UNIÓN INTERNACIONAL DE TELECOMUNICACIONES

UIT-T P.59

SECTOR DE NORMALIZACIÓN
DE LAS TELECOMUNICACIONES
DE LA UIT

(03/93)

CALIDAD DE TRANSMISIÓN TELEFÓNICA
APARATOS PARA MEDICIONES OBJETIVAS

HABLA CONVERSACIONAL ARTIFICIAL

Recomendación UIT-T P.59

(Anteriormente «Recomendación del CCITT»)

PREFACIO

El Sector de Normalización de las Telecomunicaciones de la UIT (UIT-T) es un órgano permanente de la Unión Internacional de Telecomunicaciones. El UIT-T tiene a su cargo el estudio de las cuestiones técnicas, de explotación y de tarificación y la formulación de Recomendaciones al respecto con objeto de normalizar las telecomunicaciones sobre una base mundial.

La Conferencia Mundial de Normalización de las Telecomunicaciones (CMNT), que se reúne cada cuatro años, establece los temas que habrán de abordar las Comisiones de Estudio del UIT-T, que preparan luego Recomendaciones sobre esos temas.

La Recomendación UIT-T P.59, preparada por la Comisión de Estudio XII (1988-1993) del UIT-T, fue aprobada por la CMNT (Helsinki, 1-12 de marzo de 1993).

NOTAS

1 Como consecuencia del proceso de reforma de la Unión Internacional de Telecomunicaciones (UIT), el CCITT dejó de existir el 28 de febrero de 1993. En su lugar se creó el 1 de marzo de 1993 el Sector de Normalización de las Telecomunicaciones de la UIT (UIT-T). Igualmente en este proceso de reforma, la IFRB y el CCIR han sido sustituidos por el Sector de Radiocomunicaciones.

Para no retrasar la publicación de la presente Recomendación, no se han modificado en el texto las referencias que contienen los acrónimos «CCITT», «CCIR» o «IFRB» o el nombre de sus órganos correspondientes, como la Asamblea Plenaria, la Secretaría, etc. Las ediciones futuras en la presente Recomendación contendrán la terminología adecuada en relación con la nueva estructura de la UIT.

2 Por razones de concisión, el término «Administración» se utiliza en la presente Recomendación para designar a una administración de telecomunicaciones y a una empresa de explotación reconocida.

ÍNDICE

Recomendación P.59

HABLA CONVERSACIONAL ARTIFICIAL

(Helsinki, 1993)

1 Introducción

La señal aquí descrita reproduce las características temporales on-off (activación-pausa) del habla conversacional humana para caracterizar sistemas de procesamiento de habla que disponen de detectores de habla, por ejemplo, teléfonos con altavoz, dispositivos de control de eco, equipos de multiplicación de circuitos digitales (DCME, *digital circuit multiplication equipment*), sistemas por paquetes y sistemas en modo de transferencia asíncrono (ATM, *asynchronous transfer mode*). Esta señal refleja parámetros de conversación humana como, por ejemplo, la duración de la ráfaga de habla, la pausa, la diafonía o el silencio mutuo. En los capítulos siguientes se describen estas características, así como un método para la generación de habla conversacional artificial.

NOTAS

1 La voz artificial descrita en la Recomendación P.50, que es una señal monocanal carente de pausas, es utilizada para la medición objetiva de sistemas y dispositivos de procesamiento de habla en los que el entorno conversacional es indiferente (por ejemplo, los codecs de habla).

2 El habla conversacional artificial descrita en esta Recomendación genera, durante las ráfagas de habla, la voz artificial descrita en la Recomendación P.50.

2 Características del habla conversacional humana

La duración y cadencia de las ráfagas de conversación y de las pausas varían en función de las condiciones de medición. Se especifican a continuación dos valores para cada parámetro de habla conversacional, basados en la medición de habla sin y con tiempo de retención, respectivamente.

2.1 Características medidas sin tiempo de retención

Las características descritas a continuación han sido obtenidas de [1].

1) *Características de la ráfaga de conversación*

La función de densidad de probabilidad (pdf) de la duración de las ráfagas de conversación está modelizada por dos pdf geométricas ponderadas:

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

donde

$$C_1 = 0,60278 \quad U_1 = 0,92446$$
$$C_2 = 0,39817 \quad U_2 = 0,98916$$

Cada incremento de la variable k es igual a 5 ms. En el diagrama a) de la Figura 1 se representa la función de distribución acumulativa de las duraciones de ráfaga de conversación. La duración media de ráfagas de conversación es de 227 ms.

2) *Características de pausa*

La pdf de las duraciones de pausa está también modelizada por dos pdf geométricas ponderadas:

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

donde

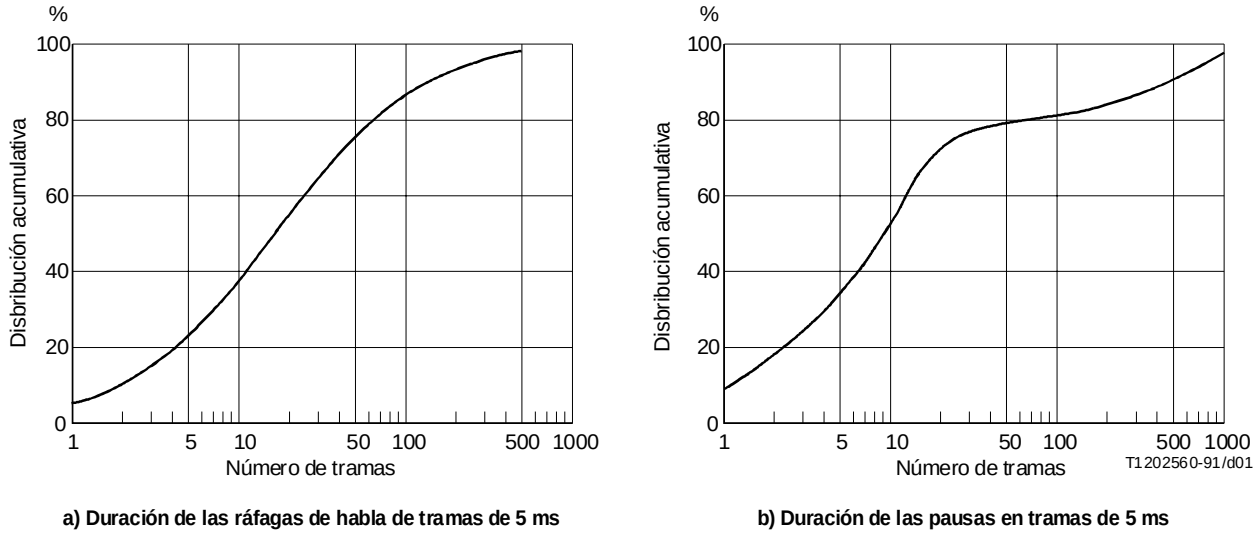
$$D_1 = 0,76693 \quad W_1 = 0,89700$$
$$D_2 = 0,23307 \quad W_2 = 0,99791$$

La función de distribución acumulativa de las duraciones de pausa está representada en el diagrama b) de la Figura 1.

3) *Factor de actividad*

Una duración media de pausas de 596 ms, junto con una duración media de ráfaga de conversación de 227 ms, proporciona un factor de actividad de habla a largo plazo de 27,6%.

NOTA – Para medir este valor se utiliza un medidor sin tiempo de retención. Sin embargo, si se utilizase un medidor conforme a la Recomendación. P.56, cabría prever un factor de actividad más elevado. (Véase el Cuadro 1.)



a) Duración de las ráfagas de habla de tramas de 5 ms **b) Duración de las pausas en tramas de 5 ms**

FIGURA 1/P.59

Distribución acumulativa de las duraciones de ráfaga de conversación y de pausa (sin tiempo de retención)

2.2 Características medidas con tiempo de retención

En el Cuadro 1 se indican los valores de algunos parámetros del habla conversacional humana. Estos valores han sido obtenidos promediando los valores contenidos en [1]-[4].

CUADRO 1/P.59

Parámetros temporales del habla conversacional (promedios de inglés, italiano y japonés)

Parámetro	Duración (s)	Proporción (%)
Ráfaga de conversación	1,004	38,53
Pausa	1,587	61,47
Diloquia	0,228	6,59
Silencio mutuo	0,508	22,48

La función de distribución acumulativa de la duración de ráfagas de conversación puede ser aproximada mediante una función exponencial, mientras que la de duración de pausas puede ser aproximada mediante una función constante más un término exponencial. Así, para las ráfagas de conversación:

$$Pts(t) = 1 - \exp(-Ats \cdot t)$$

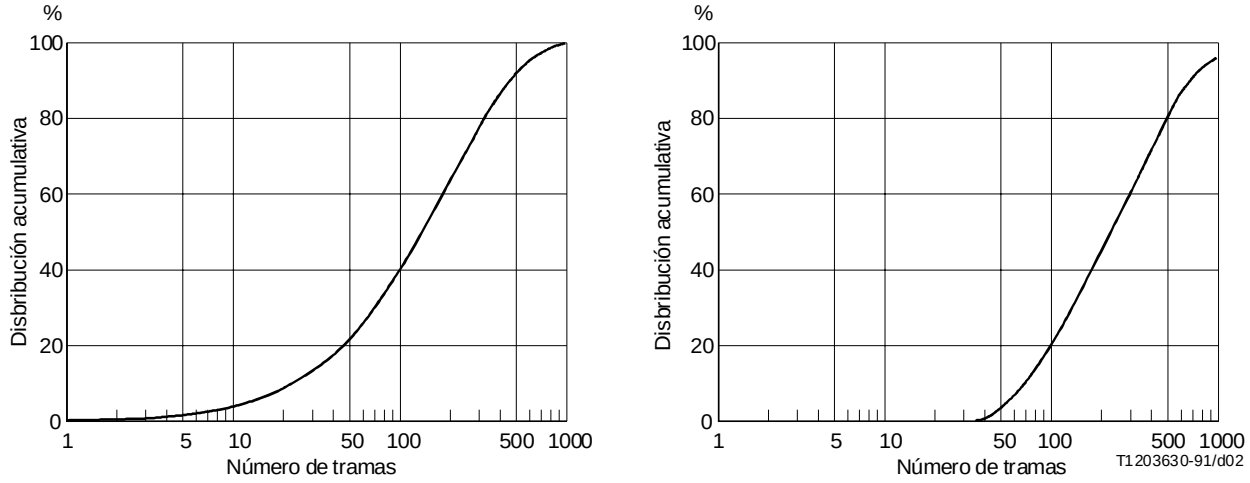
Ats = 1/ ,Tts, ,Tts: duración promediada de ráfaga de habla,

y para las pausas:

$$Pps(t) = \begin{cases} 0 & \text{para } 0 \le t \le 0,2 \\ 1 - \exp[-Aps(t - 0,2)] & \text{para } t > 0,2 \end{cases}$$

Aps = 1/(Tps – 0,2) Tps: duración promediada de pausa.

Ambas características están representadas en la Figura 2.



a) Duración de las ráfagas de habla de tramas de 5 ms **b) Duración de las pausas en tramas de 5 ms**

FIGURA 2/P.59

Distribución acumulativa de las duraciones de ráfaga de conversación y de pausa (con tiempo de retención)

3 Método de generación de habla conversacional artificial

Las ráfagas de conversación y las pausas son generadas con arreglo al modelo de transición de estados representado en la Figura 3, en el que P_1, P_2 y P_3 denotan probabilidades de transición expresadas en porcentajes. La voz artificial descrita en la Recomendación P.50 es generada durante una ráfaga de conversación.

Los valores Tst (duración de monoloquia), Tdt (duración de diloquia), y Tms (duración de silencio mutuo) varían según las ecuaciones siguientes. En estas ecuaciones, el tiempo está expresado en segundos.

$$Tst = -0,854 \ln(1 - x_1)$$
$$Tdt = -0,226 \ln(1 - x_2)$$
$$Tms = -0,456 \ln(1 - x_3)$$

$0 < x_1, x_2, x_3 < 1$: Variables aleatorias de distribución uniforme.

Cuando la duración de pausa es inferior a 200 ms, el modelo elige o bien el estado de monoloquia o el de silencio mutuo, con probabilidades de 50%, hasta que la duración de pausa excede de 200 ms. Los valores de P_1, P_2 y P_3 son 40, 50 y 50, respectivamente. La duración total del habla conversacional artificial deberá ser de al menos 10 minutos para poder ajustarse a las características especificadas en 2.2.

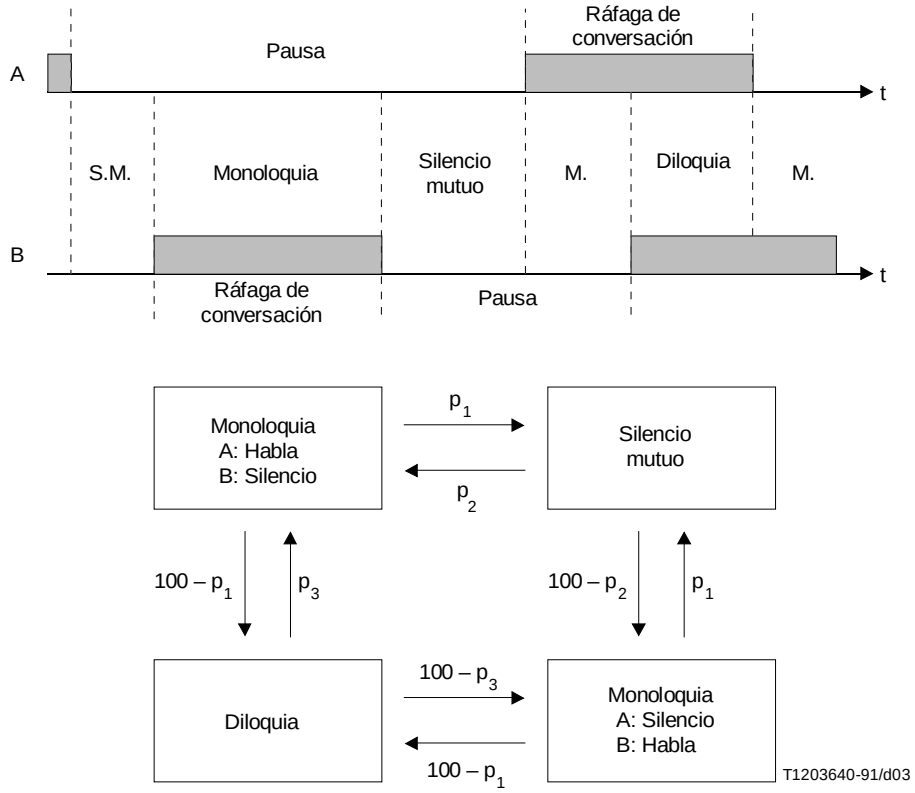


FIGURA 3/P.59

Modelo de transición de estados de la conversación

Referencias

[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, pp. 630-637, junio 1986.

[2] BRADY (P.T.): A statistical analysis of on-off patterns in 16 conversations, *BSTJ*, pp. 73-91, enero 1968.

[3] CCITT Contribución COM XII-20, *Características de activación-pausa de la voz conversacional* (CSELT), periodo de estudios 1989-1992.

[4] CCITT Contribución Tardía D.42 (WP XII/1), *Collecting procedure for on-off characteristics of conversational speech in telecommunication* (NTT), periodo de estudios 1989-1992.

[5] CCITT Contribución Tardía COM-64 (WP XII/1), *Generation of artificial voice with pauses* (NTT), periodo de estudios 1989-1992.