sauf si c'est le dernier i ; dans ce cas, aller à l'étape 8).

Sommer les puissances de stimulus:

$$p_i = p_i + x_i$$

Le volume sonore du stimulus est-il suffisant pour que le calcul de l'affaiblissement d'écho soit valide? S'il est insuffisant, ignorer la trame courante et passer à la trame suivante.

4 Si $x_i < (\text{puissance efficace du stimulus à long terme } -25 \text{ dB})$

$$j = j + 1$$

$$i = i + 1$$

Aller à 4.

Sinon:

Comparer l'écho au seuil:

Si $y_i < -65 \text{ dB}$ {5 dB au-dessus du plancher de bruit de la loi μ ou de la loi A}

Incrémenter le compteur d'événements de trame:

$$k = k + 1$$

Incrémenter la longueur d'événement de trame y compris toute coupure $< 20 \text{ ms}$:

$$n_k = n_k + 1 + \text{HAAS}$$

Réinitialiser le compteur de "Haas":

$$\text{HAAS} = 0$$

Accumuler la puissance d'écho de l'événement:

$$e_k = e_k + y_i$$

Accumuler la puissance de stimulus pendant l'événement:

$$s_k = s_k + x_i$$

Aller à 1.

Sinon:

Est-ce qu'il n'y a pas eu d'événement pendant les 20 dernières millisecondes?

Si $k = 0$

$$\text{HAAS} = 0$$

Aller à 1.

Sinon:

Il y a eu un événement pendant les 20 dernières millisecondes:

$$\text{HAAS} = \text{HAAS} + 1$$

Est-ce qu'une période de 20 ms sans événement s'est terminée après un événement récent?

Si $\text{HAAS} * 4 < 20$

Aller à 1.

Sinon:

Un événement est terminé, ajouter un événement au compteur d'événements:

$$\text{NEVENT} = \text{NEVENT} + 1$$

Incrémenter le compteur de durée totale des événements en ajoutant la durée en nombre de trames de l'événement le plus récent:

$$NSAMPS = NSAMPS + n_k$$

La durée de l'événement le plus récent était-elle inférieure à 750 ms?

$$\text{Si } n_k * 4 < 750$$

Calculer la pondération liée à l'intégration temporelle pour l'événement d'écho le plus récent:

$$WEIGHT = 8 * \log_{10}(n_k * 4) - 23$$

Incrémenter le compteur du nombre d'événements qui ont été pondérés temporellement:

$$N750 = N750 + 1$$

Sinon:

Calculer l'affaiblissement d'adaptation pour l'écho pondéré de l'événement le plus récent en dB:

$$LEVENT = 10 * \log_{10}(s_k / e_k) - WEIGHT$$

Enregistrer les affaiblissements d'adaptation pour l'écho minimal et maximal en dB:

$$\text{Si } LEVENT < MINERL; MINERL = LEVENT$$

$$\text{Si } LEVENT > MAXERL; MAXERL = LEVENT$$

Enregistrer l'affaiblissement d'adaptation pour l'écho de l'événement le plus récent en dB pour le futur calcul de sigma:

$$EVENT(NEVENT) = LEVENT$$

Reconvertir l'affaiblissement d'adaptation pour l'écho de l'événement le plus récent en valeur linéaire; recalculer la puissance d'écho linéaire pondérée:

$$e_k = s_k / (10^{(LEVENT/10)})$$

Accumuler toutes les puissances d'événement d'écho pour le calcul futur de ALERL_t et LERL_t:

$$e_i = e_i + e_k$$

Accumuler toutes les puissances de stimulus pendant les événements pour le calcul futur de ALERL_t:

$$TEMPSK = TEMPSK + s_k$$

Réinitialiser les variables relatives aux événements d'écho:

$$k = 0$$

$$n_k = 0$$

$$WEIGHT = 0$$

$$HAAS = 0$$

$$e_k = 0$$

$$s_k = 0$$

Aller à 1.

- *Etape 8*

Calculer l'affaiblissement ERL_t actif à long terme ($ALERL_t$), l'affaiblissement ERL_t à long terme ($LERL_t$), le nombre d'événements d'écho par minute (NEVMIN), le pourcentage de signaux vocaux sans événement d'écho (PER), la longueur moyenne d'un événement (AVGEVENT) et la durée pendant laquelle la parole était inactive (DUR).

NOTE – Vérifier si la valeur de e_i est nulle avant le calcul; si $e_i = 0$, mettre $ALERL_t$ et $LERL_t$ à 100 dB.

$$ALERL_t = 10 * \log_{10}(TEMPSK/e_i)$$

$$LERL_t = 10 * \log_{10}(p_i/e_i)$$

$$NEVMIN = 60 * NEVENT / [(i-j) * 0,004] \text{ \{nombre d'événements par minute\}}$$

$$PER = 100 * [(i-j) - NSAMPS] / (i-j) \text{ \{pourcentage de signaux vocaux sans écho\}}$$

$$AVGEVENT = NSAMPS * 4 / NEVENT \text{ \{longueur moyenne d'un événement en millisecondes\}}$$

$$DUR = j * 0,004$$

Calculer sigma en analysant la matrice EVENT, qui contient l'affaiblissement d'adaptation pour l'écho de chaque événement; on donne à chaque événement un poids égal dans le calcul de sigma, quelle que soit sa durée; on suppose que c'est la transition entre événements et non leur durée qui est la plus désagréable.

Boucle j de 1 à NEVENT:

$$SUM = SUM + EVENT(j)$$

$$SQ = SQ + EVENT(j) ** 2$$

Fin de la boucle

$$SIGMA = \sqrt{SQ / NEVENT - [SUM / NEVENT] ** 2}$$

Calculer la moyenne des événements:

$$MEAN = SUM / NEVENT$$

- *Etape 9*

Statistiques résultantes:

Imprimer $ALERL_t$, $LERL_t$, MINERL, MAXERL, NEVENT, NEVMIN, PER, N750, AVGEVENT, DUR,

SIGMA, MEAN

ANNEXE B

Filtres de mesure en période de parole simultanée pour la méthode A

Pour les tests en période de parole simultanée, il faut utiliser des filtres coupe-bande et des filtres passe-bande à diverses fréquences. Une implémentation recommandée est indiquée dans le tableau ci-dessous.

Les termes décrits sont les suivants:

- fpl: fréquence la plus basse de la bande passante ou de la bande coupée à -3 dB;
- fpu: fréquence la plus haute de la bande passante ou de la bande coupée à -3 dB;
- fsl: fréquence la plus basse de la bande passante ou de la bande coupée à -atten dB;
- fsu: fréquence la plus haute de la bande passante ou de la bande coupée à -atten dB;

- atten: atténuation totale spécifiée du filtre;
- atten (réelle): atténuation totale réelle du filtre;
- ripple: ondulation du filtre en dB (\pm);
- gain: gain du filtre passe-bande (linéaire) dans la bande passante;
- order: ordre du filtre passe-bande (fréquence d'échantillonnage de 8 kHz). La durée d'oscillation du filtre passe-bande correspond à l'ordre multiplié par 125 μ s. Pour le filtre coupe-bande, l'ordre renvoie à l'ordre du biquad (elliptique).

Type de filtre	Passe-bande FIR 500 Hz	Passe-bande FIR 1 kHz	Passe-bande FIR 1,75 kHz	Passe-bande FIR 2,5 kHz	Coupe-bande IIR 500 Hz	Coupe-bande IIR 1 kHz	Coupe-bande IIR 1,75 kHz	Coupe-bande IIR 2,5 kHz
fpl	495	990	1 733	2 475	400	800	1 450	2 100
fpu	505	1 010	1 767	2 525	610	1 250	2 100	2 950
fsl	435	900	1 611	2 302	435	900	1 610	2 300
fsu	570	1 100	1 900	2 715	570	1 100	1 900	2 715
atten	30	30	30	30	30	30	30	30
atten (réelle)	31	29,5	34	34	30	40		
ripple	1	1	3	1	1,5	1,5	1,5	1,5
gain	0,92	0,9	0,78	0,99				
order	160	100	80	60	6	6	6	6

La durée d'oscillation du filtre passe-bande aura une incidence sur la mesure si on n'en tient pas compte. Il ne faut commencer les mesures qu'à la fin de l'oscillation du filtre due à l'application initiale. Cela est nécessaire de manière à pouvoir réaliser une mesure de référence correcte pour les évaluations de l'atténuation et les évaluations de l'écrêtage. La durée d'oscillation la plus longue est de 20 ms pour le filtre passe-bande de 500 Hz. Comme la fenêtre pour moyenniser les mesures d'atténuation est de 8 ms, le filtre passe-bande doit être inséré à $20 + 8 = 28$ ms avant le début de la parole simultanée, ou $60 - 0$ (voir Figure 7).

ANNEXE C

Description de la séquence de conditionnement

C.1 Conditionnement de l'annuleur avant la parole simultanée

On trouvera des informations de base sur la programmation d'une conversation dans la Recommandation UIT-T P.59 (rafales de parole, pauses, parole simultanée, silence mutuel). Les gabarits de conditionnement donnés ci-dessus sont extraits de cette Recommandation.

C.1.1 Gabarits d'activité pour le conditionnement en période de parole simultanée

On va maintenant spécifier les gabarits d'amplitude exacts conjointement avec les caractéristiques d'amplitude du signal en période de parole simultanée. Chaque type de test en période de parole simultanée a des spécifications particulières en termes de durée et d'amplitude du signal en période de parole simultanée. Pour l'affaiblissement d'adaptation pour l'écho, la durée doit être suffisamment longue pour détecter toute divergence, mais pas trop longue pour ne pas épuiser les ressources mémoire du système testé et pour ne pas entraîner un temps de calcul inacceptable. Des tests ont montré qu'une durée de parole simultanée de 20 secondes est acceptable pour la mesure de

l'affaiblissement d'adaptation pour l'écho en période de parole simultanée. A la fin de la parole simultanée (le locuteur qui a déclenché la parole simultanée devient inactif), on peut poursuivre la mesure de l'affaiblissement d'adaptation pour l'écho pendant 10 secondes (le locuteur actif juste avant la parole simultanée reste actif) pour mesurer la récupération après la parole simultanée. Après cela, il faut insérer deux secondes de silence. De cette façon, on peut mesurer le bruit sur le trajet d'écho. Si on peut supposer que le bruit et l'écho ne sont pas corrélés et que le bruit est stationnaire, on peut soustraire le bruit mesuré pendant les deux secondes de l'écho plus bruit en période de parole simultanée pour obtenir une mesure plus précise de l'écho en période de parole simultanée.

La durée de la parole simultanée pendant les évaluations de l'atténuation et de l'écrêtage en période de parole simultanée peut être beaucoup plus courte. Comme toutes les constantes de temps à l'étude doivent être inférieures à 200 ms, la durée de la parole simultanée est fixée à 200 ms. On poursuit l'analyse (le locuteur actif juste avant la parole simultanée reste actif) pendant une seconde après la fin de la parole simultanée pour les évaluations de l'atténuation afin de mesurer toute suppression d'affaiblissement au moment où on repasse à la conversation simple. Il n'est pas nécessaire d'évaluer le bruit et de le corriger pour les évaluations de l'atténuation et de l'écrêtage en période de parole simultanée.

Les gabarits recommandés sont donnés sur les Figures C.1 et C.2.

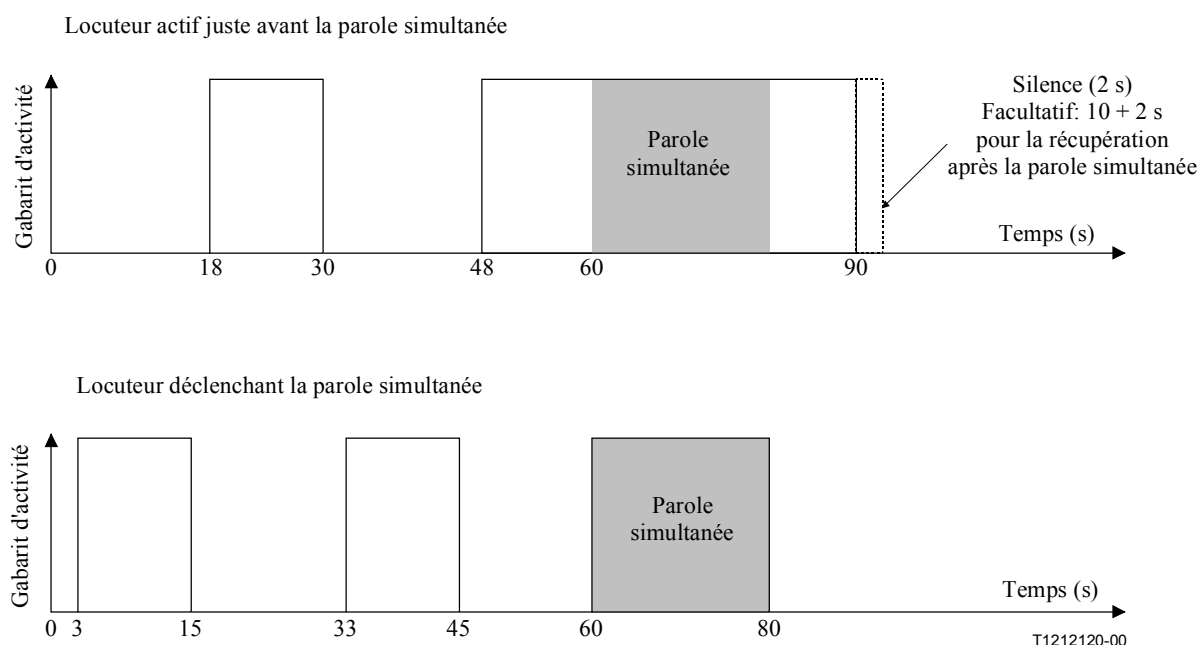


Figure C.1/P.502 – Gabarit d'activité pour la mesure de l'affaiblissement d'adaptation pour l'écho

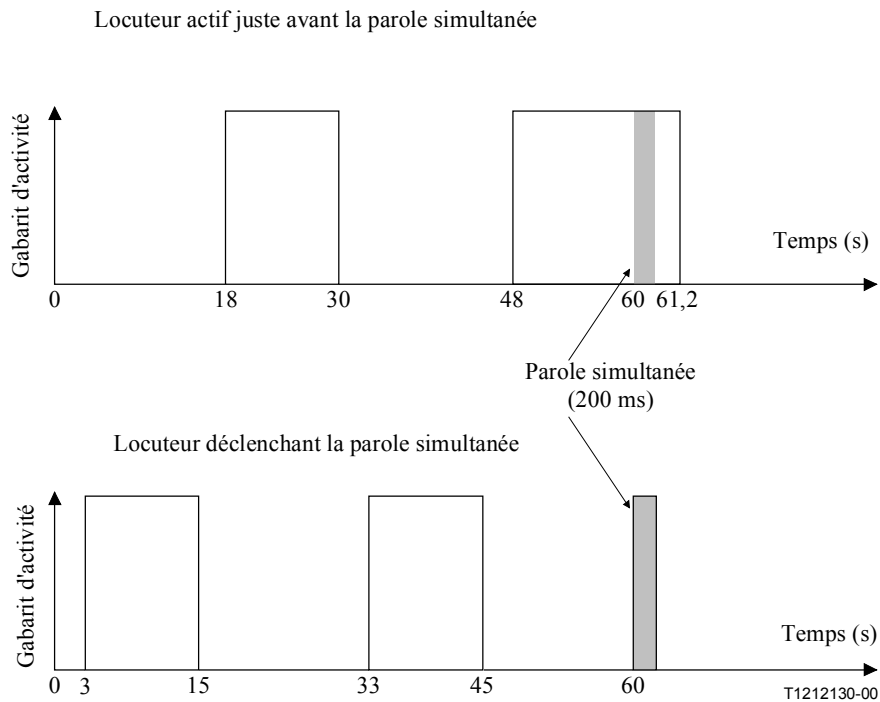


Figure C.2/P.502 – Gabarit d'activité pour la mesure de l'atténuation et celle de l'écrêtage

C.1.2 Synchronisation des gabarits d'activité pour le conditionnement en période de parole simultanée

Il faut synchroniser les gabarits pour éviter une parole simultanée prématurée. Pour cela, il faut tenir compte des 1,5 ms de retard sur le trajet aérien entre la bouche artificielle et le téléphone mains-libres. Il faut déclencher le signal du simulateur de bouche 1,5 ms avant le signal R_{in} en retardant le fichier de stimulus utilisé à la réception de 1,5 ms. On définit le début la parole simultanée comme l'instant où le microphone constate une activité valide à l'émission et le point R_{in} constate une activité valide à la réception.

Dans les diagrammes des gabarits d'activité, le point de départ à $t = 0$ correspond au début du fichier appliqué au point R_{in} . On peut estimer que le point de départ du signal du simulateur de bouche est à $t = -1,5$ ms, mais le terminal les verra synchronisés. Au point correspondant à 60 secondes, on passe au mode parole simultanée et on commence le test en période de parole simultanée.

C.1.3 Compensation associée aux filtres de mesure

Les méthodes de mesure en période de parole simultanée nécessitent que des filtres soient présents sur le trajet audio. Ces filtres auront une incidence sur la résolution temporelle et l'instant précis auquel le test en période de parole simultanée peut commencer. Si on utilise des filtres dont la durée d'oscillation est connue, on peut mettre en attente la mesure pendant cette durée.

APPENDICE I

Références bibliographiques

- [1] HEARING, GULICK, GESCHIEDER, FRISNA: *Oxford University Press*, 1989.
- [2] DAVIS (D.), DAVIS (C.): The LEDE Concept, *JAES*, 1985.
- [3] OLIVE (S.): The Detection of Reflections, *JAES*, 1987.
- [4] OLIVE (S.): Modification of Timbre by Resonance, *JAES*, 1988.
- [5] ZWICKER (E.), FASTL (H.): Psychoacoustics, *Springer Verlag*, 1990.
- [6] Enhancements of hands-free telecommunications, *Esprit Consortium, Annals of telecommunications*, 49 N^{os} 7-8, 1994.
- [7] Methodology of Evaluation and Standards, Deliverable 1.2, *Freetel*, juillet 1993.
- [8] GIERLICH (H.W.): The auditory perceived quality of hands-free telephones: auditory judgements, instrumental measurements and their relationship. *Speech Communication* 20, p. 241-254, octobre 1996,
- [9] Subjective valuation procedures for hands-free telephones – Double talk performance. *Contribution COM 12-5 de l'UIT-T*, Genève, avril 1997.
- [10] Subjective evaluation of hands-free telephones using conversational tests, specific double talk tests and listening only tests. *Contribution COM 12-6 de l'UIT-T*, Genève, avril 1997.
- [11] Double talk measurements for hands-free telephones: Measurement proposals and measurement results. *Contribution COM 12-32 de l'UIT-T*, Genève, février 1998.

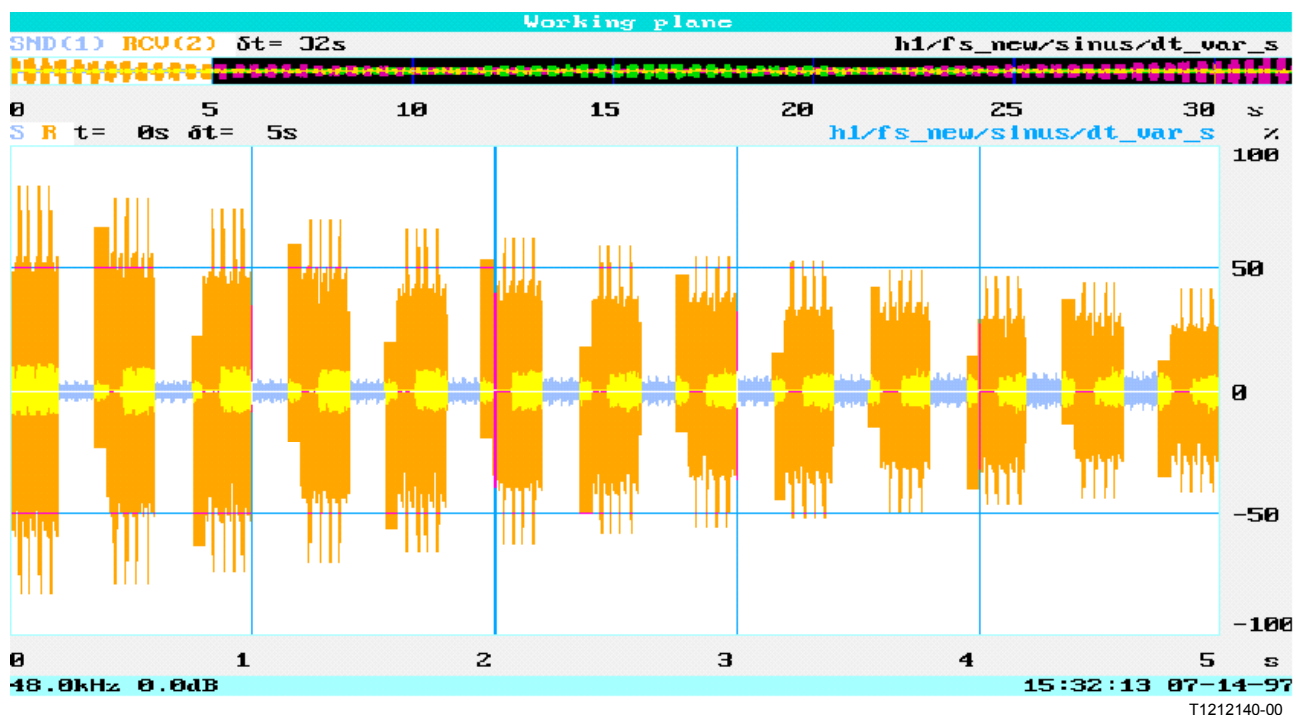
APPENDICE II

Exemples d'évaluation

II.1 Quelques exemples d'évaluation conformément au paragraphe 5

II.1.1 Réponses en fréquence en période de parole simultanée

Sur la base des évaluations de téléphones mains-libres, on donne quelques exemples d'application de la procédure dans la Figure II.1.

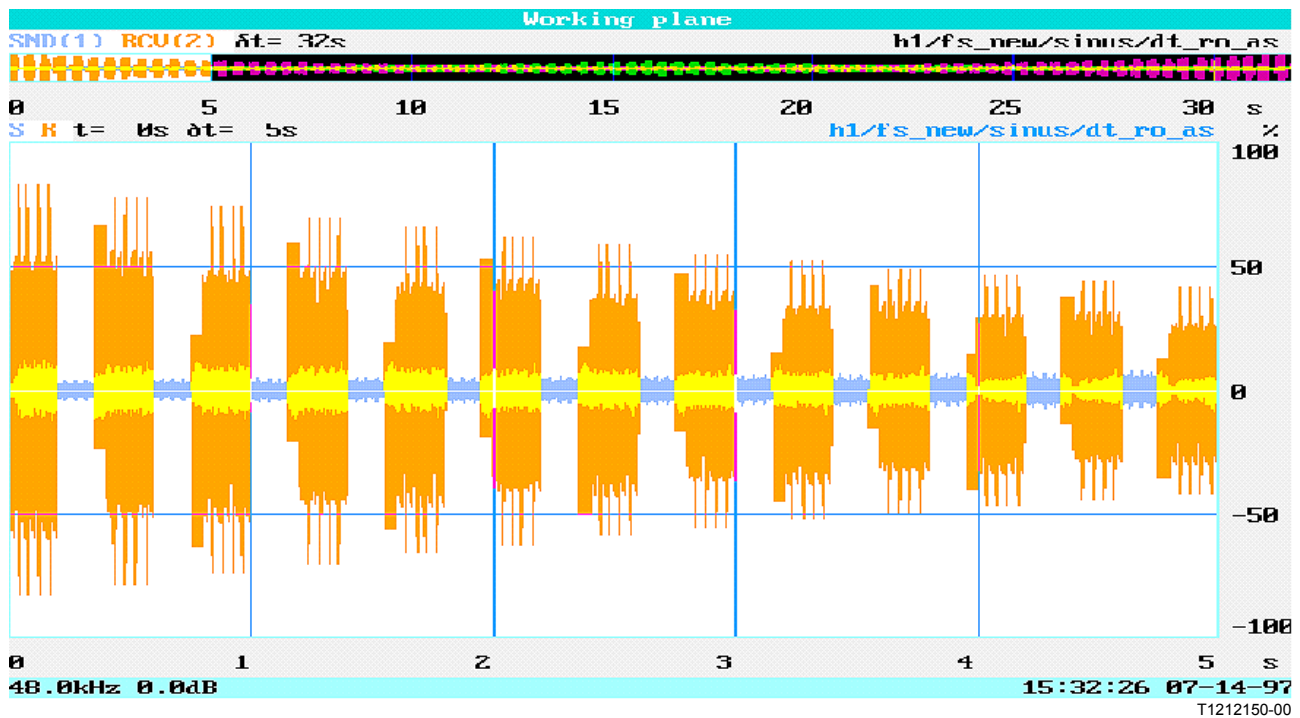


En couleur foncée: signal d'excitation dans le sens émission.

En couleur claire: signal mesuré dans le sens émission.

Les pauses sont remplies de signal de parole simultanée transféré du haut-parleur au microphone du téléphone mains-libres.

Figure II.1/P.502 – Sens émission d'un téléphone mains-libres à commutation de niveau



En couleur foncée: signal d'excitation dans le sens émission.

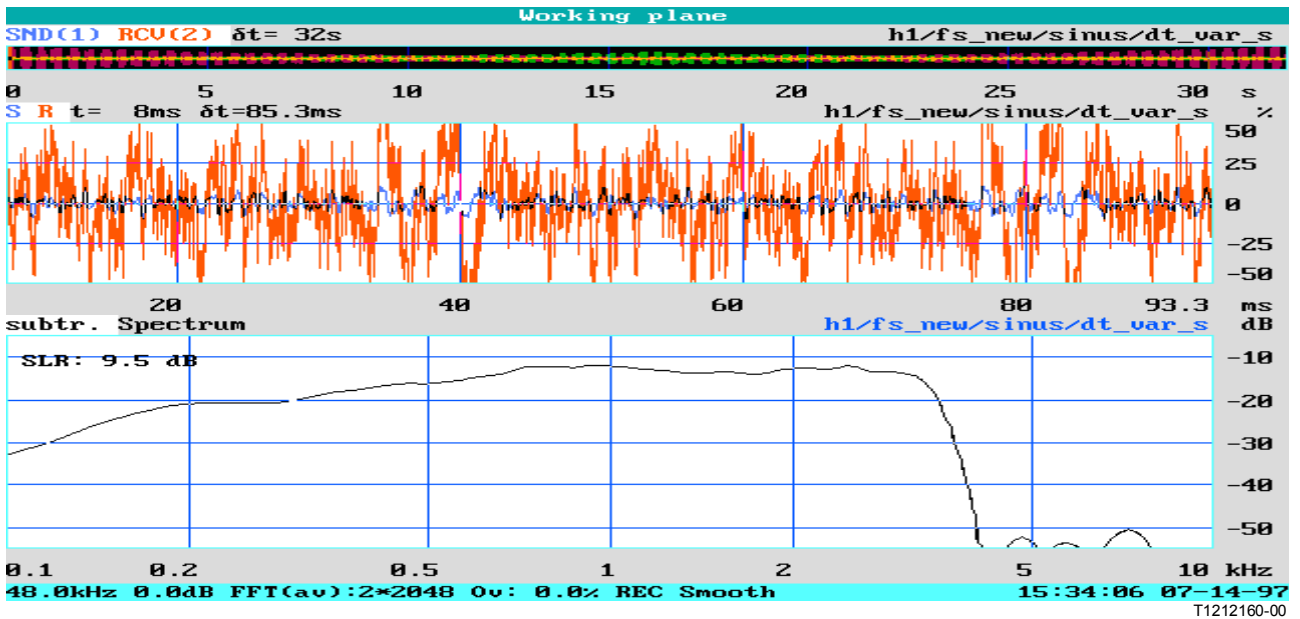
En couleur claire: signal mesuré dans le sens émission.

Les pauses sont remplies de signal de parole simultanée transféré du haut-parleur au microphone du téléphone mains-libres.

Figure II.2/P.502 – Sens émission d'un téléphone mains-libres à annulation d'écho et commutation de niveau

Les Figures II.1 et II.2 montrent les résultats de mesure lorsqu'on utilise ce type de signaux de test pour l'évaluation du sens émission de deux téléphones mains-libres. Ce sont uniquement les périodes où seul le signal dans le sens émission est présent qui sont prises en compte pour la mesure. Conformément à la Figure 3, il s'agit d'un intervalle de 150 ms qui commence 50 ms après le signal d'activation du son voisé. Pendant cet intervalle, seul le signal d'émission est présent. Les Figures II.1 et II.2 montrent les résultats de mesure simplement sous la forme d'une séquence temporelle pour les 5 premières secondes.

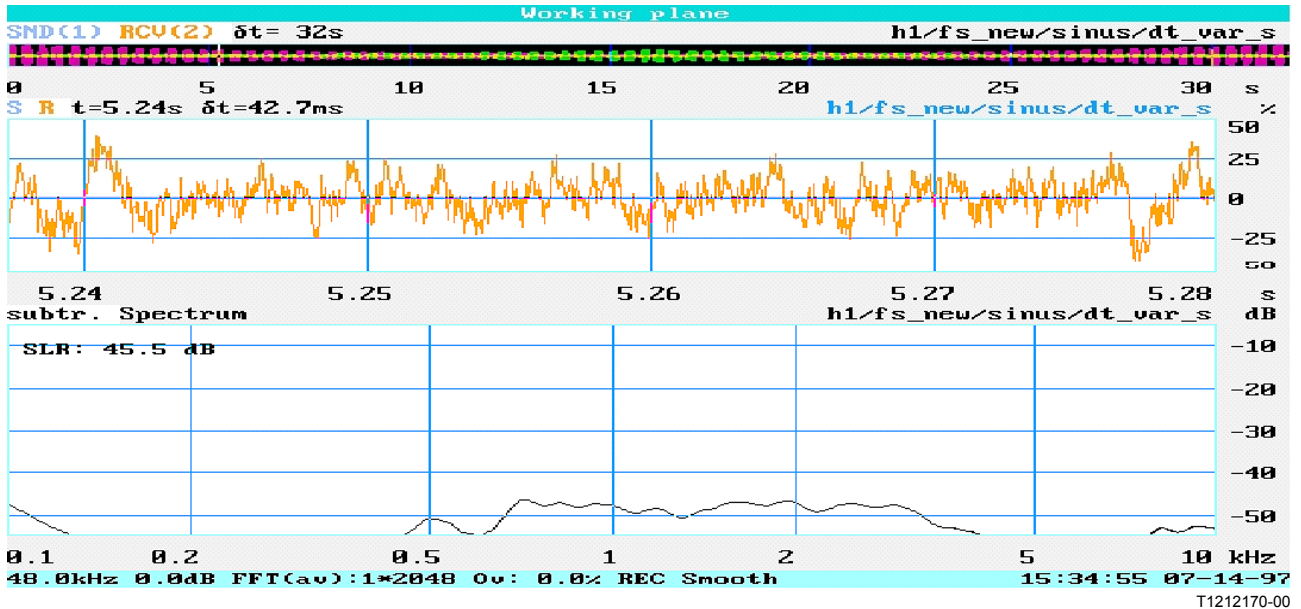
Les Figures II.3, II.4 et II.5 montrent quant à elles les résultats de mesure de la réponse en fréquence et de l'équivalent SLR pour le téléphone mains-libres à commutation de niveau dans le sens émission.



SLR = 9,5 dB

Mesure faite au début avec un niveau de signal de $-4,7$ dBPa.
 Partie supérieure de l'image: historique du signal.
 Partie inférieure: réponse en fréquence et équivalent pour la sonie.

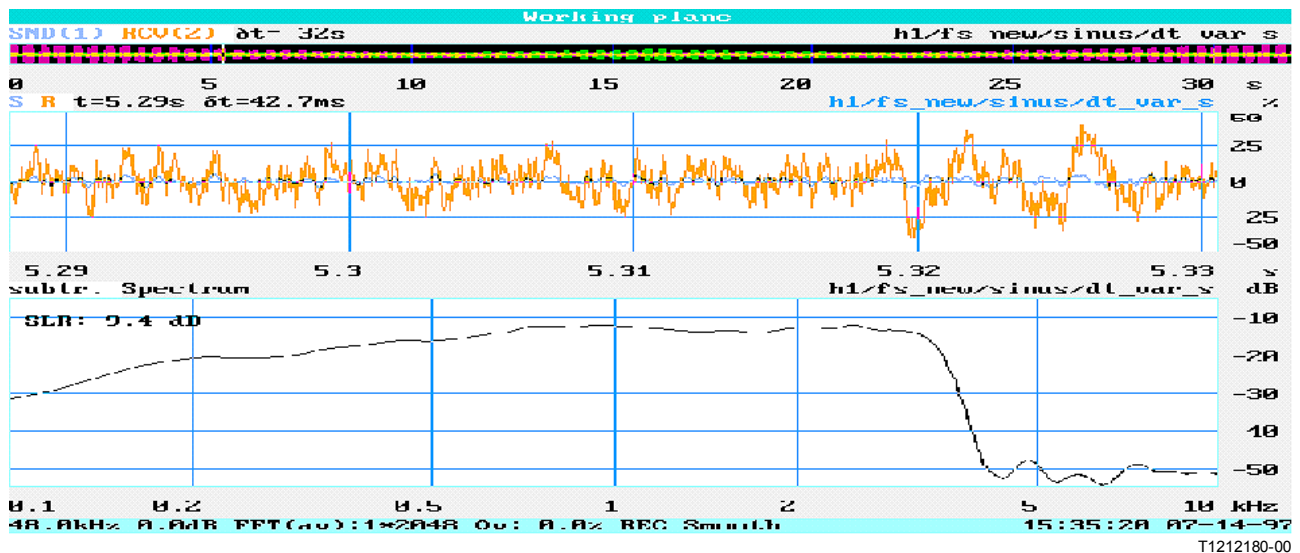
Figure II.3/P.502 – Téléphone mains-libres à commutation de niveau: caractéristiques de transfert et équivalents pour la sonie



SLR = 45,5 dB

Mesure faite après ~ 5 s, niveau du signal d'excitation de $-9,5$ dBPa, mesure avant l'activation du téléphone.
 Partie supérieure de l'image: historique du signal.
 Partie inférieure: réponse en fréquence et équivalent pour la sonie.

Figure II.4/P.502 – Téléphone mains-libres à commutation de niveau: caractéristiques de transfert et équivalents pour la sonie



SLR = 9,4 dB

Mesure faite après ~5 s, niveau du signal d'excitation de -9,5 dBPa, mesure avant l'activation du téléphone.
 Partie supérieure de l'image: historique du signal.
 Partie inférieure: réponse en fréquence et équivalent pour la sonie.

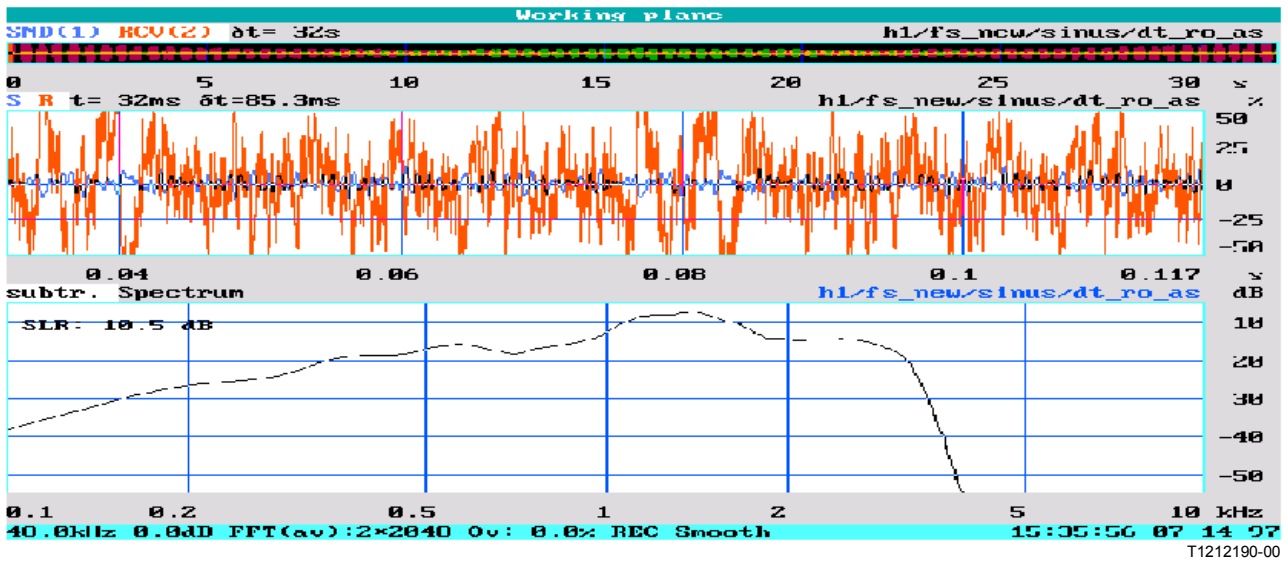
Figure II.5/P.502 – Téléphone mains-libres à commutation de niveau: caractéristiques de transfert et équivalents pour la sonie

D'après les Figures II.3 à II.5, on peut voir comment le dispositif à commutation de niveau dans le sens émission réagit sur ce type de signal. D'après l'analyse, il est clair que l'équivalent pour la sonie à l'émission au début de la séquence, lorsque le sens émission est totalement activé, est de 9,5 dB. Pendant les périodes où l'activation du système n'est pas complète, l'équivalent pour la sonie à l'émission est de 45,5 dB (voir aussi la réponse en fréquence mesurée à ce moment). On peut donc dire que l'atténuation du téléphone est d'environ 30 dB; lorsqu'on compare les réponses en fréquence, on ne constate aucune caractéristique dépendante de la fréquence.

On peut aussi mesurer le temps d'activation pour ce dispositif, en fonction du niveau à l'émission et du niveau à la réception. Par exemple, après ~5 s, il est d'environ 90 ms. Le niveau à l'émission à cet instant est de -9,5 dBPa et le niveau à la réception de -31,5 dBm.

On peut alors évaluer les caractéristiques de transfert, les équivalents pour la sonie et les temps de commutation pour d'autres combinaisons de niveau.

Les Figures II.6 à II.8 montrent le même type d'évaluation, mais cette fois-ci pour un dispositif à annulation d'écho et à commutation de niveau. Comme ce téléphone ne présente aucune commutation en période de parole simultanée, on a évalué les séquences au début, après 5 s et après 16 s.



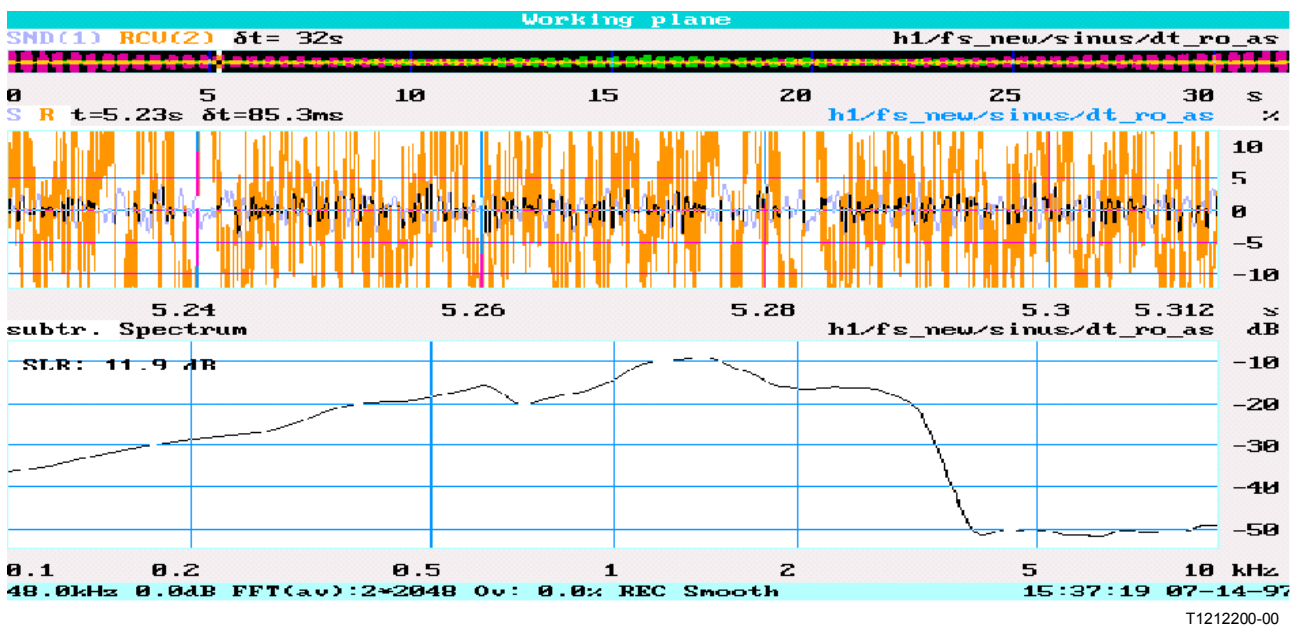
SLR = 10,5 dB

Mesure faite au début avec un niveau de signal de $-4,7$ dBPa.

Partie supérieure de l'image: historique du signal.

Partie inférieure: réponse en fréquence et équivalent pour la sonie.

Figure II.6/P.502 – Téléphone mains-libres à annulation d'écho et à commutation de niveau: caractéristiques de transfert et équivalents pour la sonie



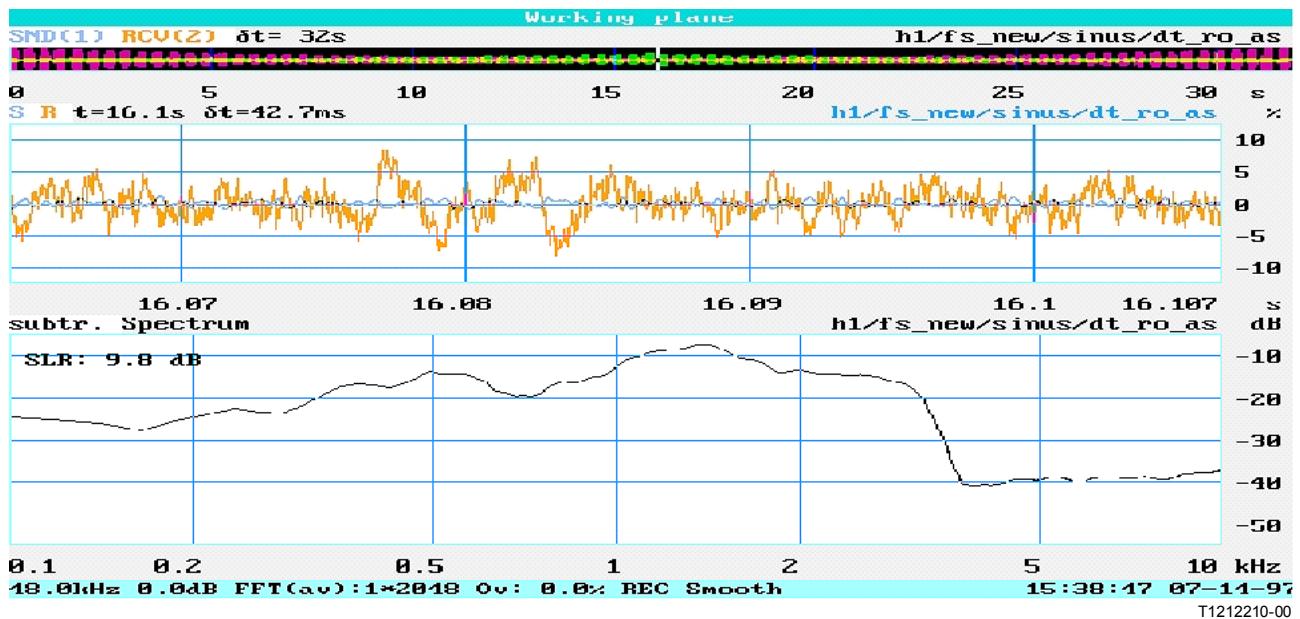
SLR = 11,9 dB

Mesure faite après 5 s, niveau du signal d'excitation de $-9,5$ dBPa, mesure avant l'activation du téléphone.

Partie supérieure de l'image: historique du signal.

Partie inférieure: réponse en fréquence et équivalent pour la sonie.

Figure II.7/P.502 – Téléphone mains-libres à annulation d'écho et à commutation de niveau: caractéristiques de transfert et équivalents pour la sonie



SLR = 9,8 dB

Mesure faite après 5 s, niveau du signal d'excitation de $-24,7$ dBPa, mesure avant l'activation du téléphone.

Partie supérieure de l'image: historique du signal.

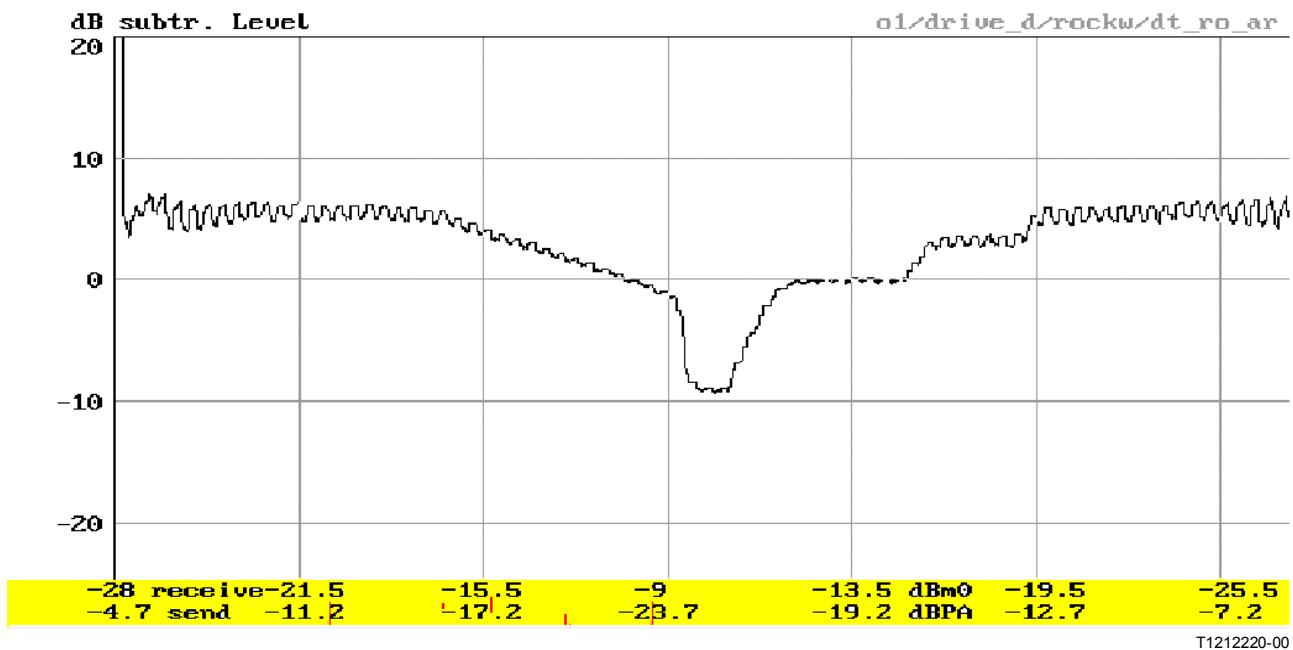
Partie inférieure: réponse en fréquence et équivalent pour la sonie.

Figure II.8/P.502 – Téléphone mains-libres à annulation d'écho et à commutation de niveau: caractéristiques de transfert et équivalents pour la sonie

L'évaluation de ce poste téléphonique montre que les caractéristiques de transfert et les équivalents pour la sonie dans toutes les conditions dans l'ensemble de la plage de niveaux allant de $-4,7$ dBPa à $-24,7$ dBPa en présence de signal de parole simultanée sont stables, aucune extinction n'est possible. Toutefois, d'après les équivalents pour la sonie et les réponses en fréquence, on peut sans conteste noter un certain type d'effets de compression-extension. L'équivalent pour la sonie n'est pas toujours le même, il présente une variation (en fonction du signal) d'environ 2 dB.

II.1.2 Variations de niveau en période de parole simultanée

La Figure II.9 donne un aperçu général du comportement d'un téléphone mains-libres en période de parole simultanée. Ici, on peut voir l'amplification du trajet de réception en fonction du niveau du signal à la réception ainsi que du niveau du signal à l'émission. La Figure 2 montre le signal de test utilisé pour cette évaluation.

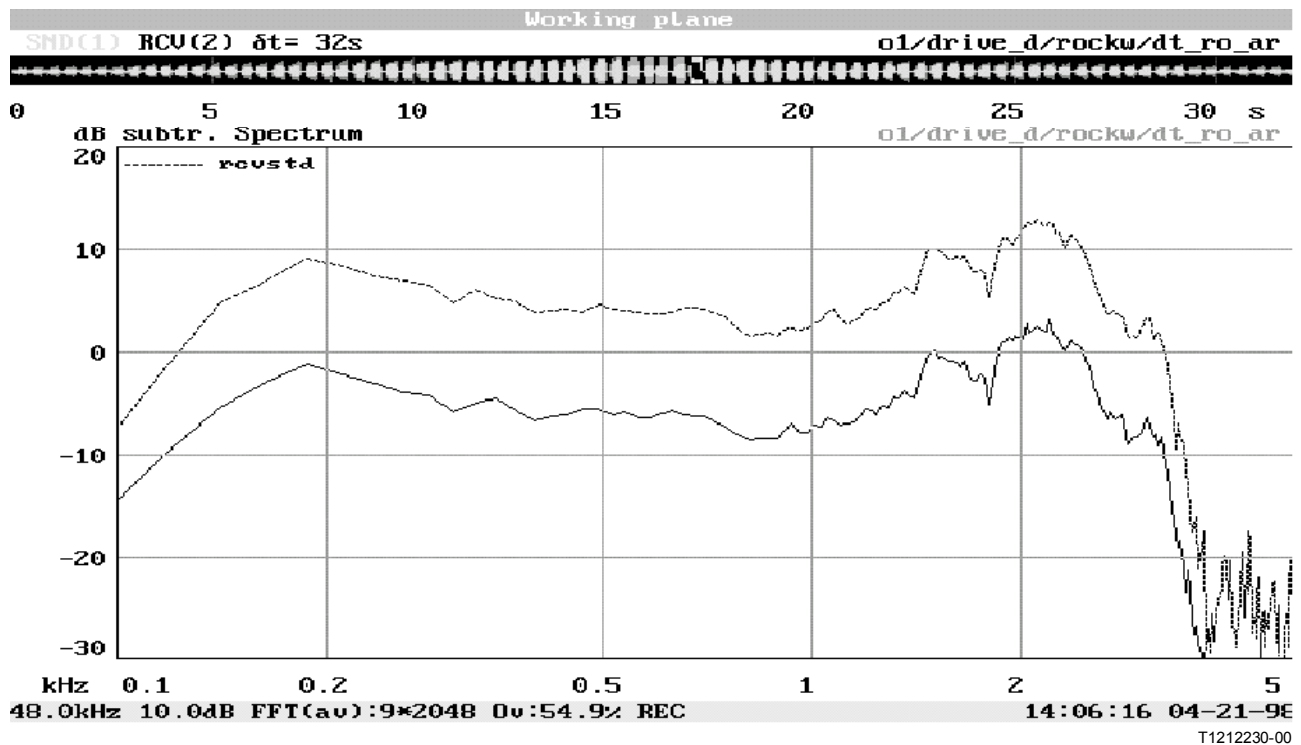


Variation de niveau:
 A la réception: -28 dBm0 ... -8 dBm0 ... -28 dBm0
 A l'émission: -4,7 dBPa ... -24,7 dBPa ... -4,7 dBPa

Figure II.9/P.502 – Variation d'amplification dans le sens réception en période de parole simultanée

Il est évident qu'une commande AGC asymétrique est active et que, pour les niveaux élevés de signal à la réception, elle aura un effet marqué sur la fluctuation du volume sonore telle qu'elle est perçue subjectivement. Lorsqu'on augmente le niveau à la réception, on constate d'abord une légère décroissance de l'amplification suivie par une atténuation brutale d'environ 10 dB. Lorsqu'on diminue à nouveau le niveau à la réception, on constate 3 pas dans l'atténuation: 10 dB, 12,5 dB et 15 dB, le pas de 10 dB aura un effet marqué sur la qualité perçue. La variation de niveau est relativement élevée (15 dB) et incorpore un changement rapide d'amplification.

La Figure II.10, qui correspond à une évaluation plus détaillée avec ce type de signal de test, représente les réponses en fréquence et les équivalents pour la sonie (pour des niveaux de signal à la réception de -28 dBm0 et -8 dBm0) qui sont mesurés dans le sens réception en période de parole simultanée, avec une pause dans le signal de parole simultanée. On a donc mesuré la réponse en fréquence et l'équivalent pour la sonie en utilisant 88 ms de signal pour l'analyse. Dans cet exemple, on peut voir la variation de niveau en fonction de la fréquence du téléphone mains-libres testé.



RLR = 6,3 dB avec:

Signal d'excitation de niveau élevé à l'émission (-4,7 dBPa) et signal d'excitation de niveau faible à la réception (-28 dBm0).

RLR = 19,5 dB avec:

Signal d'excitation de niveau faible à l'émission (-24,7 dBPa) et signal d'excitation de niveau élevé à la réception (-8 dBm0).

Figure II.10/P.502 – Réponses en fréquence et équivalents pour la sonie dans le sens réception en période de parole simultanée

II.1.3 Commutation en période de parole simultanée

La Figure II.11 donne un exemple déduit du sens émission. Dans ce sens, pour des niveaux à l'émission inférieurs à -10,7 dBPa, on constate une commutation de niveau en période de parole simultanée. Pour évaluer l'incidence de cette commutation sur le plan subjectif, il faut déterminer les constantes de temps et le gain de commutation ainsi que la variation de ces paramètres en fonction du niveau. La Figure II.11 montre le temps de commutation et le gain. 35 ms après la parole simultanée, l'amplification augmente d'environ 10 dB. Le temps de commutation est d'environ 10 ms.

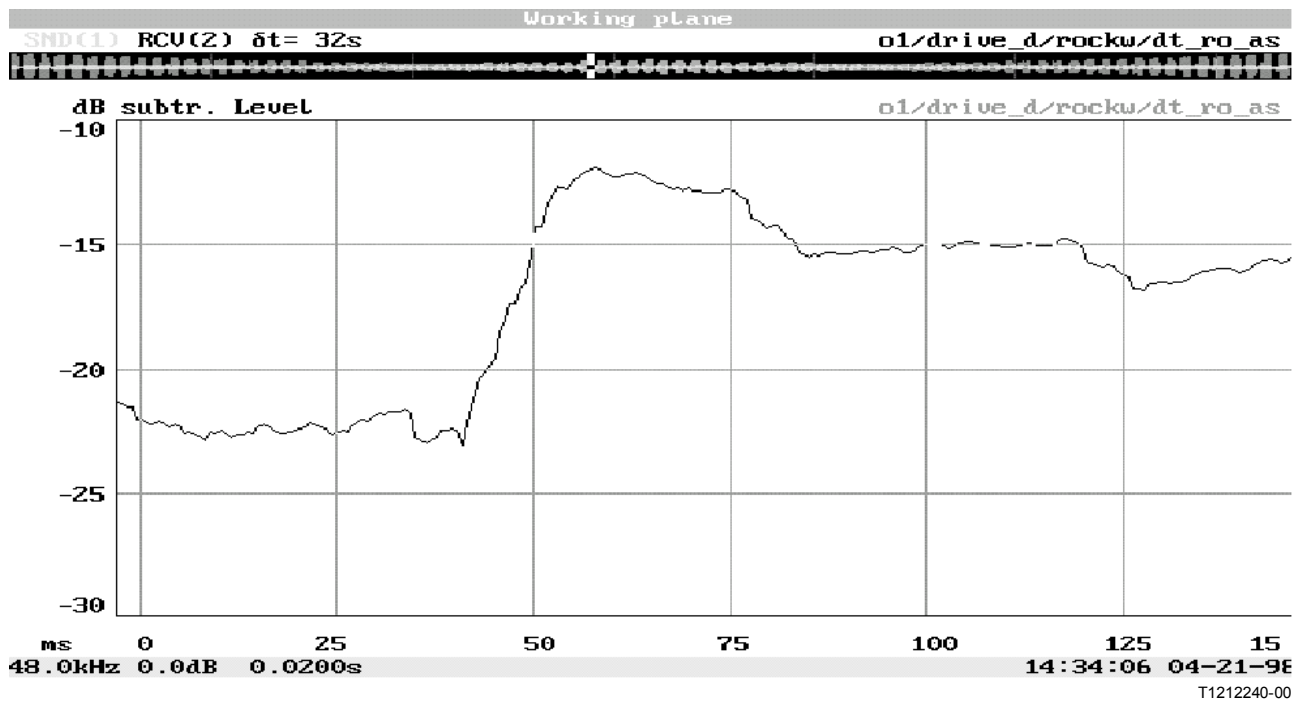


Figure II.11/P.502 – Temps de commutation et variation de niveau en période de parole simultanée

SÉRIES DES RECOMMANDATIONS UIT-T

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