



UNION INTERNATIONALE DES TÉLÉCOMMUNICATIONS

UIT-T

SECTEUR DE LA NORMALISATION
DES TÉLÉCOMMUNICATIONS
DE L'UIT

P.85

(06/94)

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

**MÉTHODE D'ÉVALUATION SUBJECTIVE
DE LA QUALITÉ DE PAROLE
DES SERVEURS VOCAUX**

Recommandation UIT-T P.85

(Antérieurement «Recommandation du ISCTT»)

AVANT-PROPOS

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

La Conférence mondiale de normalisation des télécommunications (CMNT), qui se réunit tous les quatre ans, détermine les thèmes d'études à traiter par les Commissions d'études de l'UIT-T lesquelles élaborent en retour des Recommandations sur ces thèmes.

L'approbation des Recommandations par les Membres de l'UIT-T s'effectue selon la procédure définie dans la Résolution n° 1 de la CMNT (Helsinki, 1^{er}-12 mars 1993).

La Recommandation UIT-T P.85, que l'on doit à la Commission d'études 12 (1993-1996) de l'UIT-T, a été approuvée le 21 juin 1994 selon la procédure définie dans la Résolution n° 1 de la CMNT.

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 de télécommunications.

© UIT 1994

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

TABLE DES MATIÈRES

	<i>Page</i>
1	Champ d'application..... 1
2	Méthode d'évaluation 1
2.1	Considérations générales 1
2.2	Caractéristiques essentielles de la méthode recommandée..... 1
3	Préparation du test..... 2
3.1	Stimuli..... 2
3.2	Sources..... 2
3.3	Préparation des stimuli..... 2
4	Schéma expérimental 2
4.1	Tâche des sujets 2
4.2	Echelles de notation 2
4.3	Plan d'expérience 2
4.4	Procédure d'écoute 3
5	Analyse statistique et présentation des résultats..... 3
6	Autres méthodes 4
	Annexe A – Messages 4
	Annexe B – Feuilles de réponse 5
	Annexe C – Evaluation de la parole synthétique: consignes aux auditeurs..... 8
	Références 9
	Bibliographie 9

RÉSUMÉ

Des serveurs vocaux permettent maintenant aux usagers du réseau téléphonique public commuté d'obtenir sous forme vocale divers renseignements: renseignements téléphoniques, prévisions météorologiques, données relatives à des achats par correspondance, etc. Etant produite par une machine, la parole issue de ces serveurs n'est pas d'une qualité parfaite.

La présente Recommandation définit une méthode d'évaluation subjective de la qualité de parole de serveurs vocaux qui permet de comparer plusieurs systèmes entre eux. Elle aidera les concepteurs de systèmes et les fournisseurs de services à contrôler la qualité de leurs produits.

La méthode utilise le principe des essais d'écoute. Les sujets écoutent des messages et expriment leur opinion sur une ou plusieurs échelles de notation après avoir répondu à des questions portant sur les informations contenues dans ces messages. Les résultats traduisent plusieurs aspects de la qualité et permettent ainsi de comparer l'adéquation de différents systèmes de synthèse aux objectifs poursuivis.

MÉTHODE D'ÉVALUATION SUBJECTIVE DE LA QUALITÉ DE PAROLE DES SERVEURS VOCAUX

(Genève, 1994)

1 Champ d'application

Des serveurs vocaux sont maintenant accessibles aux usagers du réseau téléphonique public commuté; ils mettent en oeuvre des annonces préenregistrées ou une parole synthétique. La parole synthétique peut être fabriquée à partir de segments de parole stockés: mots, syllabes, diphones, etc.; elle peut aussi être le produit d'une synthèse par règles, par exemple une synthèse à formants. Dans tous les cas, à cause de la compression numérique et du processus de traitement de signal comme la concaténation de segments ou des modifications d'intensité, d'intonation, de durées segmentales, la voix n'est pas d'une qualité parfaite.

S'appuyant sur la Recommandation P.80 et les expériences relatives à l'évaluation de messages vocaux synthétisés [1], [2], [3], la présente Recommandation définit une méthode de test destinée à l'évaluation de la qualité subjective de la parole synthétique. Certaines adaptations de la méthode telle qu'elle est décrite pourront se révéler nécessaires en fonction des particularités des systèmes à évaluer.

Cette méthode permet de prendre en compte aussi bien les performances des utilisateurs que leurs opinions et leurs réactions. Les opinions et réactions sont évaluées à l'aide d'échelles multiples.

La présente Recommandation traite à la fois les performances globales des systèmes et leur comportement dans des applications spécifiques; deux exemples d'application sont donnés en Annexe A.

Les informations que permet d'obtenir la méthode décrite dans la présente Recommandation représente des évaluations d'ensemble de la qualité acoustique de la sortie des machines parlantes. Des procédures pour l'évaluation spécifique de composants des systèmes de synthèse à partir du texte (par exemple le module de transcription phonétique du texte, etc.) sont à l'étude.

2 Méthode d'évaluation

2.1 Considérations générales

L'évaluation de la qualité de la parole synthétique peut s'appuyer sur les méthodes recommandées pour l'évaluation de la qualité de la parole en téléphonie que décrivent la Recommandation P.80 et le paragraphe 2.5 (Essais d'opinion) de la 2^e édition du *Manuel de téléphonométrie* [4]. La façon dont les auditeurs perçoivent la parole synthétique sera mieux décrite si l'on utilise des échelles d'opinion multiples. Du fait que la parole synthétique peut impliquer certains efforts pour être comprise, le test est conçu de façon à obliger les sujets à être attentifs au contenu informationnel des messages avant d'avoir à exprimer leur opinion.

2.2 Caractéristiques essentielles de la méthode recommandée

Plusieurs sources différentes seront présentées à l'écoute lors d'un test. Ainsi les opinions des sujets relatives à une source donnée auront été obtenues en rapport avec les autres sources. Les sources seront des systèmes de synthèse ainsi que des conditions de référence (parole naturelle avec une dégradation calibrée ou systèmes de synthèse connus).

Les sujets doivent exprimer leur opinion à l'aide d'une ou plusieurs échelles d'opinion à 5 points, comme dans les méthodes d'évaluation par catégories absolues (ACR) (*absolute category rating*) et d'évaluation par catégories de dégradation (DCR) (*degradation category rating*) décrites dans la Recommandation P.80. Outre l'échelle d'impression générale on peut utiliser des échelles d'effort d'écoute, d'agrément de la voix, etc.

Les messages émis doivent se rapporter à des applications pratiques. Concrètement, des applications différentes impliquent des sessions de test différentes.

Chaque message est présenté deux fois. Lors de la première écoute, les sujets répondent à des questions portant sur les informations contenues dans le message; lors de la seconde écoute, les sujets jugent la qualité de la parole et expriment leur opinion sur une ou plusieurs échelles de notation.

3 Préparation du test

3.1 Stimuli

Les messages doivent être d'une longueur suffisante pour que les sujets disposent du temps nécessaire, soit à l'écriture sur une feuille de réponse des informations qui leur sont demandées, soit à l'expression de leurs jugements sur les échelles de notation de l'autre feuille de réponse. On choisira une durée de message comprise entre 10 et 30 secondes.

Chaque message est constitué de parties fixes, spécifiques de l'application, et de parties variables qui changent d'un message à l'autre. Les messages doivent être construits de façon à ce que la prédictibilité des parties variables ne diffère pas significativement d'un message à l'autre. L'Annexe A donne des exemples de tels messages. On peut utiliser des stimuli présentant un autre degré de difficulté (charge de la mémoire à court terme).

3.2 Sources

On utilisera si possible au moins cinq sources différentes; le nombre de sources dépendra des systèmes à tester, des applications traitées et du schéma expérimental. L'une des sources au moins sera une voix naturelle (voix d'homme ou de femme selon les systèmes testés). Cette (ces) voix naturelle(s), dégradée(s) par un bruit multiplicatif conforme à la Recommandation P.81 (voir «Conditions de référence» en B.2.3/P.80), peut (peuvent) servir de conditions de référence. Cependant, des recherches en cours semblent indiquer que d'autres dégradations, tels le T-Reference System [6] ou le Time and Frequency Warping (TFW) [7], pourraient être mieux adaptées à l'évaluation de voix synthétiques.

3.3 Préparation des stimuli

Ce paragraphe est analogue à B.1/P.80 (Enregistrement des sources), sauf en ce qui concerne le microphone: on enregistrera la voix naturelle avec un microphone à réponse plate.

4 Schéma expérimental

4.1 Tâche des sujets

Les sujets reçoivent les instructions décrivant le test et des feuilles de réponse. On leur demande d'utiliser deux feuilles par message: l'une pour recueillir la transcription d'informations contenues dans le message, l'autre pour recueillir les réponses sur une ou plusieurs échelles d'opinion.

4.2 Echelles de notation

Les échelles de notation sont les suivantes:

– impression générale		(questionnaires de type I et questionnaires de type Q)
– effort d'écoute	}	(questionnaires de type I)
– difficultés de compréhension		
– articulation	}	(questionnaires de type Q)
– prononciation		
– débit de parole		
– agrément de la voix		
– acceptabilité		(questionnaires de type I et de type Q)

Le libellé des questions et les catégories des échelles sont donnés dans l'Annexe B.

4.3 Plan d'expérience

4.3.1 Si le nombre de conditions est suffisant, c'est-à-dire au moins 7, le plan d'expérience sera constitué à partir de carrés gréco-latins (GL) (*graeco latin squares*). Les quatre facteurs sont les suivants: source, message, ordre de présentation, groupe de sujets.

4.3.2 A l'intérieur d'une session, les messages sont relatifs à une seule et même application. Si le test doit être dupliqué on utilisera des messages analogues mais différents.

4.3.3 Un message ne doit pas être réutilisé après avoir été présenté deux fois.

4.3.4 Lorsque l'ensemble des échelles est utilisé, on divisera une session en deux blocs, chaque bloc correspondant à un questionnaire de type I ou de type Q (voir l'Annexe 8). En cas d'utilisation de carrés gréco-latins, chacun des deux blocs d'une session sera organisé en fonction de deux carrés gréco-latins distincts.

4.3.5 Une séance peut être constituée d'une ou plusieurs sessions. Les sessions de test proprement dites seront précédées d'une session d'apprentissage comportant six messages au moins émis par des sources choisies de façon à ce qu'ils correspondent à l'éventail de qualité rencontré dans le test.

4.3.6 Si l'on utilise des carrés gréco-latins, le nombre de sujets sera supérieur ou égal à 4 fois la dimension des carrés. En d'autres termes, le nombre de sujets d'un groupe sera au moins égal à 4.

4.3.7 L'intervalle de temps séparant les deux présentations d'un même message sera de l'ordre de 8 secondes; celui séparant des paires de messages différents sera de l'ordre de 20 secondes. Ils dépendent en fait de la durée effective des messages du test.

4.3.8 La durée totale du test pour les sujets sera comprise entre 40 et 60 minutes, consignes aux sujets, apprentissage et pause compris.

4.3.9 Si des voix naturelles figurent parmi les sources, l'une d'elles sera incluse dans la session d'apprentissage.

4.4 Procédure d'écoute

4.4.1 *Conditions d'écoute* – Alinéa analogue au B.4.1/P.80.

4.4.2 *Système d'écoute* – Alinéa analogue au B.4.2/P.80. Toutes les sources seront soumises au même filtre passe-bande (en fonction des conditions prévues pour l'application, par exemple 300-3400 Hz).

4.4.3 *Niveau d'écoute* – L'idéal est que les messages soient présentés au niveau préféré pour la parole synthétique. Si celui-ci n'est pas connu, on prendra le niveau préféré pour la parole codée (79 dB acoustique, -15 dB/Pa, voir le paragraphe 2.5.8.1 du *Manuel de Téléphonométrie*). Dans la mesure du possible, on présentera aux mêmes sujets un ou plusieurs blocs du test à deux niveaux supplémentaires, l'un supérieur, l'autre inférieur au niveau préféré.

4.4.4 *Auditeurs* – Alinéa analogue au B 4.4/P.80.

4.4.5 *Instructions aux sujets* – L'Annexe C donne un exemple de consignes aux sujets. Les instructions doivent être fournies sous forme écrite. Elles peuvent être présentées aussi oralement; dans ce cas on utilisera de préférence un enregistrement.

5 Analyse statistique et présentation des résultats

La présentation synthétique des résultats se fera sous forme d'histogrammes et/ou de distributions cumulées pour chaque échelle d'opinion.

Pour comparer différentes sources on tracera les distributions cumulées de chacune de ces sources à raison d'un diagramme par échelle (voir la Figure 1).

On peut aussi calculer, pour l'échelle d'impression générale et pour l'échelle d'effort d'écoute, les notes moyennes d'opinion (MOS) (*mean opinion score*) de chaque source pour chaque type de message. Une analyse de la variance et des tests de comparaison multiple différences honnêtement significatives (HSD) (*honestly significant difference*) seront effectués pour chaque MOS calculée.

Aucune procédure particulière n'est recommandée pour l'analyse des réponses concernant les informations contenues dans les messages. Si toutefois une source particulière présente des performances nettement moins bonnes que les autres (au niveau du taux de réponses exactes par exemple) on pourra en tirer certaines conclusions.

Les résultats d'acceptabilité seront donnés en pourcentages de réponses positives et de réponses négatives.

Les résultats de la session d'apprentissage ne seront pas considérés.

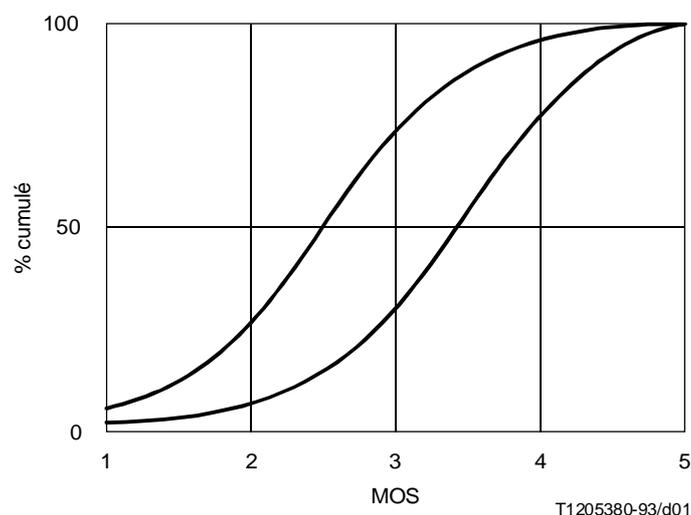


FIGURE 1/P.85
Distributions cumulées des MOS

6 Autres méthodes

Si l'on souhaite avoir une estimation quantitative de l'intelligibilité globale d'un système de synthèse à partir du texte (TTS) (*text-to-speech*), il est particulièrement indiqué de procéder à des tests portant sur des phrases. Un tel test a été élaboré dans le cadre d'un projet européen consacré à l'évaluation multilingue de la synthèse et de la reconnaissance de la parole (projet Esprit n° 2589, «SAM»). Ce test, dit «SUS» (Semantically Unpredictable Sentences, Phrases sémantiquement non prévisibles), est essentiellement destiné à l'évaluation de systèmes TTS en cours d'étude [5].

Annexe A

Messages

(Cette annexe fait partie intégrante de la présente Recommandation)

Cette annexe présente des exemples de messages tirés de l'expérience décrite dans les références [3].

L'expérience mettait en jeu deux applications: commande par correspondance (M) et renseignements ferroviaires (R). On donne trois messages pour chacune d'elles:

- M1:** Madame Robert, les chaussures de sport, couleur blanc, pointure 42-44, référence 501-97-52, au prix de 319 francs, vous seront livrées dans 1 semaine.
- M2:** Monsieur Gauthier, le téléviseur multistandard Pal-Secam, télécommande multifonction, modèle 36 cm, référence 811-61-32, au prix de 2 492 francs, vous sera livré dans 3 semaines.
- M3:** Monsieur Rousseau, la perceuse électrique D162, puissance 550 watts, 2 vitesses, référence 481-20-30, au prix de 499 francs, vous sera livrée dans 2 semaines.
- R1:** Le train n° 9783 en provenance de Poitiers arrivera à 9 heures 24, quai 3, voie G.
- R2:** Le train n° 7826 en direction de Nice partira à 12 heures 20, quai 9, voie A.
- R3:** Le train n° 4320 en provenance de Clermont-Ferrand arrivera à 17 heures 44, quai 2, voie C.

Annexe B

Feuilles de réponse

(Cette annexe fait partie intégrante de la présente Recommandation)

Les figures ci-dessous donnent des exemples de feuilles de réponse. Dans les Figures B.1 et B.2, les applications sont les mêmes que celle de l'Annexe A. Voir les Figures B.3 et B.4.

Nom	<input type="text"/>	
Désignation de l'article (1 à 3 mois)	<input type="text"/>	
Référence	<input type="text"/>	
Prix	<input type="text"/>	francs
Délai de livraison	<input type="text"/>	semaines

FIGURE B.1/P.85

**Les cinq questions relatives à l'application
de commande par correspondance**

Train n°	<input type="text"/>
Destination ou provenance	<input type="text"/>
Horaire	<input type="text" value="Heure"/>
Quai	<input type="text"/>
Voie	<input type="text"/>

FIGURE B.2/P.85

**Les cinq questions relatives à l'application
de renseignements ferroviaires**

<p>Impression générale</p> <p><i>Comment jugez-vous la qualité sonore de ce que vous venez d'entendre?</i></p> <p><input type="checkbox"/> Excellente</p> <p><input type="checkbox"/> Bonne</p> <p><input type="checkbox"/> Passable</p> <p><input type="checkbox"/> Médiocre</p> <p><input type="checkbox"/> Mauvaise</p>		
<p>Effort d'écoute</p> <p><i>Comment qualifiez-vous l'effort d'écoute nécessaire pour comprendre le message?</i></p> <p><input type="checkbox"/> Détente absolue: aucun effort</p> <p><input type="checkbox"/> Attention nécessaire; pas d'effort appréciable</p> <p><input type="checkbox"/> Effort modéré</p> <p><input type="checkbox"/> Effort considérable</p> <p><input type="checkbox"/> Signification incompréhensible en dépit de tous les efforts possibles</p>	<p>Difficultés de compréhension</p> <p><i>Avez-vous éprouvé de la difficulté pour comprendre certains mots?</i></p> <p><input type="checkbox"/> Jamais</p> <p><input type="checkbox"/> Rarement</p> <p><input type="checkbox"/> De temps en temps</p> <p><input type="checkbox"/> Souvent</p> <p><input type="checkbox"/> Tout le temps</p>	<p>Articulation</p> <p><i>L'articulation vous paraît-elle claire?</i></p> <p><input type="checkbox"/> Très claire</p> <p><input type="checkbox"/> Suffisamment claire</p> <p><input type="checkbox"/> Moyennement claire</p> <p><input type="checkbox"/> Assez confuse</p> <p><input type="checkbox"/> Très confuse</p>
<p>Acceptabilité</p> <p><i>Pensez-vous que cette voix serait acceptable dans un service vocal téléphonique de ce type ?</i></p> <p><input type="checkbox"/> Oui</p> <p><input type="checkbox"/> Non</p>		
<p>Observations:</p>		

FIGURE B.3/P.85

Questionnaire de type I avec toutes les échelles

Impression générale		
<i>Comment jugez-vous la qualité sonore de ce que vous venez d'entendre?</i>		
<input type="checkbox"/> Excellente <input type="checkbox"/> Bonne <input type="checkbox"/> Passable <input type="checkbox"/> Médiocre <input type="checkbox"/> Mauvaise		
Qualité de prononciation	Débit	Agrément
<i>Déterminez-vous des anomalies dans la prononciation?</i>	<i>Le débit moyen de l'énoncé était:</i>	<i>Comment qualifiez-vous la voix?</i>
<input type="checkbox"/> Non <input type="checkbox"/> Oui, mais pas gênantes <input type="checkbox"/> Oui, un peu gênantes <input type="checkbox"/> Oui, gênantes <input type="checkbox"/> Oui, très gênantes	<input type="checkbox"/> Beaucoup trop rapide <input type="checkbox"/> Un peu trop rapide <input type="checkbox"/> Satisfaisant <input type="checkbox"/> Un peu trop lent <input type="checkbox"/> Beaucoup trop lent	<input type="checkbox"/> Très agréable <input type="checkbox"/> Agréable <input type="checkbox"/> Passable <input type="checkbox"/> Désagréable <input type="checkbox"/> Très désagréable
Acceptabilité		
<i>Pensez-vous que cette voix serait acceptable dans un service vocal téléphonique de ce type?</i>		
<input type="checkbox"/> Oui <input type="checkbox"/> Non		
Observations:		

FIGURE B.4/P.85

Questionnaire de type Q avec toutes les échelles

Annexe C

Evaluation de la parole synthétique: consignes aux auditeurs¹⁾

(Cette annexe fait partie intégrante de la présente Recommandation)

Vous allez participer à une expérience d'évaluation de différentes machines parlantes.

Vous entendrez deux sortes de messages: des messages relatifs à la vente par correspondance et des messages relatifs au trafic ferroviaire. Voici un exemple de chaque sorte de message:

- Madame Morin, la montre Data-Bank 50 mémoires, bracelet résine, référence 811.19.04, au prix de 479 francs, vous sera livrée dans 3 semaines.
- Le train numéro 4119 en provenance d'Orléans arrivera à 12 heures 23, quai 8, voie H.

Chaque message sera prononcé deux fois. Lors de la première écoute, vous aurez à remplir des cases qui ont pour but de vérifier la bonne compréhension des informations présentes dans le message. Lors de la seconde écoute, on vous demandera de juger la qualité vocale du message en répondant à un questionnaire qui comporte cinq questions. Certaines questions varient d'une partie à l'autre du test.

Les jugements d'opinions qui vous sont demandés portent sur les éléments suivants:

Impression générale – Nous vous demandons d'exprimer votre opinion en imaginant ce que serait votre réaction dans une application réelle : message téléphonique provenant d'un organisme de vente par correspondance, ou demande de renseignement à une agence de voyages.

Effort d'écoute – Votre réponse doit traduire l'effort qui vous est nécessaire pour comprendre l'ensemble du message et en extraire les informations que vous aviez à transcrire.

Difficultés de compréhension – Vous devez indiquer dans quelle mesure le contenu du message est difficile à comprendre. La question ne concerne pas seulement les mots que vous avez transcrits, elle porte sur tous les mots du message.

Articulation – Vous devez donner votre opinion sur la netteté de la prononciation (selon que vous distinguez plus ou moins bien les sons).

Acceptabilité – Veuillez indiquer si vous considérez ou non que la voix que vous avez entendue serait acceptable dans un service vocal téléphonique de ce type.

Impression générale – Nous vous demandons d'exprimer votre opinion en imaginant ce que serait votre réaction dans une application réelle : message téléphonique provenant d'un organisme de vente par correspondance, ou demande de renseignement à une agence de voyage.

Qualité de prononciation – Cette question est relative à d'éventuels écarts de la prononciation naturelle (intonation, rythme, découpage, etc.).

Débit – Votre réponse doit traduire ce que serait votre réaction à la vitesse d'élocution si vous étiez dans la situation réelle.

Agrément – Cette question est relative à l'agrément de la voix telle qu'elle vous parvient.

Acceptabilité – Veuillez indiquer si vous considérez ou non que la voix que vous avez entendue serait acceptable dans un service vocal téléphonique de ce type.

Vous disposerez de deux feuilles de réponse par message. La première contient les cadres réservés à la transcription des informations qui vous sont demandées. La seconde contient le questionnaire d'opinion. (*N.B.*: nous vous demandons d'utiliser une feuille par présentation et de changer de feuille entre les deux présentations du message; vous ne devez pas revenir à une feuille antérieure.)

Pour vous familiariser avec l'écoute des messages et la réponse aux questionnaires, la séance commencera par 6 messages d'apprentissage. Ils vous fourniront un éventail des systèmes et des voix que vous aurez à entendre, et vous donneront l'occasion de répondre aux différents types de questionnaire. Une pause après ces six messages vous permettra de signaler si vous rencontrez des difficultés et souhaitez des explications complémentaires.

¹⁾ Cet exemple met en jeu les deux mêmes applications que les Annexes A et B; toutes les échelles sont utilisées. Le test se décompose en 2 × 2 blocs (2 applications, 2 types de questionnaire).

La séance proprement dite sera partagée en deux sessions séparées par quelques minutes de repos. Au cours de la première session, vous entendrez quatorze messages de vente par correspondance: sept seront évalués avec l'un des ensembles de questions et, après une courte pause, sept avec l'autre ensemble de questions. Au cours de la deuxième session, vous entendrez deux blocs de sept messages relatifs à la circulation des trains.

Merci de votre participation.

Références

- UIT-T Recommandation P.80 – *Méthodes d'évaluation subjective de la qualité de transmission.*
- UIT-T Recommandation P.81 – *Appareil de référence à bruit modulé.*

Bibliographie

- [1] CCITT – Annexe du Rapport COM XII-R 12 (1986), Subjective assessment of automatic voice answering devices, CSELT (Italy).
- [2] CCITT – Contribution COM XII-176 (1987), Subjective quality assessment of synthetic speech, Swedish Telecom.
- [3] CARTIER (M.), EMERARD (F.), PASCAL (D.), COMBESCURE (P.) et SOUBIGOU (A.): Une méthode d'évaluation multicritère de sorties vocales; application au test de 4 systèmes de synthèse à partir du texte, 19^{es} Journées d'Etude sur la Parole (*Société Française d'Acoustique et Association Belge des Acousticiens*) Bruxelles, 19-22 mai 1992.
- [4] Manuel de téléphonométrie, 2^e édition, UIT (à paraître)
- [5] BENOÎT (C.), GRICE (M.) et HAZAN (V.): The SUS test: a method for the assessment of test-to-speech synthesis intelligibility. (Article soumis pour publication dans *Speech Communication*).
 - In ESPRIT Project 1541 (SAM), Multilingual Speech Input/Output: Assessment, Methodology and Standardization; Extension Phase Final Report (1 April 1988-28 February 1989); compiled and edited by HARLAND (G.), FOURCIN (A.), BARRY (W.J.) and GRICE (M.), University College London, 344 p., February 1989.
 - Un programme de génération automatique de phrases de test en 6 langues est disponible sur demande. Contacter BENOÎT (C.), Institut de la Communication Parlée, Université de Stendhal, B.P. 25X, 38040 Grenoble, France.
- [6] Bill Cotton: New reference condition for very low bit rate coder evaluation, *Globecom'92 Conference Record*, Vol. 3, p. 1719-1722, 6-9 décembre 1992.
- [7] UIT-T Contribution COM 12-18 (1993), An on-going series of subjective experiments to assess speech output from text-to-speech systems, *British Telecom*.