



UNION INTERNATIONALE DES TÉLÉCOMMUNICATIONS

Complément au Manuel de téléphonométrie



UIT-T

SECTEUR DE LA
NORMALISATION DES
TÉLÉCOMMUNICATIONS
DE L'UIT

1999



UNION INTERNATIONALE DES TÉLÉCOMMUNICATIONS

Complément au Manuel de téléphonométrie

UIT-T
SECTEUR DE LA
NORMALISATION DES
TÉLÉCOMMUNICATIONS
DE L'UIT

1999

© UIT 1999

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.

COMPLÉMENT AU MANUEL DE TÉLÉPHONOMÉTRIE

TABLE DES MATIÈRES

	<i>Page</i>
Complément au paragraphe 2.3.6 du Manuel de téléphonométrie.....	1
Complément au paragraphe 3.2.5 du Manuel de téléphonométrie.....	7
Complément au paragraphe 4.8 du Manuel de téléphonométrie.....	9

QUELQUES CARACTÉRISTIQUES DE L'EFFET LOCAL

1 Introduction

Depuis plusieurs années, l'effet local est à l'étude et d'importantes conclusions ont été formulées en ce qui concerne, du point de vue de l'abonné, son rôle en tant que locuteur et en tant qu'auditeur. Ces conclusions ont trait à l'influence de l'effet local pour un abonné quand il entend sa voix, à la modification corrélative du niveau de la parole émise par cet abonné et aux caractéristiques de l'effet local quand l'abonné écoute dans des conditions de bruit de salle moyen à fort. Ces caractéristiques sont résumées sur les Figures 1 à 3.

La relation entre l'effet local pour le locuteur et l'effet local pour l'auditeur pour un poste téléphonique donné dépend principalement de deux facteurs:

- a) la géométrie du combiné;
- b) l'existence sur le trajet d'effet local de caractéristiques de gain ou d'affaiblissement non linéaire.

Le paragraphe 4 offre quelques indications à l'intention des concepteurs d'appareils téléphoniques.

Certaines informations sont également présentées au sujet de la fréquence toujours plus grande d'un écho à bref retard pour le locuteur, qui peut être perçu par ce dernier comme un effet local désagréable.

2 Effet local pour le locuteur

La Figure 1 montre qu'il existe un intervalle préféré d'effet local lorsque l'abonné parle dans une ambiance peu bruyante et que la différence entre un effet local gênant et un effet local trop bas est de l'ordre de 20 dB. (Ces résultats ont été fournis par des essais de locution uniquement; ils demandent à être confirmés par des essais de conversation.) L'intervalle préféré se situe entre des affaiblissements d'effet local par la méthode de masquage (STMR, *sidetone masking rating*) de 7 à 12 dB (Recommandation P.76 [1]).

L'intervalle acceptable est plus étendu; il est compris entre des affaiblissements d'effet local par la méthode de masquage (STMR) de 1 à 17 dB (mais il faut noter qu'une augmentation du STMR au-delà de 17 dB n'affectera probablement que le niveau de la parole émise, et encore marginalement). Cet intervalle correspond à la différence entre les deux courbes au niveau de 50% d'appréciations. Il n'est pas proposé que la valeur 17 dB soit considérée comme une valeur maximale. Néanmoins, pour un niveau de STMR au-dessus de 20 dB, la communication est complètement assourdie.

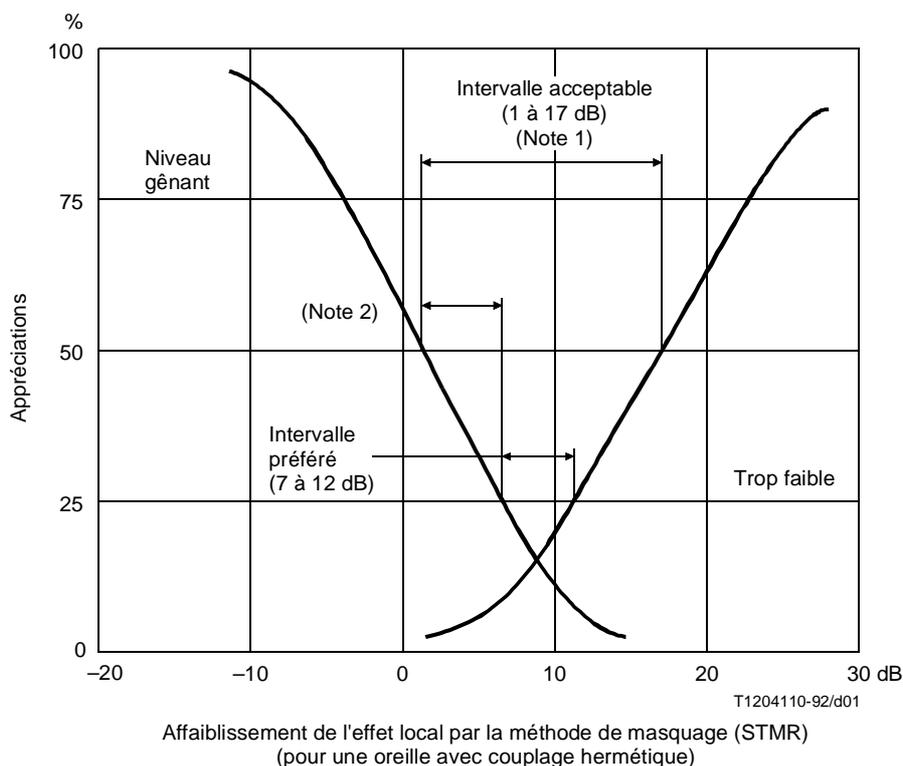
La Figure 2 montre comment le niveau des sons vocaux varie en fonction de l'effet local. Ces résultats ont été fournis par des essais de conversation, dans une connexion dont l'équivalent était très proche de la valeur préférée. S'agissant de connexions téléphoniques donnant lieu à un équivalent global pour la sonie (OLR) situé dans l'intervalle préféré, les valeurs de STMR peuvent être situées dans l'intervalle préféré de STMR mentionné plus haut. Cependant, pour les connexions subissant un fort affaiblissement, la valeur de STMR doit approcher, voire dépasser, 12 dB pour inciter la personne à parler plus fort. Pour les connexions subissant un affaiblissement peu élevé, on peut permettre dans certains cas une valeur de STMR inférieure à 7 dB, mais elle ne devra que rarement descendre jusqu'à 1 dB, par exemple, s'agissant d'appareils téléphoniques avec régulation du volume de réception. La Recommandation G.121 [2] interprète ces résultats pour les besoins de planification.

La tension des sons vocaux dépend également du niveau du bruit de salle. En ce qui concerne les postes téléphoniques modernes à transducteurs linéaires, il peut être souhaitable que la valeur de STMR soit comprise entre 10 et 15 dB si ces postes téléphoniques doivent être utilisés dans un environnement bruyant.

3 Effet local pour l'auditeur

Un fort bruit de salle dans l'environnement perturbe de deux façons la réception vocale:

- i) par le bruit recueilli par le microphone du combiné et transmis au récepteur du combiné par l'intermédiaire du trajet d'effet local électrique;
- ii) par le bruit fuyant dans le pavillon du récepteur du combiné.



NOTE 1 – Les conditions de conversation détermineront la portion de cet intervalle acceptable pour une communication donnée.

NOTE 2 – Cette portion (1 à 7 dB) de l'intervalle acceptable doit être prise en considération avec prudence, par exemple s'agissant de connexions à affaiblissement réduit (voir la Recommandation G.121) ou bénéficiant d'une régulation du volume de réception.

Figure 1 – Courbes indiquant les niveaux d'effet local gênants et trop faibles pour le locuteur, ainsi que l'intervalle préféré

Les travaux ont montré qu'aux basses fréquences, le trajet de fuite du pavillon de l'écouteur domine le trajet d'effet local électrique à peu près de la même manière que le signal d'effet local humain domine l'effet local du locuteur. La pondération appliquée au calcul de la sonie de STMR est par conséquent applicable, l'affaiblissement d'effet local pour la personne qui écoute (LSTR, *listener sidetone rating*) (Recommandation P.76) a été développé et on a mis au point l'affaiblissement de l'effet local du bruit de salle (voir le paragraphe 9/P.64 [3]) dans la méthode de calcul du STMR (voir la Recommandation P.79 [4]).

Les résultats des essais subjectifs du MOS en fonction du LSTR (reposant sur l'emploi d'un barème de notes d'opinion de 0 à 10) sont donnés dans la Figure 3. Dans chaque cas, on a utilisé, pour calculer le LSTR, Δ_{Sm} (voir les Recommandations P.10 [5], P.64, P.79 et le § 3.3.17, Partie C du Manuel de téléphonométrie [6]) pour convertir les efficacités d'effet local S_{meST} à S_{RNST} avant de calculer le LSTR, ou appliqué comme une correction pondérée au STMR, comme le décrit A.4.3.3/G.111 [7]. Les niveaux de bruit de salle étaient comparables à 55-59 dBA.

Sur la base de ces résultats, la Recommandation G.121 recommande d'atteindre si possible une valeur de LSTR de 13 dB.

La valeur 13 dB résulte d'un LSTR de 10 dB (qui peut être considéré comme une valeur minimale), quand l'augmentation du LSTR n'a pas permis d'améliorer encore la note moyenne d'opinion (voir la Figure 3), à quoi on ajoute une tolérance de 3 dB pour tenir compte du fait que dans certains bureaux le bruit de salle peut dépasser les valeurs servant à ces essais. D'autres essais ont aussi montré qu'une valeur plus élevée serait peut-être préférable.

La valeur satisfaisante pour une connexion téléphonique donnée dépend de différents facteurs: niveau de bruit de salle, OLR de la connexion, niveau utilisé pour la parole, etc. En particulier, les postes modernes à transducteurs linéaires sont plus sensibles au bruit de salle de fond. Dans ce cas, il peut être souhaitable que la valeur de LSTR soit supérieure à 15 dB.

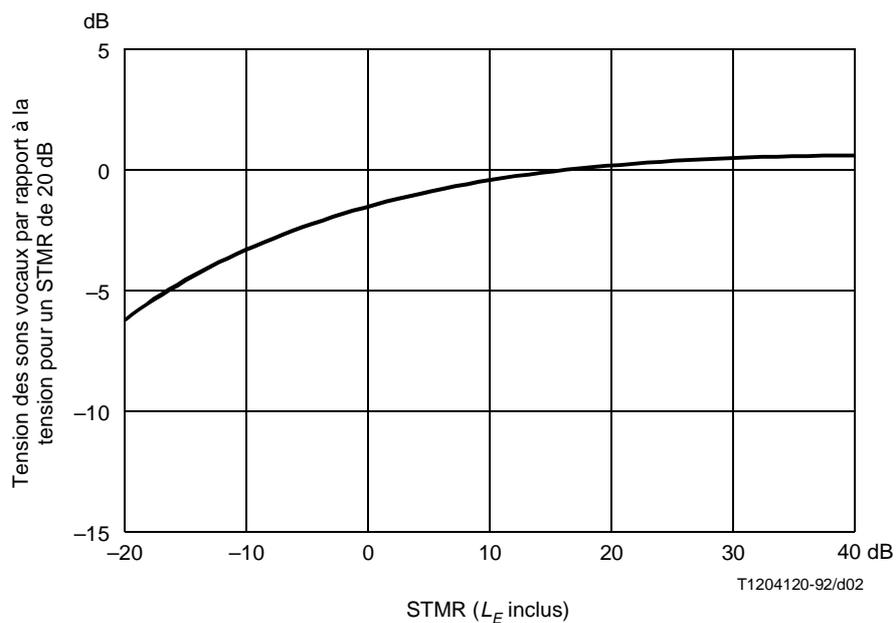


Figure 2 – Tension des sons vocaux en fonction du STMR

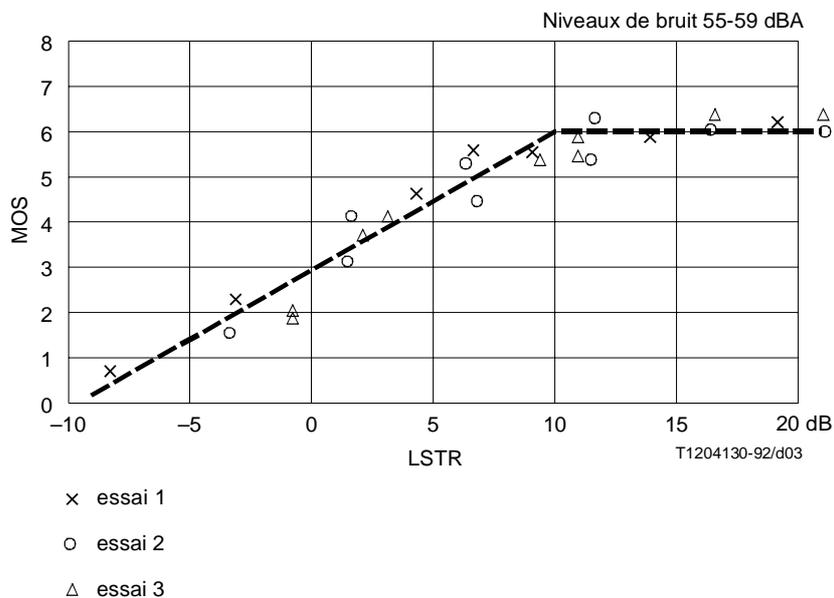


Figure 3 – MOS en fonction du LSTR calculé à partir des résultats d'essais différents

4 Relation entre l'effet local pour le locuteur et l'effet local pour l'auditeur

4.1 Postes téléphoniques ayant des caractéristiques d'effet local linéaire

Dans le cas d'un poste téléphonique ayant des caractéristiques d'affaiblissement ou de gain linéaire sur le trajet d'effet local, la relation entre les niveaux d'effet local pour le locuteur, d'une part, et l'auditeur, d'autre part, dépend de la géométrie du combiné. Cette dernière semble importante sur deux points: la distance qui sépare la bouche du locuteur et l'accès de l'émetteur, d'une part, et la taille de l'obstacle créé par l'extrémité de l'émetteur, d'autre part.

Pour des signaux de parole à l'entrée, un combiné, dont l'extrémité de l'émetteur est grande et positionnée près de la bouche, subit une pression acoustique plus importante à l'accès de son émetteur qu'un combiné dont l'extrémité de l'émetteur est positionnée plus loin de la bouche (effet de distance) ou qu'un combiné dont l'extrémité de l'émetteur est petite (effet d'obstacle). Cependant, en présence d'un champ diffus de bruit de salle, la pression acoustique à l'accès de l'émetteur est indépendante de la taille et de la forme du combiné. Ainsi, si le niveau STMR est le même pour les deux combinés, le gain électrique sur le trajet d'effet local sera moindre pour celui dont l'extrémité de l'émetteur est large et positionnée près de la bouche, ce qui aboutira à une valeur LSTR plus élevée.

Il a été montré que la différence des niveaux LSTR et STMR pour un échantillon de 26 appareils téléphoniques linéaires correspond au logarithme de la distance entre l'accès de l'émetteur (centre de l'ouverture externe du microphone à la surface du combiné) et le centre de l'anneau de garde de la bouche artificielle lorsque le combiné est placé dans la position d'essai LRGP, position de l'anneau de garde pour l'évaluation de l'équivalent pour la sonie (voir la Recommandation P.64). Cette différence est définie par la relation empirique suivante:

$$LSTR - STMR = 33 - 20 \log(d)$$

dans laquelle la distance d de l'accès de l'émetteur au centre de l'anneau de garde est exprimée en mm. De petites perturbations, de l'ordre de ± 1 dB, peuvent se produire selon la taille de l'obstacle représenté par l'extrémité de l'émetteur du combiné.

NOTE – Cette relation est fondée sur des mesures concernant des postes téléphoniques dotés de combinés relativement classiques. Elle peut ne pas être applicable aux combinés dont les formes sont extrêmes ou aux appareils téléphoniques d'opérateur dont l'accès à l'émetteur est situé derrière la position équivalente des lèvres.

4.2 Postes téléphoniques ayant des caractéristiques d'effet local non linéaire

Des caractéristiques de gain ou d'affaiblissement non linéaire peuvent être utilisées sur le trajet d'effet local électrique afin d'augmenter la différence LSTR – STMR. Les émetteurs au carbone, par exemple, sont souvent moins sensibles à l'entrée à des niveaux de bruit de salle plus faibles qu'ils ne le sont à des niveaux de parole plus élevés à l'entrée. On peut ajouter une telle caractéristique à des téléphones ayant des microphones linéaires en utilisant divers circuits de gain non linéaire.

Si la même fonction de gain non linéaire est utilisée à la fois sur le trajet d'effet local et sur celui d'émission du téléphone, on peut alors évaluer approximativement la différence LSTR – STMR en mesurant la différence entre les efficacités à l'émission due à des données de bruit de salle et de signaux de parole, DELSM (voir la Recommandation P.64). Une différence STMR peut alors être calculée selon la méthode donnée dans l'Annexe A/G.111. Cependant, si les trajets d'émission et d'effet local n'ont pas les mêmes caractéristiques de gain non linéaire (par exemple, la commande automatique de gain de circuits sur le trajet de réception qui influe sur l'effet local), la méthode DELSM donnera des résultats erronés. Dans ce cas, les valeurs de LSTR et STMR doivent être mesurées directement.

5 Echo à bref retard pour le locuteur, perçu comme un effet local

L'écho pour le locuteur peut avoir une influence défavorable sur la qualité de transmission même si le retard, de quelques millisecondes, n'est pas assez long pour être perçu comme un signal d'écho distinct de l'effet local. De tels échos peuvent, par exemple, se produire à cause de réflexions issues de l'accès au circuit de jonction analogique d'un commutateur numérique privé, ou lors de communications analogiques locales par l'intermédiaire d'un commutateur numérique. En cas de mauvaise adaptation du circuit différentiel qui fait passer le commutateur numérique privé ou public de 4 fils aux 2 fils d'un circuit analogique, un certain effet de réflexion se produira. Etant donné les durées de traitement numérique mises en jeu, ces signaux d'écho pour le locuteur auront un retard de quelques millisecondes. Aux faibles niveaux, cet écho est compensé par l'effet local; mais lorsqu'il augmente, il entre en interaction avec l'effet local de manière désagréable (réverbération de l'effet local, effet de tonneau, etc.).

L'effet objectivement mesurable de l'écho à bref retard pour le locuteur est l'injection d'ondulations dans la courbe de réponse de l'effet local. Le signal réfléchi de l'écho pour le locuteur est ajouté au signal d'effet local direct avec un déphasage qui fait augmenter ce signal à certaines fréquences et le fait décroître à d'autres. L'espacement entre ces ondulations est inversement proportionnel au retard. Lorsque le signal réfléchi de l'écho pour le locuteur est petit par rapport à l'effet local direct, les ondulations sont faibles. Au fur et à mesure de l'augmentation en amplitude du signal d'écho pour le locuteur, les ondulations augmentent en amplitude jusqu'à atteindre des crêtes de 6 dB au-dessus du signal en phase, avec des creux très profonds en raison de l'annulation des signaux en opposition de phase presque exacte. A des niveaux encore plus élevés de l'écho pour le locuteur (ou à des niveaux plus faibles de l'effet local), l'amplitude des ondulations décroît de nouveau mais le signal prédominant devient l'écho réverbéré vers le locuteur.

Les utilisateurs ont une perception différente entre, d'une part, un écho à bref retard pour le locuteur combiné à l'effet local et, d'autre part, un niveau équivalent d'effet local proprement dit, même s'ils ne sont pas toujours en mesure de détecter la présence d'un signal d'écho distinct. Une simple mesure de l'effet local telle que le STMR (affaiblissement d'effet local par la méthode de masquage) ne convient pas pour décrire l'effet du signal combiné. L'écho pour le locuteur, même s'il est affecté de très brefs retards, doit impérativement être traité comme une dégradation distincte de la qualité de transmission. La Recommandation P.11 [8] et la référence [9] donnent des indications sur la manière dont aussi bien l'effet local que l'écho pour le locuteur peuvent être pris en considération lors de la prédiction de la qualité d'une chaîne de connexion téléphonique; mais ce sujet reste à l'étude.

Références normatives

- [1] Recommandation P.76 du CCITT (1988), *Détermination des équivalents pour la sonie; principes fondamentaux.*
- [2] Recommandation UIT-T G.121 (1993), *Equivalents pour la sonie des systèmes nationaux.*
- [3] Recommandation UIT-T P.64 (1997), *Détermination des caractéristiques d'efficacité en fonction de la fréquence des systèmes téléphoniques locaux.*
- [4] Recommandation UIT-T P.79 (1993), *Calcul des équivalents pour la sonie des postes téléphoniques.*
- [5] Recommandation UIT-T P.10 (1998), *Vocabulaire des termes relatifs à la qualité de la transmission téléphonique et aux appareils téléphoniques.*
- [6] *Manuel de téléphonométrie*, UIT, Genève, 1993.
- [7] Recommandation UIT-T G.111 (1993), *Equivalents pour la sonie dans une connexion internationale.*
- [8] Recommandation UIT-T P.11 (1993), *Effet des dégradations de la transmission.*
- [9] Complément au paragraphe 3.2.5 du Manuel de téléphonométrie: conversation vocale artificielle. UIT, Genève, 1999.

CONVERSATION VOCALE ARTIFICIELLE

La voix artificielle décrite dans la Recommandation P.50 est utilisée comme signal d'essai unidirectionnel pour l'évaluation de la qualité de transmission d'équipements tels que les codecs vocaux à faible débit. Il est nécessaire d'élargir le domaine d'application de ces signaux artificiels afin d'évaluer la qualité d'équipements commandés par des signaux vocaux comme les annuleurs d'écho et les commutateurs vocaux intégrés à des terminaux téléphoniques à haut-parleur ou à des concentrateurs numériques de la parole (DSI, *digital speech interpolation*). Les signaux artificiels doivent, pour cela, simuler des conversations entre humains. Ils doivent donc contenir, non seulement des périodes «d'émission de paroles», exprimées sous la forme de voix artificielles dans la Recommandation P.50, mais aussi des périodes de «pause». Il est également nécessaire que les signaux artificiels soient transmis dans les deux sens afin de simuler un environnement «de locuteur et d'auditeur».

Un signal qui fait alterner sur l'axe des temps des séquences arbitraires de voix artificielle selon la Recommandation P.50 et de silence (séquences nulles) ne convient pas, car la conversation humaine possède des caractéristiques propres. Par exemple, les êtres humains ne peuvent pas maintenir une longue période de locution sans pause et il arrive parfois que deux personnes parlent en même temps pendant une longue période.

Il convient donc de simuler, dans les signaux artificiels, certaines caractéristiques temporelles statistiques d'une conversation réelle. Un signal artificiel qui répond à cette condition est appelé *conversation vocale artificielle*.

La Figure 1 montre des courbes temporelles de puissance de deux interlocuteurs en conversation. Les caractéristiques statistiques de la conversation vocale artificielle correspondent bien à celles d'une conversation réelle en termes de fréquences d'apparition, de valeurs moyennes et de durées cumulées calculées sur les périodes de locution, de pause, de double parole et de silence mutuel.

La Figure 2 montre un modèle de transition d'états pour la génération de conversations vocales artificielles. Une conversation alterne entre quatre états: monologue (un des deux interlocuteurs parle et l'autre reste silencieux et vice versa), parole simultanée et silence simultané; s'y ajoute un modèle de transition entre ces quatre états. Après une durée T dans un état donné, le passage à un autre état s'effectue selon la probabilité de transition p_i . Les distributions cumulées de durée T dans chaque état sont considérées comme étant exponentielles et la durée T de chaque état varie en fonction d'une variable aléatoire x . Dans les périodes de silence, des séquences nulles sont émises. Dans les périodes de locution, des séquences de voix artificielle selon la Recommandation P.50 sont émises.

Les paramètres à optimiser dans la Figure 2 sont les suivants: la durée moyenne estimée en monologue, en parole simultanée et en silence; la probabilité de transition p_1 . Les probabilités de transition p_2 et p_3 sont fixées à 50%. Ces paramètres ont été optimisés de façon à minimiser l'écart type des grandeurs suivantes de la conversation artificielle, par rapport à une conversation naturelle: durée et fréquence moyenne de locution, de pause, de parole simultanée et de silence simultané. Comme valeurs cibles pour l'optimisation des paramètres, on a utilisé les valeurs moyennes mesurées sur des conversations japonaises, anglo-américaines et italiennes. Les valeurs cibles pour l'optimisation des paramètres et les valeurs paramétriques optimales sont spécifiées dans la Recommandation P.59.

Les caractéristiques statistiques temporelles de la conversation vocale artificielle tendent à se stabiliser au fur et à mesure que la durée totale des signaux augmente. La conversation vocale artificielle présente une durée de convergence des caractéristiques à long terme de 10 min, ce qui correspond aux caractéristiques à long terme d'une conversation humaine d'environ 10 h. Cela montre l'efficacité, en termes de mesures, de l'utilisation de la conversation vocale artificielle au lieu d'une conversation humaine.

Normalement, la conversation vocale artificielle n'est pas générée par la mise en œuvre en temps réel des processus de la Recommandation P.59: ces signaux sont habituellement préenregistrés soit sur cédérom soit sur support magnétique ou optique. Dans les deux cas, des moyens d'étalonnage sont habituellement prévus. Il s'agit de tonalités traînantes qui ont une relation connue avec le niveau de la conversation vocale artificielle.

Procédure

La conversation vocale artificielle ne nécessite aucun étalonnage. Il suffit d'en régler le niveau au moyen de tonalités traînantes enregistrées sur support numérique (ROM, CD, DAT, MO, MIC vidéo) ou analogique (bandes magnétiques).

Les signaux de conversation vocale artificielle peuvent être utilisés soit en tant que signaux électriques ou comme entrée dans la bouche artificielle pour obtenir un signal acoustique de conversation artificielle. Dans ce dernier cas, la bouche artificielle doit être équipée d'un égaliseur vérifié par mesure, au point MRP, du spectre moyen du signal acoustique de conversation artificielle.

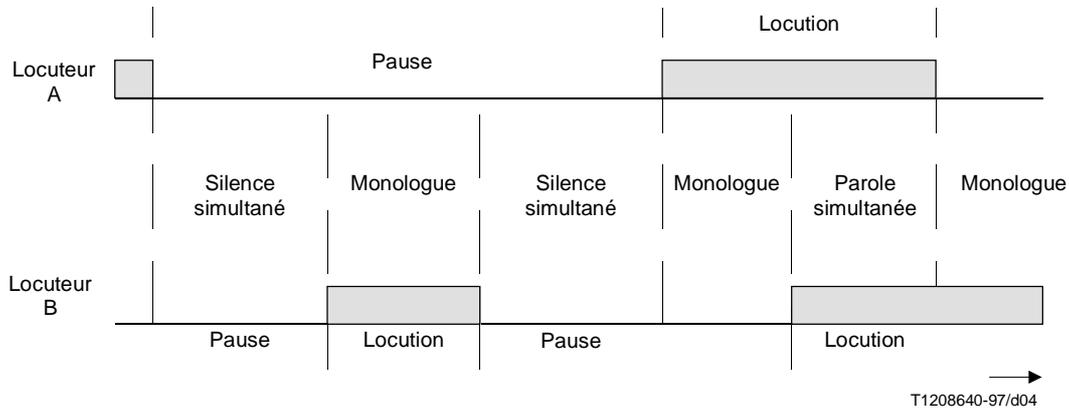


Figure 1 – Structure des courbes temporelles de puissance en conversation

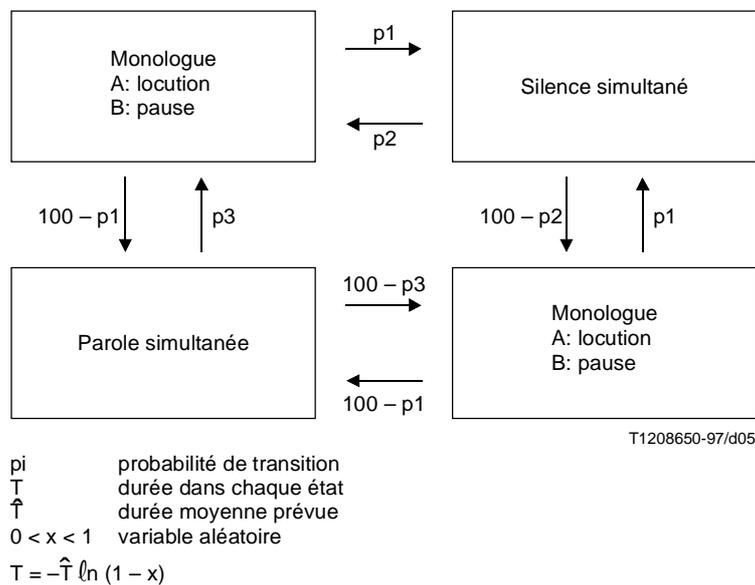


Figure 2 – Modèle de transition d'états pour la conversation

MODÈLES MATHÉMATIQUES D'ESTIMATION DES AVIS DES CLIENTS SUR LA QUALITÉ DES COMMUNICATIONS VOCALES DANS LES RÉSEAUX TÉLÉPHONIQUES

1 Introduction

Il importe d'arriver à une probabilité élevée que les utilisateurs bénéficient d'une qualité de transmission adéquate en téléphonie vocale. Les méthodes traditionnelles de planification de la transmission – consistant à affecter des étendues admissibles aux divers paramètres de transmission – sont maintenant jugées imparfaitement adaptées aux réseaux modernes. Entre autres raisons, on peut citer l'introduction de nouvelles techniques, par exemple les codecs à faible débit, ainsi que la déréglementation continue du marché des télécommunications.

Dans ces circonstances mouvantes, les opérateurs de réseau ont besoin de directives et d'une méthode satisfaisante pour évaluer la qualité de transmission des connexions dès le stade de la planification et pour prédire la satisfaction des usagers. Ces méthodes doivent se fonder sur des valeurs calculées et planifiées, car il n'est pas possible d'effectuer des mesures pour contrôler la qualité de tous les itinéraires possibles dans des réseaux réels ou simulés.

Il en a résulté un renouveau d'intérêt pour les évaluations d'opinion au moyen de modèles mathématiques capables de traiter les effets combinés de différents types de dégradation de transmission. Une brève étude sera présentée ci-dessous au sujet des modèles qui ont été utilisés dans le domaine technique de l'UIT-T, à titre d'aide pour la planification des transmissions. Avant de présenter les divers modèles mathématiques, on analysera le concept d'«avis de la clientèle» afin d'éviter tout malentendu concernant les essais subjectifs et les études de marché.

2 «Avis de la clientèle», base des modèles mathématiques

La phase critique du développement d'un modèle mathématique consiste à obtenir les matières de base sur les avis de la clientèle au sujet de la qualité de communication vocale dans le service téléphonique, en particulier dans l'étendue utile des paramètres de transmission mis en jeu. Plusieurs difficultés se posent dans ce processus.

Un problème général est que les avis quant à la qualité de la parole varient d'emblée selon les personnes ainsi qu'au fur et à mesure du temps et selon le contexte de communication environnant.

La façon dont les questions relatives à la qualité sont posées aura également une influence sur le contenu de la réponse. Un exemple classique est celui des niveaux d'écoute. Des expériences conçues pour déterminer «le niveau d'écoute le plus confortable» donnent un résultat inférieur de 3 dB par rapport à la question «à quel niveau la qualité de la parole est-elle maximale?».

Une méthode directe pour obtenir les avis des clients consiste évidemment à *conduire des études de marché* dans des réseaux réels. Au cours d'une communication téléphonique, les clients n'évaluent normalement pas consciemment la qualité de transmission des paroles mais considèrent un certain niveau comme allant de soi, contrairement à ce qui se passe lors d'une écoute musicale sur un système sonore à haute fidélité. (Ce n'est en général que si la qualité téléphonique tombe souvent au-dessous d'un certain seuil que les usagers réagissent, sans doute de manière assez véhémente, par des réclamations auprès de l'opérateur du réseau.) Théoriquement, il faudrait interroger un client immédiatement après la fin d'une communication dont on connaît plus ou moins entièrement les paramètres de transmission. De telles recherches ont été faites et continuent à être faites, parfois avec introduction délibérée de dégradations de transmission. Ce processus est assez laborieux et coûteux, mais il devrait fournir les données les plus fiables.

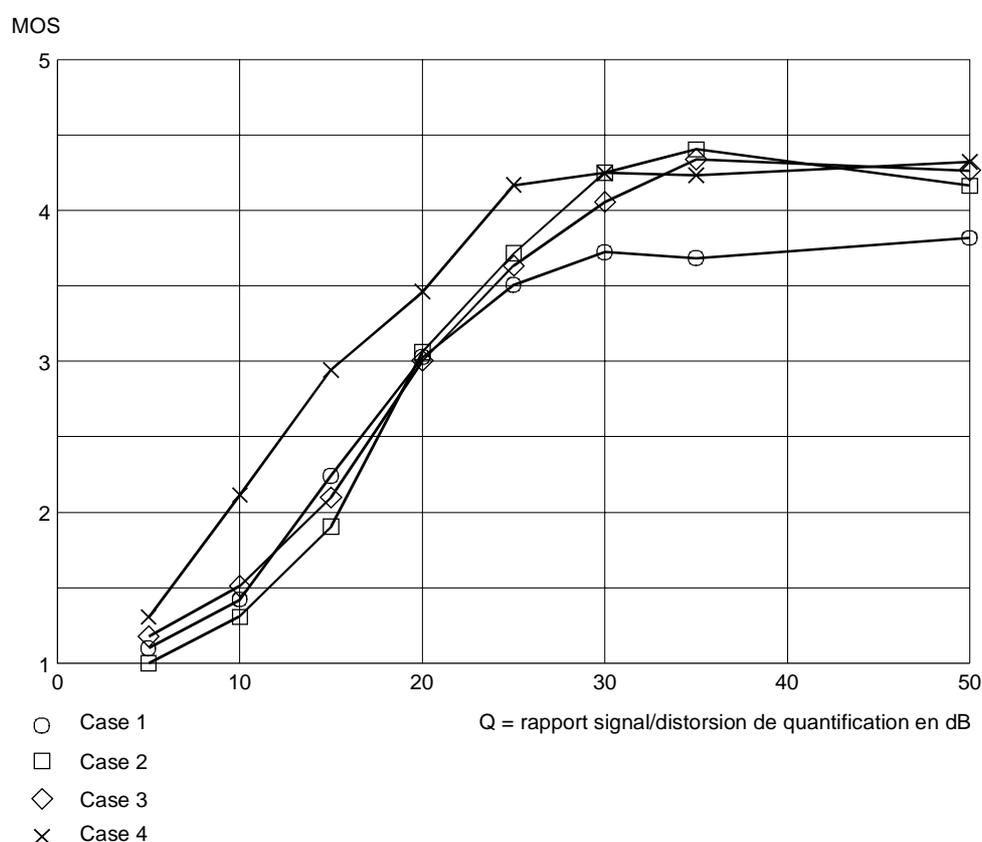
Lors de l'évaluation subjective d'une dégradation de transmission, la première étape consiste généralement à utiliser une méthode plus économique, sous la forme *d'essais subjectifs en conditions contrôlées* de laboratoire. Ces essais sont conduits avec un ensemble limité et précis de variations des paramètres de transmission et avec un nombre limité de sujets. (L'organisation de tels essais est décrite dans les Recommandations de la série P.) Le plus souvent, le résultat est présenté sous la forme de notes moyennes d'opinion (MOS, *mean opinion scores*). De grandes précautions sont toutefois requises lors de la mise au point des conditions d'essai et de l'interprétation des résultats.

Les notes MOS d'un essai subjectif sont souvent présentées avec des «limites de confiance» statistiques. Il ne faut toutefois pas perdre de vue que ces limites ne s'appliquent qu'à cette instance expérimentale particulière. On ne doit pas considérer une note MOS, issue d'un essai particulier, comme étant une valeur absolue. Il faut au contraire la considérer comme une valeur relative. Un essai subjectif soigneusement préparé doit donc comporter des conditions de référence, c'est-à-dire que les sujets sont priés d'évaluer également des dégradations de nature connue. On normalise les notes MOS réelles au moyen de ces notes MOS de référence, de façon que l'on puisse comparer les résultats issus de différents laboratoires d'essai (autrement dit, les équipes expérimentales sont «étalonnées»). Pour produire la dégradation de référence, on utilise souvent un appareil de référence à bruit modulé (MNRU, *modulated noise reference unit*): c'est devenu la procédure normale lors de l'évaluation par essais subjectifs de codecs à faible débit. (Une source d'incertitude est que la «dégradation de référence» est souvent jugée assez différente de la «dégradation d'essai».)

Par exemple, la Figure 1 montre quatre courbes de notes MOS de référence obtenues par appareil MNRU dans quatre laboratoires d'essai différents, lors d'un projet commun visant à évaluer un certain codec LD-CELP à faible débit (voir la Recommandation G.729). Comme on peut le voir, la dispersion des valeurs MOS atteint presque 1,0 unité, ce qui révèle la nécessité de normaliser les valeurs MOS mesurées par les différents laboratoires pour le codec réel mis à l'essai.

Tout bien considéré, un modèle mathématique permettant de prédire des opinions doit être adapté selon un compromis entre différents essais subjectifs et différentes études de marché. La première étape consiste à utiliser des résultats issus d'essais subjectifs en laboratoire avec un certain degré de «jugement technique». Une deuxième étape peut consister à vérifier les résultats tirés du modèle en les comparant à ceux qui ont été recueillis lors d'études de marché effectuées en conditions de réseau réel.

La littérature décrit un certain nombre de modèles d'évaluation d'opinion. Ils seront examinés ci-dessous. Compte tenu de la variabilité des résultats d'essais subjectifs et des études, il ne devrait pas être surprenant qu'ils prédisent des résultats légèrement différents pour un même état du réseau.



T1208630-97/d06

Figure 1 – Courbes de notes MOS de référence obtenues par MNRU dans quatre laboratoires d'essais subjectifs différents

3 Modèles mathématiques déjà décrits dans des publications de l'UIT-T

3.1 Généralités

Le Supplément 3 (1993) des Recommandations de la série P décrit quatre modèles mathématiques différents pour la prédiction de la qualité de transmission à partir de mesures objectives, à savoir:

- le modèle «d'évaluation de l'indice de transmission» (contribution des Etats-Unis et du Canada);
- le modèle «CATNAP83» (de British Telecom);
- le modèle de «l'indice d'information» (contribution de la France);
- le modèle «OPINE» (de NTT).

L'Annexe A/P.11 (*Livre bleu*) décrit le modèle de l'«indice de qualité de transmission», qui est une version consolidée des quatre modèles présentés dans le Supplément 4 des Recommandations de la série P mais avec des simplifications et un nombre restreint de paramètres considérés.

Tous ces modèles mathématiques ont été élaborés avant la tendance actuelle vers la déréglementation des marchés de télécommunication, c'est-à-dire lorsque les Administrations étaient en mesure de régir leurs réseaux jusqu'aux moindres détails techniques. L'affaiblissement et le bruit étaient les principales dégradations et l'influence des codecs évolués à faible débit n'était pas prise en compte, ou ne l'était qu'avec très peu de précision.

L'Appendice I/G.101 (1996) décrit le «modèle E», qui a été mis au point par un groupe de travail interne de l'ETSI (Institut européen des normes de télécommunication). Ce modèle fait appel à certains algorithmes et concepts issus des modèles décrits dans le Supplément 3 des Recommandations de la série P, en les complétant des résultats d'essais subjectifs récents. Entre autres caractéristiques nouvelles qui ont été incluses, on peut citer les effets subjectifs des codecs à faible débit et la possibilité d'évaluer les opinions des clients en matière de qualité de la parole compte tenu de leurs attentes quant à la qualité de transmission du support de communication.

3.2 Le modèle «d'évaluation de l'indice de transmission»

Les paramètres sont les équivalents globaux pour la sonie (OR), le bruit de circuit, le bruit de salle du côté réception, l'affaiblissement d'effet local par la méthode de masquage du côté du locuteur (STMT), les unités de distorsion de quantification (qdu), les fréquences de coupure, la distorsion d'affaiblissement, l'écho pour l'auditeur, l'écho pour le locuteur.

Les résultats des calculs pour une connexion sont présentés sous la forme «d'indices de transmission» (R) qui peuvent être convertis en pourcentages estimés de clients qui trouvent la connexion «au moins bonne» (GOB, *good or better*) ou «au mieux médiocre» (POW, *poor or worse*).

Les observations suivantes, figurant dans la description du modèle, sont particulièrement intéressantes: «Une des principales raisons de l'utilisation de cette échelle (d'appréciation de l'indice de transmission) réside dans le fait qu'il a été reconnu que divers facteurs peuvent influencer sur les résultats des essais subjectifs: par exemple le groupe de sujets, le type d'essai et la gamme des conditions concernant un essai donné. On a constaté que ces facteurs peuvent faire varier à la fois les notes d'opinion moyennes pour une condition donnée et l'écart type. Il est donc difficile d'établir une relation unique entre une condition de transmission donnée et une opinion subjective sous la forme d'une note d'opinion moyenne ou d'un pourcentage d'appréciation «bon» ou «excellent». L'utilisation d'une échelle d'indices de transmission tend à atténuer cette difficulté, ...».

Les données d'opinion subjective ont été recueillies au moyen d'essais en laboratoire ainsi que d'enquêtes sur le terrain assez poussées.

On notera que tous les essais ont été conduits avec des combinés de type Western Electric –500 ou équivalent.

3.3 Le modèle «CATNAP83»

Ce modèle traite des effets subjectifs de l'affaiblissement de circuit, de la distorsion d'affaiblissement en fonction de la fréquence, du bruit de circuit, de la distorsion de quantification, du bruit de salle et des trajets d'effet local dans une gamme assez étendue de valeurs et de combinaisons de ces caractéristiques.

Il est précisé que la structure de ce modèle est telle que le processus d'évaluation reflète les relations de cause à effet qui conduisent de l'entrée (propriétés de la connexion, environnement acoustique, caractéristiques auditives des participants, sons vocaux et systèmes langagiers, etc.) à la sortie (satisfaction des participants ou évaluation de la qualité de transmission).

Dans l'application pratique de ce modèle (sous la forme d'un programme informatique), les connexions doivent être spécifiées en termes d'éléments et de grandeurs comme les niveaux de bruit, les types particuliers d'appareils téléphoniques, les longueurs de câble avec indication de la résistance et de la capacité linéiques, les affaiblisseurs avec indication de leur effet. A partir de ces données, le programme calcule les équivalents pour la sonie, les niveaux des signaux vocaux et les notes d'opinion. Celles-ci sont exprimées sous la forme d'une note moyenne d'effort d'écoute moyen (Y_{le}) et d'une note moyenne d'opinion sur la conversation (Y_c).

3.4 Le modèle de «l'indice d'information»

Comme le modèle CATNAP, la théorie de ce modèle tient compte également des relations fondamentales de cause à effet à l'entrée: affaiblissement de transmission, bruit de circuit, bruit de salle, distorsion d'affaiblissement, effet local et diverses distorsions se produisant lors d'une transmission numérique.

A la sortie de ce modèle, le signal est «l'indice d'information» (I). La description du modèle est complétée d'exemples comparant les valeurs calculées de I avec les valeurs de notes moyennes d'opinion subjective (MOS).

3.5 Le modèle «OPINE»

Le modèle OPINE tient compte de l'affaiblissement de transmission, du bruit de circuit, de la distorsion d'affaiblissement en fonction de la fréquence, de la distorsion de quantification, de l'écho pour le locuteur et de l'effet local. Il modélise le processus psycho-auditif de l'évaluation de la qualité de transmission téléphonique par des humains sur la base de ces facteurs. (Il correspond donc au même type de modèle général que les modèles CATNAP et Indice d'information.)

Cinq facteurs psychologiques touchant à la qualité de la parole téléphonique ont été retenus:

- 1) distorsion de la parole pour la distorsion d'affaiblissement en fonction de la fréquence;
- 2) affaiblissement ou excès de sonie effective pendant la conversation;
- 3) bruit pendant les intervalles de conversation et les intervalles de non-conversation;
- 4) dégradation causée par l'écho pour le locuteur;
- 5) dégradation causée par l'effet local.

Chacun des facteurs ci-dessus est associé à un indice de qualité (PI, *performance index*). La note MOS pour une connexion est estimée à partir de la somme de tous les indices PI.

3.6 Le modèle de «l'indice de qualité de transmission»

Comme indiqué dans le Supplément 3 des Recommandations de la série P, il s'agit d'un modèle d'opinion de conversation simple, destiné à la prévision des effets combinés de l'équivalent global pour la sonie (OLR) et du bruit psophométrique. Il tient compte également des effets de l'affaiblissement d'effet local par la méthode de masquage (STMR), du bruit de salle et de la distorsion d'affaiblissement.

Ce modèle prédit les notes MOS en fonction des pourcentages «bon et excellent» et «médiocre et mauvais».

3.7 Le modèle «E»

A l'ETSI (Institut européen des normes de télécommunication), un groupe de travail a élaboré (sous la référence ETR 250) un Rapport technique sur la qualité de transmission de la parole de bout en bout dans les réseaux téléphoniques. Dans ce cadre, on a mis au point un modèle mathématique permettant d'estimer la perception par l'utilisateur de la qualité d'une communication vocale pour une connexion, en utilisant les paramètres de transmission «classiques» ainsi que des «facteurs de dégradation d'équipement» pour les codecs étudiés. Les résultats ont également été introduits dans les normes de l'UIT-T, lors des révisions des Recommandations G.101 et G.113.

Le modèle de l'ETSI (ou modèle E) reprend dans une large mesure le modèle «d'évaluation de l'indice de transmission» (TR, *transmission rating*) décrit au 3.2 ci-dessus, bien qu'il fasse appel à certaines caractéristiques issues des autres modèles décrits dans le Supplément 3 des Recommandations de la série P. Sa structure est toutefois différente du modèle TR. Le principe fondamental du modèle de l'ETSI repose sur une notion établie il y a plus de vingt ans par J. Allnatt et utilisée par exemple dans le modèle OPINE de NTT: «les facteurs psychologiques sont additifs sur l'échelle psychologique».

Le modèle de l'ETSI combine l'effet des divers paramètres de transmission pour construire un facteur d'évaluation R qui permet de prédire la réaction de l'utilisateur, par exemple sous la forme de pourcentages d'utilisateurs trouvant la connexion «au moins bonne», «au mieux médiocre» ou même mauvaise au point de procéder à une libération prématurée de la communication. Ce modèle permet également d'évaluer les notes qui seraient données lors d'une expérience de notes moyennes d'opinion (MOS). Le facteur d'évaluation R se compose des termes suivants:

$$R = R_o - I_s - I_d - I_e + A$$

où:

R_o est en principe le rapport de base signal vocal sur bruit;

I_s, I_d, I_e sont les «facteurs de dégradation»;

I_s est une dégradation se produisant en simultanéité avec le signal vocal, comme un volume téléphonique trop important, un effet local puissant, une distorsion de quantification due à la MIC;

I_d est une dégradation due à un retard, comme l'écho pour le locuteur ou pour l'auditeur ainsi qu'un temps de propagation absolu trop long;

I_e est une dégradation de transmission due à des équipements spéciaux tels que certains codecs à faible débit, à des équipements de multiplication de circuit numérique (DCME, *digital circuit multiplication equipment*), etc. (ce facteur est un nouveau concept);

A est un terme appelé *facteur d'expectative*. Il représente un «avantage d'accès» que certains systèmes possèdent par rapport aux habituels systèmes de communication câblés. [La perception d'une «haute qualité» est intimement liée à la mesure dans laquelle les attentes du client sont satisfaites, ce qui correspond à des notions telles que «la facilité d'utilisation (ou employabilité)», «l'utilité» et «les attitudes». Le facteur d'expectative est un nouveau concept qui n'a pas encore été utilisé dans les modèles mathématiques.]

Le modèle E ne tient pas compte des effets spécifiques de la distorsion d'affaiblissement en fonction de la fréquence et des réponses en fréquence des combinés téléphoniques. La raison en est que, dans un marché des télécommunications déréglementé, les opérateurs ne connaissent pas ces paramètres très précisément, de sorte qu'il faut se contenter des valeurs de planification en termes d'équivalents pour la sonie. Par ailleurs, la manipulation du combiné par l'utilisateur introduit de toute façon une grande variabilité dans la réponse en fréquence par fuites au pavillon.

En pratique, le modèle de l'ETSI donne à peu près les mêmes valeurs R que le modèle TR pour les dégradations «classiques». De même, le modèle TR pourra être remanié de façon à avoir la même structure des termes additifs pour les dégradations dont il se compose. Les résultats obtenus par l'application du modèle ETSI dans des connexions typiques concordent bien avec ceux qui ont été obtenus au moyen d'autres modèles ou avec ceux d'essais subjectifs de la littérature. Les dégradations dues à des codecs à faible débit peuvent en particulier être prédites par ce modèle avec une meilleure exactitude que par la méthode des «unités de distorsion de quantification» utilisée jusqu'à maintenant. (En planification concrète, la manière la plus pratique d'utiliser le modèle E consiste à fixer des limites à la somme des facteurs de dégradation, comme c'est le cas dans la Recommandation G.113.)

L'Annexe E/G.113 décrit la façon d'évaluer les codecs à faible débit, c'est-à-dire la façon de transformer des valeurs de notes MOS obtenues subjectivement en «facteurs de dégradation d'équipement».

La structure additive du modèle E facilite sa mise à jour éventuelle, si de nouvelles données issues d'essais subjectifs avec les «anciens» paramètres la rendent nécessaire ou si de nouveaux types de dégradation doivent être pris en compte.

4 Evolution future des modèles «cognitifs et perceptifs»

L'évolution des modèles cognitifs et perceptifs est un domaine auquel beaucoup de recherches sont consacrées. C'est également un sujet d'étude dans la Commission d'études 12 de l'UIT-T. La raison principale est la nécessité de mettre au point des méthodes fiables et objectives pour évaluer les codecs à faible débit (dont de nouveaux types semblent constamment apparaître). Actuellement, les performances de ces codecs doivent être évaluées par de coûteux essais subjectifs afin d'obtenir des résultats réalistes et fiables.

