



UNION INTERNATIONALE DES TÉLÉCOMMUNICATIONS

UIT-T

SECTEUR DE LA NORMALISATION
DES TÉLÉCOMMUNICATIONS
DE L'UIT

P.59

(03/93)

QUALITÉ DE LA TRANSMISSION TÉLÉPHONIQUE APPAREILS DE MESURES OBJECTIVES

VOIX CONVERSATIONNELLE ARTIFICIELLE

Recommandation UIT-T P.59

(Antérieurement «Recommandation du CCITT»)

AVANT-PROPOS

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

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

La Recommandation UIT-T P.59, élaborée par la Commission d'études XII (1988-1993) de l'UIT-T, a été approuvée par la CMNT (Helsinki, 1-12 mars 1993).

NOTES

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

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

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

TABLE DES MATIÈRES

	<i>Page</i>
1 Introduction	1
2 Caractéristiques de la conversation humaine	1
2.1 Caractéristiques mesurées sans temps de maintien.....	1
2.2 Caractéristiques mesurées avec temps de maintien	2
3 Méthode de production de la conversation artificielle	3
Références	4

VOIX CONVERSATIONNELLE ARTIFICIELLE

(Helsinki, 1993)

1 Introduction

Le signal décrit ici reproduit les caractéristiques temporelles (présence/absence de signal) de la conversation humaine servant à caractériser des systèmes de traitement de la parole qui ont des détecteurs vocaux tels que les appareils téléphoniques à haut-parleur, les dispositifs de réduction d'écho, les équipements de multiplication de circuits numériques (DCME) (*digital circuit multiplication equipment*), les systèmes de transmission par paquets et les systèmes en mode de transfert asynchrone (ATM) (*asynchronous transfer mode*). Ce signal reflète les paramètres de la conversation humaine tels que la longueur de la salve de parole, de la pause, de la double parole et du silence mutuel. Les chapitres qui suivent décrivent ces caractéristiques et une méthode de production de la conversation artificielle.

NOTES

1 La voix artificielle décrite dans la Recommandation P.50 est un signal monovoie sans pauses qui est utilisé pour des mesures objectives de systèmes et de dispositifs de traitement de la parole dans lesquels le mode de conversation n'entre pas en ligne de compte, par exemple les codecs vocaux.

2 La conversation artificielle décrite dans la présente Recommandation produit la voix artificielle décrite dans la Recommandation P.50 pendant les salves de parole.

2 Caractéristiques de la conversation humaine

Les durées et les débits des salves de parole et des pauses varient selon les conditions de mesure. Deux valeurs pour chaque paramètre de la conversation sont spécifiées ci-après. L'une est basée sur la mesure de la voix sans temps de maintien tandis que l'autre est basée sur une mesure effectuée avec temps de maintien.

2.1 Caractéristiques mesurées sans temps de maintien

Les caractéristiques décrites ci-dessous ont été extraites de la référence [1].

1) Caractéristiques des salves de parole

La fonction de densité de probabilité (pdf) de la durée des salves de parole est modélisée par deux pdf géométriques pondérées:

$$f_1(k) = C_1(1 - U_1) U_1^{k-1} + C_2(1 - U_2) U_2^{k-1}, k = 1, 2, 3, \dots$$

où

$$C_1 = 0,60278 \quad U_1 = 0,92446$$

$$C_2 = 0,39817 \quad U_2 = 0,98916$$

Chaque incrément de la variable k est égal à 5 ms. La fonction de distribution cumulative des durées des salves de parole est représentée dans le diagramme a) de la Figure 1. La durée moyenne des salves de parole est de 227 ms.

2) Caractéristiques des pauses

La pdf de la durée des pauses est également modélisée par deux pdf géométriques pondérées:

$$f_p(k) = D_1(1 - W_1) W_1^{k-1} + D_2(1 - W_2) W_2^{k-1}, k = 1, 2, 3, \dots$$

où

$$D_1 = 0,76693 \quad W_1 = 0,89700$$

$$D_2 = 0,23307 \quad W_2 = 0,99791$$

La fonction de distribution cumulative de la durée des pauses est représentée dans le diagramme b) de la Figure 1.

3) Facteur d'activité

La durée moyenne des pauses égale à 596 ms combinée à la durée moyenne des salves de parole égale à 227 ms, donne un facteur d'activité de parole à long terme de 27,6%.

NOTE – Cette valeur est mesurée au moyen d'un appareil de mesure sans temps de maintien. Cependant, si on utilise un appareil de mesure conforme à la Recommandation P.56, on obtiendra probablement un facteur d'activité plus élevé (voir le Tableau 1).

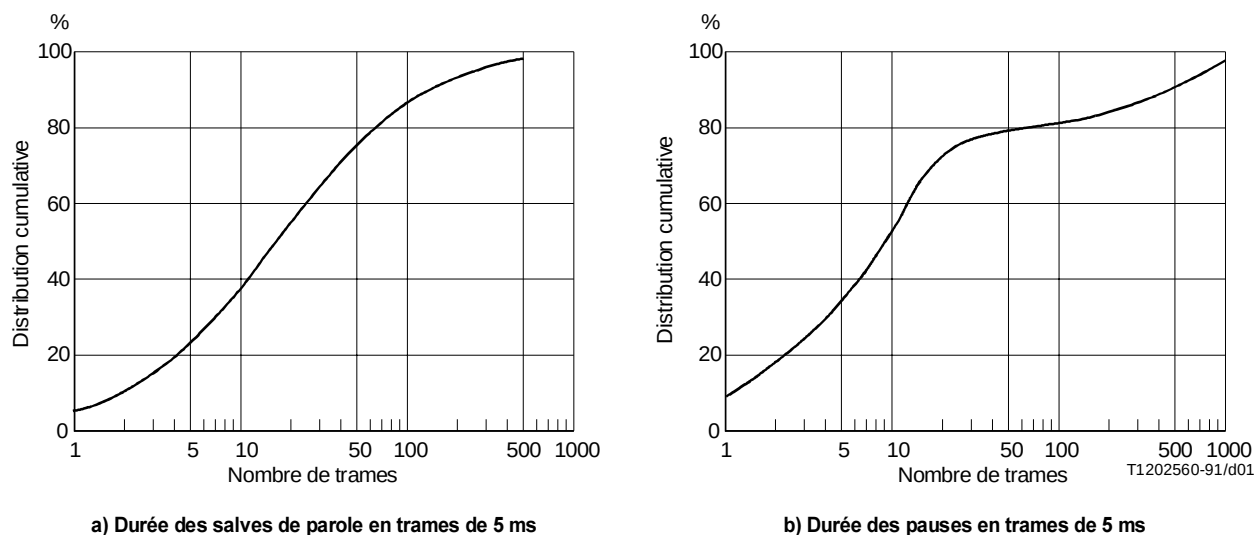


FIGURE 1/P.59

**Distribution cumulative des durées des salves de parole et des pauses
(sans temps de maintien)**

2.2 Caractéristiques mesurées avec temps de maintien

Le Tableau 1 énumère les valeurs des principaux paramètres de la conversation humaine. On a obtenu ces valeurs en établissant la moyenne des valeurs indiquées dans les Références [1] à [4].

TABLEAU 1/P.59

**Paramètres temporels de la conversation
(moyenne pour l'anglais, l'italien et le japonais)**

Paramètre	Durée (s)	Taux (%)
Salve de parole	1,004	38,53
Pause	1,587	61,47
Double parole	0,228	6,59
Silence mutuel	0,508	22,48

La fonction de distribution cumulative de la durée des salves de parole est calculée approximativement à l'aide d'une fonction exponentielle et celle de la durée des pauses à l'aide d'une fonction constante-plus-exponentielle, soit, pour les salves de parole:

$$P_{ts}(t) = 1 - \exp(-A_{ts} \cdot t)$$

$$A_{ts} = 1/\bar{T}_{ts}$$

$\bar{T}_{ts}$: durée moyenne des salves de parole et pour les pauses,

$$P_{ps}(t) = \begin{cases} 0 & \text{pour } 0 \leq t \leq 0,2 \\ 1 - \exp[-A_{ps}(t - 0,2)] & \text{pour } t > 0,2 \end{cases}$$

$$A_{ps} = 1/(\bar{T}_{ps} - 0,2) \quad \bar{T}_{ps}: \text{durée moyenne des pauses.}$$

Les deux caractéristiques sont représentées sur la Figure 2.

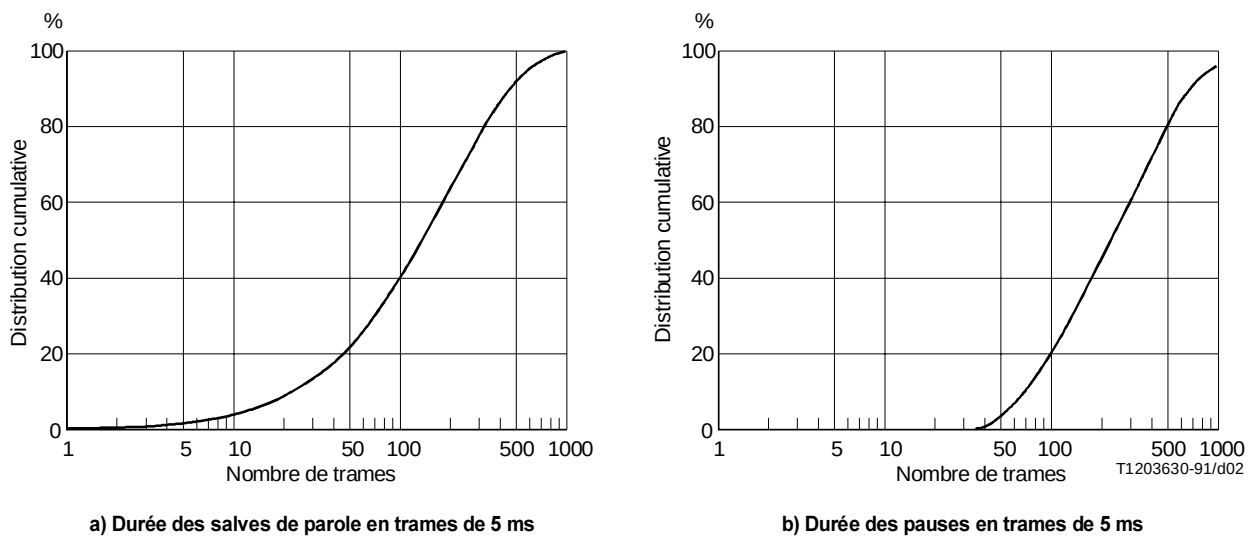


FIGURE 2/P.59

**Distribution cumulative des durées des salves de parole et des pauses
(avec temps de maintien)**

3 Méthode de production de la conversation artificielle

Les salves de parole et les pauses sont produites selon le modèle de transition d'état représenté sur la Figure 3 dans lequel P_1 , P_2 et P_3 désignent les probabilités de transition exprimées en pourcentage. La voix artificielle décrite dans la Recommandation P.50 est produite pendant une salve de parole.

T_{st} (durée de la simple parole), T_{dt} (durée de la double parole) et T_{ms} (durée du silence mutuel) varient selon les équations suivantes. Les temps dans ces équations sont exprimés en secondes.

$$T_{st} = -0,854 \ln(1 - x_1)$$

$$T_{dt} = -0,226 \ln(1 - x_2)$$

$$T_{ms} = -0,456 \ln(1 - x_3)$$

$0 < x_1, x_2, x_3 < 1$: variables aléatoires avec distribution uniforme.

Si la durée des pauses est inférieure à 200 ms, le modèle choisit soit l'état de simple parole, soit l'état de silence mutuel avec des probabilités de 50% jusqu'à ce que la durée des pauses dépasse 200 ms. Les valeurs de P_1 , P_2 , P_3 sont égales respectivement à 40, 50 et 50. La durée totale de la conversation artificielle doit être d'au moins 10 minutes pour se conformer aux caractéristiques spécifiées en 2.2.

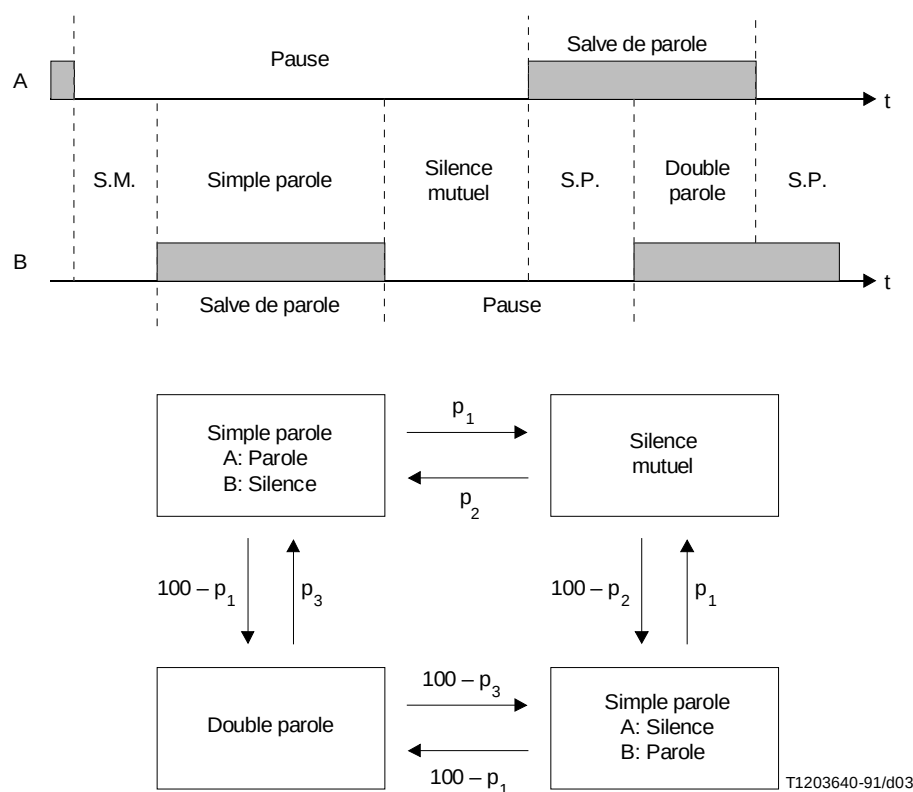


FIGURE 3/P.59
Modèle de transition d'état pour la conversation

Références

- [1] LEE (H. H.), UN (C. K.): A study of on-off characteristics of conversational speech, *IEEE Trans. on Comm.*, Vol. COM-34, n° 6, p. 630-637, juin 1986.
- [2] BRADY (P. T.): A statistical analysis of on-off patterns in 16 conversations, *BSTJ*, p. 73-91, janvier 1968.
- [3] CCITT Contribution COM XII-20 *Caractéristiques des conversations en présence/en l'absence de signaux vocaux* (CSELT), période d'études 1989-1992.
- [4] CCITT Contribution tardive D.42 (GT XII/1) *Collecting procedure for on-off characteristics of conversational speech in telecommunication* (NTT), période d'études 1989-1992.
- [5] CCITT Contribution tardive COM 64 (GT XII/1) *Generation of artificial voice with pauses* (NTT), période d'études 1989-1992.