



UNIÓN INTERNACIONAL DE TELECOMUNICACIONES

UIT-T

SECTOR DE NORMALIZACIÓN
DE LAS TELECOMUNICACIONES
DE LA UIT

P.50

(09/99)

SERIE P: CALIDAD DE TRANSMISIÓN TELEFÓNICA,
INSTALACIONES TELEFÓNICAS Y REDES LOCALES

Aparatos para mediciones objetivas

Voces artificiales

Recomendación UIT-T P.50

(Anteriormente Recomendación del CCITT)

RECOMENDACIONES UIT-T DE LA SERIE P

CALIDAD DE TRANSMISIÓN TELEFÓNICA, INSTALACIONES TELEFÓNICAS Y REDES LOCALES

Vocabulario y efectos de los parámetros de transmisión sobre la opinión de los clientes	Serie	P.10
Líneas y aparatos de abonado	Serie	P.30
		P.300
Patrones de transmisión	Serie	P.40
Aparatos para mediciones objetivas	Serie	P.50
		P.500
Medidas electroacústicas objetivas	Serie	P.60
Medidas relativas a la sonoridad vocal	Serie	P.70
Métodos de evaluación objetiva y subjetiva de la calidad	Serie	P.80
		P.800
Calidad audiovisual en servicios multimedios	Serie	P.900

Para más información, véase la Lista de Recomendaciones del UIT-T.

VOCES ARTIFICIALES

Resumen

La "voz artificial" descrita en esta Recomendación reproduce las características de la voz humana, para la caracterización de los sistemas y dispositivos de telecomunicaciones, lineales y no lineales, diseñados para la transducción o la transmisión de señales vocales. La voz artificial es una señal matemáticamente definida que reproduce las características temporales y espectrales de la voz que afectan significativamente al funcionamiento de los sistemas de telecomunicación. Se definen dos categorías de voz artificial, que reproducen respectivamente las características de la voz femenina y masculina.

Orígenes

La Recomendación UIT-T P.50, ha sido revisada por la Comisión de Estudio 12 (1997-2000) del UIT-T y fue aprobada por el procedimiento de la Resolución N.º 1 de la CMNT el 30 de septiembre de 1999.

PREFACIO

La UIT (Unión Internacional de Telecomunicaciones) es el organismo especializado de las Naciones Unidas en el campo de las telecomunicaciones. El UIT-T (Sector de Normalización de las Telecomunicaciones de la UIT) es un órgano permanente de la UIT. Este órgano estudia los aspectos técnicos, de explotación y tarifarios y publica Recomendaciones sobre los mismos, con miras a la normalización de las telecomunicaciones en el plano mundial.

La Conferencia Mundial de Normalización de las Telecomunicaciones (CMNT), que se celebra cada cuatro años, establece los temas que han de estudiar las Comisiones de Estudio del UIT-T, que a su vez producen Recomendaciones sobre dichos temas.

La aprobación de Recomendaciones por los Miembros del UIT-T es el objeto del procedimiento establecido en la Resolución N.º 1 de la CMNT.

En ciertos sectores de la tecnología de la información que corresponden a la esfera de competencia del UIT-T, se preparan las normas necesarias en colaboración con la ISO y la CEI.

NOTA

En esta Recomendación, la expresión *empresa de explotación reconocida (EER)* designa a toda persona, compañía, empresa u organización gubernamental que explote un servicio de correspondencia pública. Los términos *Administración*, *EER* y *correspondencia pública* están definidos en la *Constitución de la UIT* (Ginebra, 1992).

PROPIEDAD INTELECTUAL

La UIT señala a la atención la posibilidad de que la utilización o aplicación de la presente Recomendación suponga el empleo de un derecho de propiedad intelectual reivindicado. La UIT no adopta ninguna posición en cuanto a la demostración, validez o aplicabilidad de los derechos de propiedad intelectual reivindicados, ya sea por los miembros de la UIT o por terceros ajenos al proceso de elaboración de Recomendaciones.

En la fecha de aprobación de la presente Recomendación, la UIT no ha recibido notificación de propiedad intelectual, protegida por patente, que puede ser necesaria para aplicar esta Recomendación. Sin embargo, debe señalarse a los usuarios que puede que esta información no se encuentre totalmente actualizada al respecto, por lo que se les insta encarecidamente a consultar la base de datos sobre patentes de la TSB.

© UIT 2000

Es propiedad. Ninguna parte de esta publicación puede reproducirse o utilizarse, de ninguna forma o por ningún medio, sea éste electrónico o mecánico, de fotocopia o de microfilm, sin previa autorización escrita por parte de la UIT.

ÍNDICE

	Página
1 Introducción	1
2 Alcance, objeto y definición	1
2.1 Alcance y objeto.....	1
2.2 Definición	2
3 Terminología.....	2
3.1 voz artificial eléctrica.....	2
3.2 señal de excitación de la boca artificial	2
3.3 voz artificial acústica	2
4 Características	3
4.1 Espectro medio a largo plazo	3
4.2 Espectro a corto plazo	3
4.3 Distribución de amplitudes instantáneas.....	4
4.4 Distribución del nivel de potencia por segmentos	4
4.5 Espectro de la envolvente de modulación.....	5
4.6 Convergencia temporal	6
5 Método de generación.....	6
5.1 Señal de fuente de excitación.....	7
5.2 Excitación glotal	7
5.3 Sonidos no vocalizados.....	7
5.4 Envolvente de potencia	8
5.5 Filtro de conformación del espectro	9
6 Bibliografía	12
Anexo A – Características del espectro a corto plazo de la voz artificial.....	12

Recomendación P.50

VOCES ARTIFICIALES

(Melbourne, 1988; modificada en Helsinki, 1993; Ginebra, 1999)

1 Introducción

La señal descrita a continuación reproduce las características de la voz humana, para la caracterización de los sistemas y dispositivos de telecomunicaciones, lineales y no lineales, diseñados para la transducción o la transmisión de señales vocales. Se sabe que para ciertos fines, tales como las mediciones objetivas de índices de sonoridad, pueden utilizarse igualmente señales más sencillas. Ejemplos de tales señales son el ruido rosa o el ruido gaussiano de espectro conformado; para los fines de esta Recomendación tales señales no podrán denominarse "voces artificiales".

La voz artificial es una señal matemáticamente definida que reproduce las características temporales y espectrales de la voz que afectan significativamente al funcionamiento de los sistemas de telecomunicación. Se definen dos categorías de voz artificial, que reproducen respectivamente las características de la voz masculina y femenina.

La voz artificial reproduce las siguientes características temporales y espectrales de la voz humana:

- a) espectro medio a largo plazo;
- b) espectro a corto plazo;
- c) distribución de amplitudes instantáneas;
- d) estructura vocalizada y no vocalizada de la forma de onda vocal;
- e) envolvente silábica.

El apéndice I/P.50 incluye un CD-ROM que contiene señales de prueba de utilidad. Las señales de este CD-ROM incluyen la señal descrita en la Recomendación P.50 y otras señales que han sido consideradas de utilidad por algunas Administraciones. Además, la base de datos vocales completa que se utilizó para elaborar la Recomendación P.50 figura también en este CD-ROM. El apéndice I/P.50 se publica aparte.

2 Alcance, objeto y definición

2.1 Alcance y objeto

La voz artificial tiene por objeto reproducir las características de la voz humana en la banda de 100 Hz a 8 kHz. Se puede utilizar para caracterizar muchos dispositivos, por ejemplo, micrófonos de carbón, aparatos telefónicos con altavoz, codificadores no lineales, dispositivos de control del eco, compansores silábicos y sistemas no lineales, en general.

La voz artificial descrita en esta Recomendación se utiliza principalmente para la evaluación objetiva de los sistemas y dispositivos de procesamiento de la palabra en los cuales una señal monocanal con actividad continua (es decir, sin pausas) basta para medir las características. Un ejemplo es la evaluación de los codecs de señales vocales. Para las evaluaciones objetivas que requieren dos señales con pausas (por ejemplo, evaluación de dispositivos con detectores de habla), debe utilizarse la señal de habla conversacional artificial descrita en la Recomendación P.59.

La utilización de voz artificial en vez de la humana tiene las ventajas tanto de una generación más sencilla como de una menor variedad de muestras que la voz humana.

Por supuesto, cuando se prueba un sistema dado, deben considerarse las características del trayecto de transmisión que le precede. Hay que producir entonces la señal real de prueba como una convolución entre la voz artificial y la respuesta en el trayecto.

2.2 Definición

La **voz artificial** es una señal matemáticamente definida, que reproduce todas las características de la voz humana apropiadas para la caracterización de los sistemas de telecomunicación lineales y no lineales. Su objeto es proporcionar una correlación satisfactoria entre las medidas objetivas y las pruebas con voz humana.

3 Terminología

Se puede producir la voz artificial en forma de señal eléctrica o acústica, según el sistema o dispositivo probado (canales de comunicación, codificadores, micrófonos). Las siguientes definiciones se refieren a las señales de la leyenda de la figura 1.

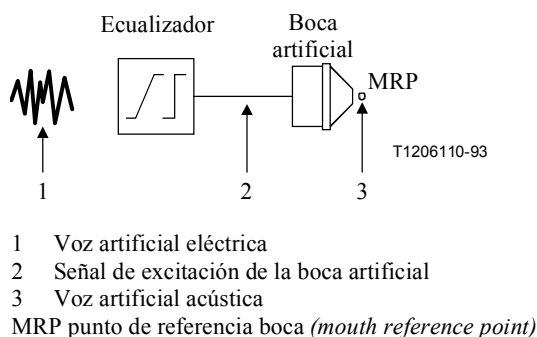


Figura 1/P.50

3.1 voz artificial eléctrica

Voz artificial producida como una señal eléctrica para probar canales de transmisión u otros dispositivos eléctricos.

3.2 señal de excitación de la boca artificial

Señal aplicada a la boca artificial para producir la voz artificial acústica. Se obtiene ecualizando la voz artificial eléctrica para compensar la característica sensibilidad/frecuencia de la boca.

NOTA – La ecualización depende de la boca artificial que se emplee y se puede efectuar eléctrica o matemáticamente dentro del proceso de generación de señales.

3.3 voz artificial acústica

Señal acústica en el punto de referencia boca (MRP), de la boca artificial. Se ajusta a las características temporales y espectrales de la voz artificial eléctrica.

4 Características

4.1 Espectro medio a largo plazo

En la figura 2 y en el cuadro 1 se indica el espectro medio a largo plazo de la voz artificial filtrado a un tercio de octava, normalizado para un nivel de presión sonora en banda ancha de $-4,7$ dBPa.

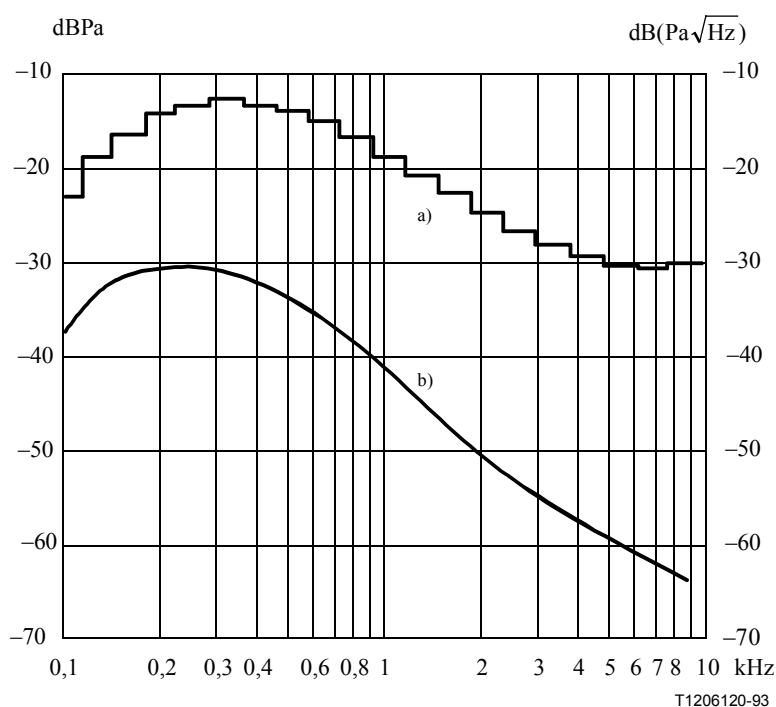
Los valores del espectro a largo plazo de la voz artificial en el MRP pueden deducirse de la ecuación:

$$S(f) = -376,44 + 465,439(\log_{10}f) - 157,745(\log_{10}f)^2 + 16,7124(\log_{10}f)^3 \quad (4-1)$$

donde $S(f)$ es la densidad espectral en dB, con respecto a una intensidad sonora de 1 pW/m^2 por hertzios a la frecuencia f . La gama de frecuencias de definición va de 100 Hz a 8 kHz .

La figura 2 muestra la curva del espectro. Los valores de $S(f)$ para frecuencias ISO a un tercio de octava vienen dados en la cuarta columna del cuadro 1. Las tolerancias vienen dadas en la quinta columna del cuadro 1. Las tolerancias por debajo de 200 Hz se aplicarán únicamente a la voz artificial masculina.

El nivel total de la presión sonora del espectro definida en la ecuación (4-1) es de $-4,7 \text{ dBPa}$. No obstante, este espectro es aplicable también para los niveles comprendidos entre $-19,7$ y $+10,3 \text{ dBPa}$. En otras palabras, el primer término de la ecuación (4-1) puede estar comprendido entre $-391,44$ y $-361,44$.



a) Espectro a un tercio de octava [columna (3) del cuadro 1].

b) Densidad espectral [columna (3)-(2) del cuadro 1].

Figura 2/P.50 – Espectro a largo plazo de la voz artificial

4.2 Espectro a corto plazo

Las características espectrales de las voces artificiales masculina y femenina se describen en el anexo A.

4.3 Distribución de amplitudes instantáneas

La distribución de la densidad de probabilidad de la amplitud instantánea de la voz artificial aparece en la figura 3.

4.4 Distribución del nivel de potencia por segmentos

La figura 4 muestra la distribución del nivel de potencia por segmentos de la voz artificial, medida en ventanas de tiempo de 16 ms. También se indican los límites superiores e inferiores de tolerancia.

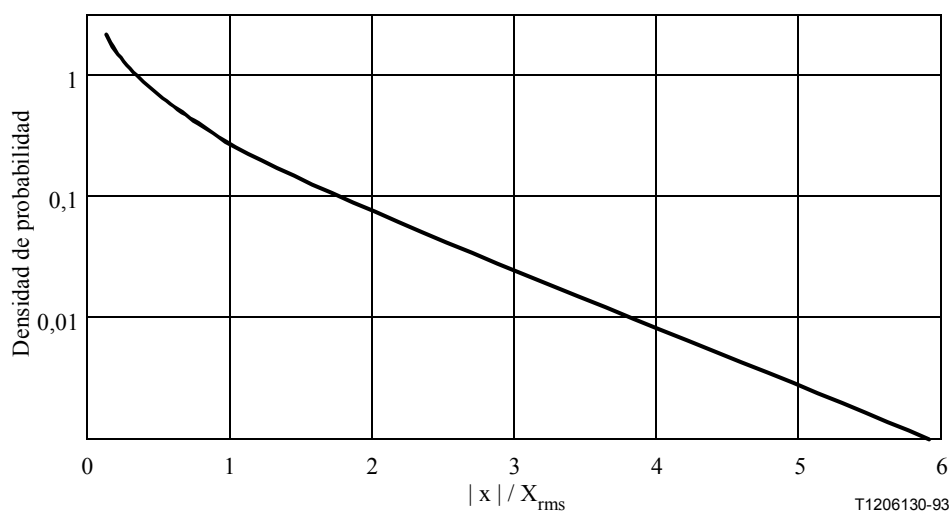
NOTA – El límite superior de tolerancia representa la distribución típica del nivel de potencia por segmentos de la conversión normal, mientras que el límite inferior representa la palabra continua (frases telefónicas).

Cuadro 1/P.50 – Espectro a largo plazo de la voz artificial

Frecuencia central del tercio de octava	Factor de corrección de anchura de banda $10 \log_{10} \Delta f$	Nivel de presión sonora (tercio de octava)	Densidad espectral	Tolerancia
(Hz) (1)	(dB) (2)	(dBPa) (3)	(dB) (3) – (2)	(dB)
100	13,6	–23,1	–36,7	–
125	14,6	–19,2	–33,8	+3, –6 ^{a)}
160	15,6	–16,4	–32,7	+3, –6 ^{a)}
200	16,6	–14,4	–31,7	+3, –6
250	17,6	–13,4	–31,7	±3,0
315	18,6	–13,0	–31,6	±3,0
400	19,6	–13,3	–32,9	±3,0
500	20,6	–14,1	–34,7	±3,0
630	21,6	–15,4	–37,7	±3,0
800	22,6	–17,0	–39,6	±3,0
1000	23,6	–18,9	–42,5	±3,0
1250	24,6	–21,0	–45,6	±3,0
1600	25,6	–23,0	–48,6	±3,0
2000	26,6	–25,1	–51,7	±3,0
2500	27,6	–26,9	–54,5	±3,0
3150	28,6	–28,6	–57,2	±3,0
4000	29,6	–29,8	–59,4	±6,0
5000	30,6	–30,6	–61,2	±6,0
6300	31,6	–30,9	–62,5	±6,0
8000	32,6	–30,5	–63,1	–
a) Las tolerancias dadas se aplican al espectro a largo plazo de la voz masculina y deben satisfacerse igualmente por los ruidos conformados por señales vocales. Sin embargo no se aplican al espectro vocal femenino, cuyo contenido energético en esta gama de frecuencias es virtualmente despreciable.				

4.5 Espectro de la envolvente de modulación

En la figura 5 se muestra el espectro de la forma de onda de la envolvente de modulación que debería reproducirse con una tolerancia de ± 5 dB en toda la gama de frecuencias.



$|x|$ Valor absoluto de la amplitud instantánea
 X_{rms} Valor cuadrático medio de la señal (valor eficaz)

Figura 3/P.50 – Distribución de amplitudes instantáneas

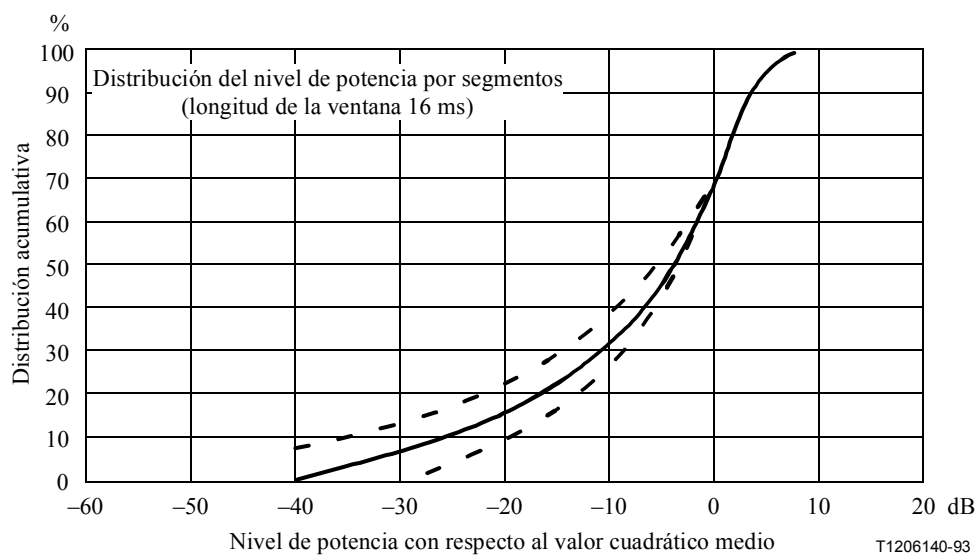


Figura 4/P.50 – Distribución del nivel de potencia por segmentos

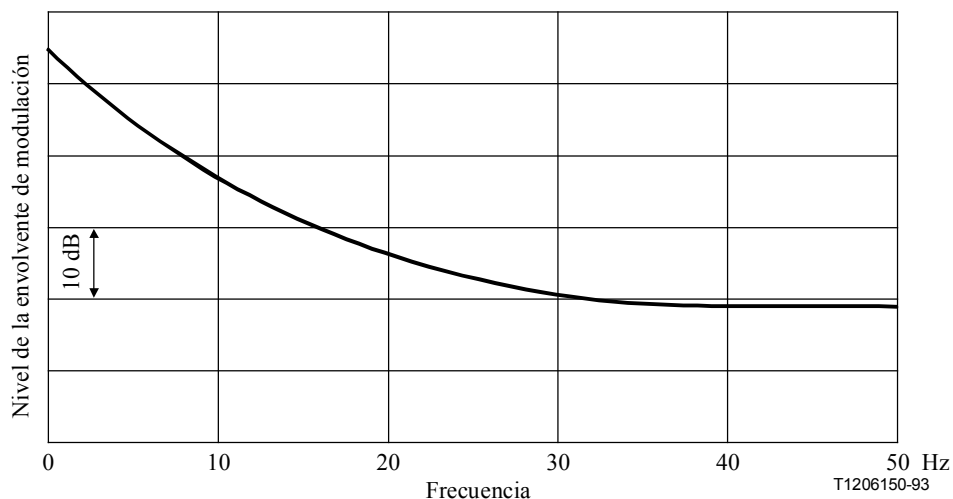


Figura 5/P.50 – Espectro de la envolvente de modulación

4.6 Convergencia temporal

La voz artificial debe tener unas características lo más semejantes posible a la voz humana. En especial, debería ser posible obtener las características de espectro a largo plazo y de distribución de amplitudes en 10 segundos.

5 Método de generación

La figura 6 muestra un diagrama de bloques del proceso de generación de la voz artificial. Se genera aplicando dos tipos diferentes de señales de fuente de excitación, una señal de excitación glotal y un ruido aleatorio, a un filtro de conformación del espectro de respuesta variable con el tiempo. La voz artificial generada por la señal de excitación glotal y por el ruido aleatorio corresponde respectivamente a los sonidos vocalizados y no vocalizados. La respuesta en frecuencia del filtro de conformación del espectro simula las características de transmisión del canal vocal.

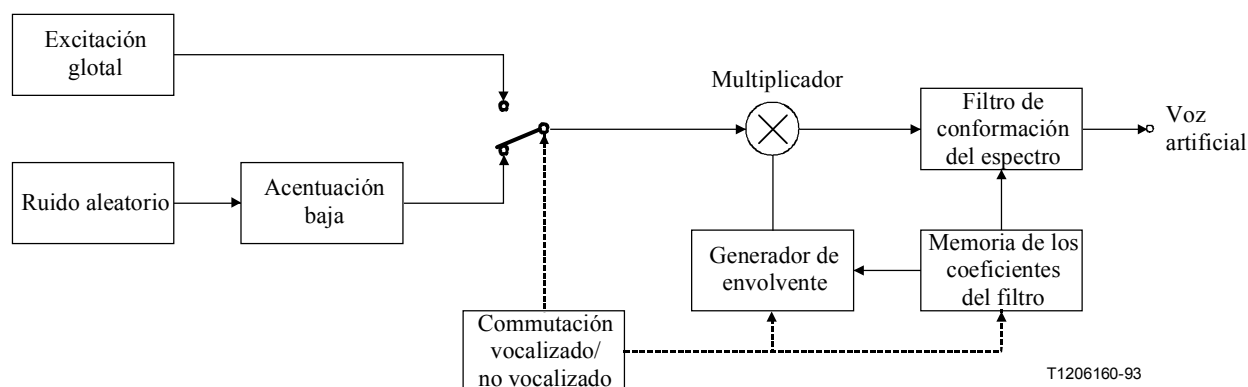


Figura 6/P.50 – Proceso de generación de la voz artificial

5.1 Señal de fuente de excitación

La voz artificial se obtiene alternando aleatoriamente los cuatro elementos unitarios básicos, cada uno de los cuales contiene segmentos vocalizados y no vocalizados. Mientras que un elemento unitario comienza con un sonido no vocalizado, seguido de uno vocalizado, los otros tres elementos comienzan con un sonido vocalizado, seguido de uno no vocalizado y terminan de nuevo con un sonido vocalizado (véase la figura 9). La relación entre la duración de los sonidos no vocalizados T_{uv} y la duración total de los segmentos vocalizados T_v para cada elemento unitario es de 0,25. La duración $T = T_{uv} + T_v$ de los elementos unitarios varía de acuerdo con la siguiente ecuación:

$$T = -3,486 \ln(r)$$

donde r designa un número aleatorio uniformemente distribuido tal que $(0,371 \leq r \leq 0,609)$.

Las duraciones de los sonidos vocalizados y no vocalizados de los cuatro elementos unitarios son las siguientes:

Elemento a: no vocalizado (T_{uv}) + vocalizado (T_v)

Elemento b: vocalizado ($T_v/4$) + no vocalizado (T_{uv}) + vocalizado ($3T_v/4$)

Elemento c: vocalizado ($T_v/2$) + no vocalizado (T_{uv}) + vocalizado ($T_v/2$)

Elemento d: vocalizado ($3T_v/4$) + no vocalizado (T_{uv}) + vocalizado ($T_v/4$).

Los elementos unitarios deben repetirse aleatoriamente durante al menos 10 segundos para satisfacer las características de la voz artificial según se especifica en la cláusula 4.

5.2 Excitación glotal

La señal de excitación glotal presenta una forma de onda periódica según se muestra en la figura 7. La frecuencia del tono ($1/T_0$ en la figura 7) varía de acuerdo con el esquema de variación mostrado en la figura 8 durante el periodo T_v . El valor inicial de la frecuencia del tono (F_s de la figura 8) se determina de acuerdo con las siguientes relaciones:

$F_s = F_c - 31,82 T_v + 39,4 R$ para la voz artificial masculina

$F_s = F_c - 51,85 T_v + 64,2 R$ para la voz artificial femenina.

donde F_c y R designan respectivamente la frecuencia central y una variable aleatoria uniformemente distribuida ($0 < R < 1$). F_c es 128 Hz para la voz artificial masculina y 215 Hz para la voz artificial femenina. En el trapecioide del esquema de variación de la frecuencia del tono, la superficie del trapecioide por encima de F_c debe ser igual a la correspondiente por debajo de F_c (sombreada en la figura 8). En los elementos de las partes b), c) y d) de la figura 9 el esquema de variación de la frecuencia del tono se aplica a la combinación de las dos partes vocalizadas, independientemente de donde se inserte el segmento no vocalizado.

5.3 Sonidos no vocalizados

La función de transferencia del filtro paso bajo situado después del generador de ruido aleatorio (baja acentuación) es $1/(1 - (1 - \epsilon) z^{-1})$, $\epsilon = 0,001$ donde z^{-1} designa el retardo unitario. El generador de ruido aleatorio genera ruido blanco gaussiano mediante un proceso aleatorio discreto de media cero cuyas muestras no están correlacionadas.

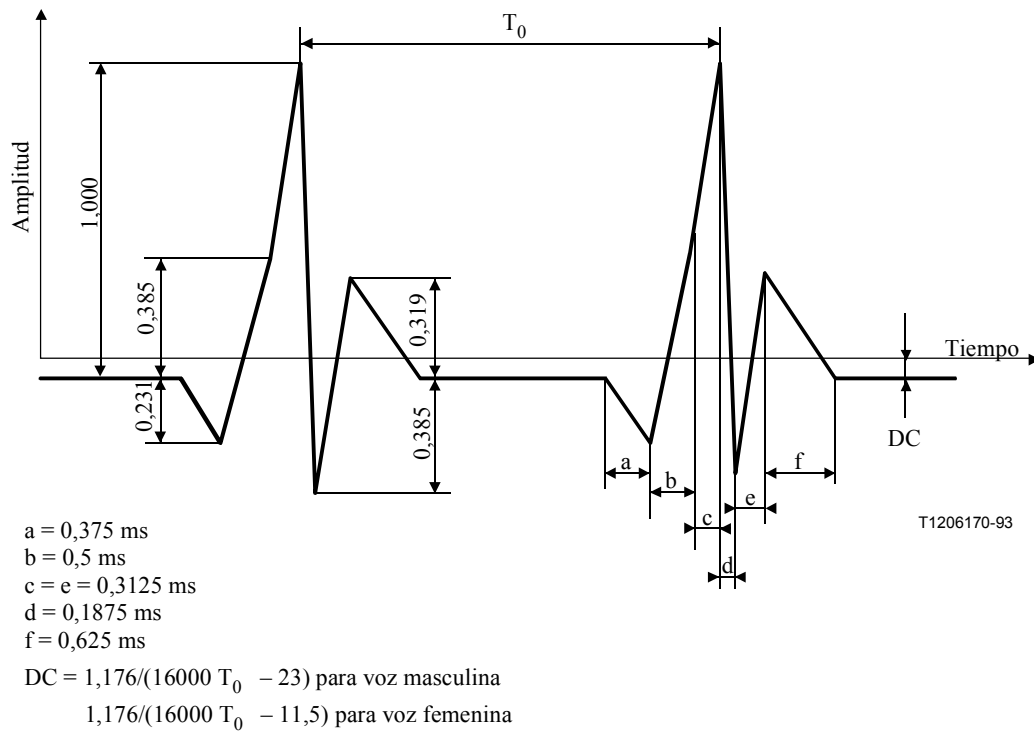


Figura 7/P.50 – Señal de excitación glotal

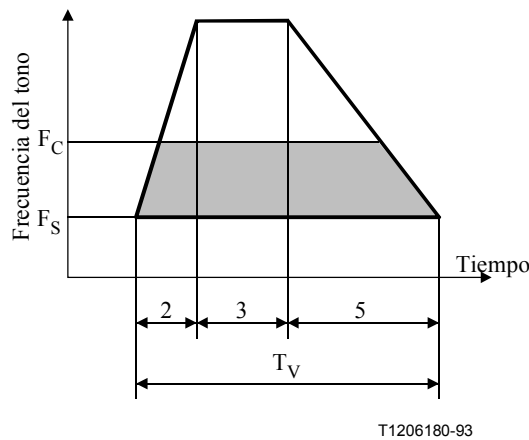


Figura 8/P.50 – Esquema de variación de la frecuencia del tono

5.4 Envoltente de potencia

La envoltente de potencia de cada elemento unitario de la señal de fuente de excitación está tan controlada que la potencia por segmentos a corto plazo (evaluada en intervalos de 2 ms) de la voz artificial varía de acuerdo con los esquemas mostrados en la figura 9. Esto se obtiene utilizando las siguientes relaciones que proporcionan las señales de entrada y salida del filtro de conformación del espectro:

$$P_{en} = P_{sal} \prod_{i=1}^{12} (1 - k_i^2)$$

donde

P_{en} es la potencia de entrada al filtro de conformación del espectro,

P_{sal} es la potencia de salida del filtro de conformación del espectro,

K_i es el i ésimo coeficiente del filtro de conformación del espectro.

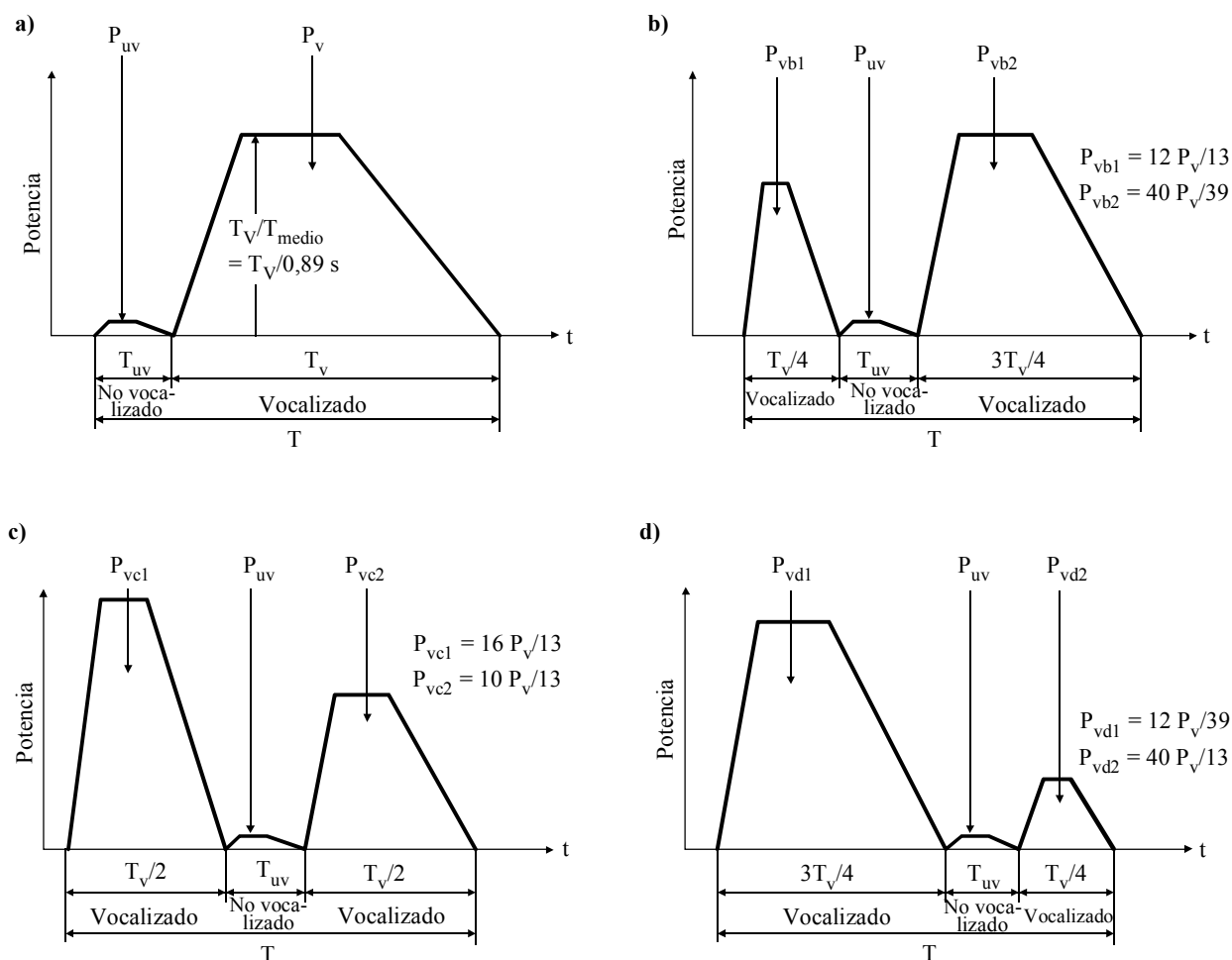
Los tiempos de subida, estacionario, y de caída de cada uno de los trapezoides de la figura 9 estarán mutuamente relacionados a través de los mismos coeficientes de proporcionalidad (2:3:5) del esquema de variación de la frecuencia del tono mostrado en la figura 8. Para cada elemento unitario, la potencia media de los sonidos no vocalizados (P_{uv}) será 17,5 dB inferior a la potencia media de los sonidos vocalizados (P_v).

5.5 Filtro de conformación del espectro

El filtro de conformación del espectro tiene una estructura en celosía de orden 12 como se indica en la figura 10. Se definen 16 grupos, cada uno con 12 coeficientes de filtrado ($k_1 - k_{12}$); se utilizarán 13 grupos para generar la parte vocalizada mientras que se utilizarán tres grupos para generar la parte no vocalizada. En el cuadro 2 se numeran estos coeficientes para las voces artificiales tanto masculina como femenina.

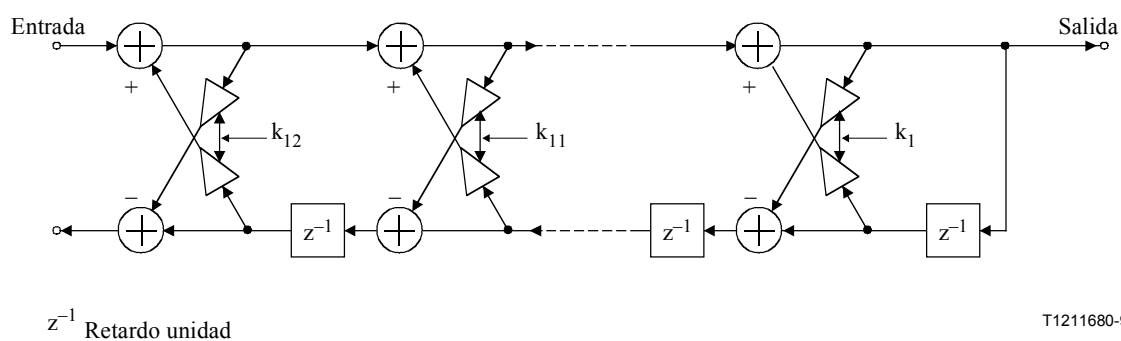
Los doce coeficientes del filtro se actualizarán cada 60 ms mientras se genera la señal. Más concretamente, durante cada periodo de 60 ms, deben actualizarse los coeficientes reales de filtrado cada 2 ms mediante una interpolación lineal entre los dos conjuntos de valores adoptados para los intervalos siguientes de 60 ms. En la parte de sonido vocalizado, se escogerá cada uno de los 13 grupos de coeficientes de forma aleatoria cada 780 ms ($= 60 \text{ ms} \times 13$) y en la parte de sonidos no vocalizados se escogerá cada uno de los tres grupos de coeficientes de forma aleatoria una vez cada 180 ms ($= 60 \text{ ms} \times 3$).

NOTA – La implementación descrita del filtro de conformación debe considerarse como un ejemplo y no es parte integrante de esta Recomendación. Cualquier otra implementación que proporcione la misma función de transferencia puede utilizarse como una alternativa. Una frecuencia de muestreo de 16 000 Hz corresponde al cuadro 2a y al cuadro 2b.



T1206190-93

Figura 9/P.50 – Esquemas de variación de la potencia a corto plazo de los cuatro elementos unitarios utilizados para generar la voz artificial



T1211680-99

Figura 10/P.50 – Filtro de conformación del espectro

NOTA – El puerto de salida del filtro de la parte inferior izquierda de la figura 10 no se utiliza; se muestra para resaltar la simetría de la estructura del filtro.

Cuadro 2a/P.50 – Coeficientes k_i para voz artificial masculina

		k_1	k_2	k_3	k_4	k_5	k_6	k_7	k_8	k_9	k_{10}	k_{11}	k_{12}
No vocalizado	1	-0,471	-0,108	0,024	-0,048	0,140	0,036	0,054	0,004	0,123	0,044	0,099	-0,003
	2	-0,284	-0,468	0,030	0,090	0,124	-0,020	0,087	0,067	0,131	0,011	0,076	-0,024
	3	-0,025	-0,496	-0,176	0,162	0,236	-0,012	0,068	0,001	0,096	0,029	0,086	-0,018
Vocalizado	1	0,974	0,219	0,025	-0,123	-0,132	-0,203	-0,103	-0,174	-0,079	-0,153	-0,010	-0,061
	2	0,629	-0,152	-0,138	-0,142	-0,118	-0,135	0,147	0,019	0,077	-0,040	0,029	-0,007
	3	0,599	-0,119	0,067	0,051	0,103	0,023	0,106	0,036	-0,006	-0,133	-0,052	-0,094
	4	0,164	-0,364	-0,248	-0,076	0,168	0,072	0,103	0,045	0,112	0,010	0,048	-0,034
	5	0,842	0,022	0,171	0,173	0,067	-0,057	0,089	-0,045	-0,039	-0,134	-0,034	-0,122
	6	0,933	-0,537	-0,137	-0,161	-0,216	-0,139	0,115	-0,042	0,027	-0,163	0,102	-0,107
	7	0,937	-0,413	0,132	-0,059	-0,103	-0,134	0,047	-0,115	-0,105	-0,097	0,039	-0,108
	8	0,965	-0,034	0,032	0,001	-0,107	-0,189	-0,057	-0,175	-0,109	-0,163	-0,003	-0,055
	9	0,870	-0,476	-0,016	-0,136	-0,125	-0,107	0,091	-0,008	0,021	-0,128	0,042	-0,069
	10	0,686	-0,030	0,178	0,197	0,155	-0,026	0,078	0,004	-0,001	-0,128	-0,004	-0,102
	11	0,963	-0,232	0,086	-0,018	-0,147	-0,192	-0,040	-0,179	-0,144	-0,133	0,042	-0,042
	12	0,930	-0,461	0,071	-0,144	-0,122	-0,096	0,034	-0,066	-0,021	-0,171	0,067	-0,091
	13	0,949	-0,334	0,143	-0,040	-0,112	-0,161	0,010	-0,156	-0,123	-0,119	0,049	-0,070

Cuadro 2b/P.50 – Coeficientes k_i para voz artificial femenina

		k_1	k_2	k_3	k_4	k_5	k_6	k_7	k_8	k_9	k_{10}	k_{11}	k_{12}
No vocalizado	1	0,488	-0,388	0,145	0,053	0,122	0,027	0,135	0,035	0,080	0,017	0,068	0,028
	2	-0,093	-0,444	-0,102	0,121	0,154	0,009	0,102	-0,031	0,084	0,019	0,101	-0,020
	3	-0,709	-0,179	0,134	0,007	0,142	0,027	0,099	0,000	0,115	0,007	0,075	-0,037
Vocalizado	1	0,355	-0,247	-0,092	-0,043	0,032	0,046	0,113	-0,023	0,071	-0,030	-0,000	-0,116
	2	0,976	0,150	-0,062	-0,187	-0,172	-0,200	-0,122	-0,207	-0,054	-0,127	0,012	-0,111
	3	0,737	-0,324	-0,175	-0,197	-0,153	0,023	0,110	-0,018	0,040	-0,062	0,034	-0,091
	4	0,598	0,234	0,126	0,011	-0,005	-0,026	0,131	0,032	0,073	-0,063	0,011	-0,088
	5	0,808	0,118	0,262	0,139	0,063	-0,024	0,001	-0,184	-0,056	-0,100	0,014	-0,115
	6	0,914	-0,500	-0,051	-0,115	-0,211	-0,012	-0,077	-0,179	0,064	-0,102	0,037	-0,092
	7	0,933	-0,359	0,089	-0,107	-0,178	-0,050	-0,137	-0,206	0,046	-0,088	-0,004	-0,074
	8	0,966	-0,023	0,044	-0,105	-0,178	-0,195	-0,150	-0,233	-0,045	-0,092	0,029	-0,097
	9	0,870	-0,469	-0,244	-0,107	-0,140	-0,037	0,084	-0,131	0,021	-0,066	-0,003	-0,091
	10	0,673	-0,292	0,392	0,158	0,143	0,160	0,019	-0,281	-0,105	-0,195	-0,156	-0,185
	11	0,962	-0,191	0,030	-0,089	-0,207	-0,133	-0,141	-0,263	0,007	-0,054	0,014	-0,074
	12	0,879	-0,340	0,046	-0,049	-0,071	-0,024	-0,039	-0,188	0,017	-0,078	-0,014	-0,117
	13	0,941	-0,258	0,122	-0,073	-0,163	-0,089	-0,151	-0,250	0,025	-0,062	-0,006	-0,093

6 Bibliografía

- [1] Contribución del CCITT COM XII-N.º 76 (1982), Una voz artificial (Italia).
- [2] Contribución del CCITT COM XII-N.º 108 (1982), Proyecto de texto parcial del punto 3 de la Recomendación P.51 (Relator especial).
- [3] Contribución del CCITT COM XII-N.º 11 (1981), Measurement of carbon microphone using real voice and artificial test signals (Canadá).
- [4] Contribución del CCITT COM XII-N.º 150 (1983), Mejoramiento de la voz artificial insertado sonidos fricativos (Italia).
- [5] Contribución del CCITT COM XII-N.º 132 (1983), Evaluation of non-linear distortion using simulated speech signals (Canadá).

ANEXO A

Características del espectro a corto plazo de la voz artificial

Se genera la voz artificial seleccionando aleatoriamente cada uno de los 16 esquemas del espectro a corto plazo cada 960 ms (=60 ms × 16 esquemas). La densidad espectral de cada uno de los esquemas viene dada por la ecuación (A-1) y el cuadro A.1 y el espectro a corto plazo de la señal durante el intervalo de 60 ms que va entre cualesquiera dos selecciones sucesivas de esquemas varía suavemente de un esquema al siguiente.

NOTA – Los esquemas del espectro de la ecuación (A-1) y el cuadro A.1 se expresan en forma de potencias normalizadas.

$$\text{Densidad del espectro } S_i(f) = \frac{1}{A_{i0} + 2 \sum_{j=1}^{12} A_{ij} [\cos(2\pi j f)]}, i = 1, 2, \dots, 16 \quad (\text{A-1})$$

Cuadro A.1a/P.50 – Coeficientes A_{ij} para voz artificial masculina

i	j												
	0	1	2	3	4	5	6	7	8	9	10	11	12
1	2,09230	-1,33222	1,32175	-1,14200	0,99352	-0,94634	0,72684	-0,63263	0,41196	-0,42858	0,22070	-0,19746	0,10900
2	9,34810	-8,55934	7,35732	-6,35320	5,33999	-4,47238	3,62417	-2,85246	2,12260	-1,49424	0,93988	-0,44998	0,12400
3	11,69068	-10,91138	9,46588	-8,11729	6,94160	-5,90977	4,95137	-3,89587	2,88750	-1,97671	1,14892	-0,50255	0,12100
4	12,56830	-11,81209	10,36030	-8,82879	7,37947	-6,01017	4,66740	-3,46913	2,42182	-1,60880	0,91652	-0,39648	0,12000
5	6,83438	-6,18275	5,59089	-4,71866	4,06004	-3,44767	2,65380	-2,12140	1,50334	-1,07904	0,64553	-0,31816	0,11500
6	12,37251	-11,52358	9,89962	-8,31774	6,99062	-5,86272	4,69809	-3,56806	2,53340	-1,70522	0,99232	-0,45403	0,13400
7	21,07637	-19,62125	16,56781	-13,67518	11,41379	-9,61940	7,93529	-6,32841	4,92443	-3,53539	2,09095	-0,86543	0,18100
8	30,77371	-29,17365	25,52254	-21,51978	17,80583	-14,30488	10,87190	-7,71572	5,14643	-3,20113	1,72149	-0,68054	0,14400
9	4,18618	-3,36611	3,36793	-2,92133	2,38452	-2,06047	1,57550	-1,34240	0,84994	-0,70462	0,38685	-0,21857	0,12100
10	14,12359	-13,14611	11,25804	-9,47510	7,97588	-6,70717	5,44803	-4,23843	3,10807	-2,12879	1,25096	-0,53230	0,12600
11	26,36971	-24,95984	21,80496	-18,41045	15,30642	-12,49415	9,84879	-7,40287	5,29262	-3,43906	1,84980	-0,71546	0,14800
12	11,50808	-10,74609	9,34328	-7,91953	6,66959	-5,54500	4,34328	-3,27036	2,33714	-1,61333	0,96597	-0,44666	0,13500
13	5,32020	-4,61998	4,29145	-3,62118	3,01310	-2,67071	2,13992	-1,72147	1,22163	-0,93163	0,53317	-0,28989	0,11900
14	20,61945	-19,39682	16,80034	-14,14817	11,84307	-9,78712	7,73534	-5,77921	4,06200	-2,66324	1,49831	-0,59887	0,12600
15	30,02641	-28,42244	24,75314	-20,70178	16,98199	-13,72247	10,81050	-8,20966	5,94148	-3,90501	2,11507	-0,81306	0,16400
16	27,62370	-26,17896	22,93678	-19,42253	16,18997	-13,17171	10,19859	-7,42299	5,07437	-3,21481	1,73980	-0,67818	0,14000

Cuadro A.1b/P.50 – Coeficientes A_{ij} para voz artificial femenina

i	j												
	0	1	2	3	4	5	6	7	8	9	10	11	12
1	8,92953	-8,28905	7,23150	-6,06571	5,06663	-4,16883	3,34820	-2,64174	1,91152	-1,27122	0,74358	-0,35347	0,13100
2	9,11050	-8,29868	7,05018	-6,03862	5,02156	-4,15784	3,37442	-2,70084	2,04257	-1,41928	0,90339	-0,47240	0,14100
3	13,69058	-12,75539	10,87390	-8,98976	7,28838	-5,87662	4,78110	-3,72852	2,71831	-1,81828	1,06757	-0,48907	0,12900
4	6,95118	-6,29134	5,35757	-4,64910	3,87250	-3,23132	2,69856	-2,12367	1,58326	-1,13885	0,70609	-0,37778	0,12800
5	2,73454	-1,80664	1,95283	-1,79464	1,43897	-1,31656	0,93268	-0,87398	0,53694	-0,47562	0,24159	-0,15438	0,11800
6	10,82358	-10,07808	8,78565	-7,43643	6,14765	-5,10041	4,10027	-3,23241	2,34220	-1,54676	0,91918	-0,45059	0,14200
7	21,58481	-19,92676	16,21532	-12,43566	9,61057	-7,93982	6,97363	-6,00866	4,68271	-3,12797	1,70804	-0,67499	0,14200
8	23,73912	-22,20897	18,74416	-15,03715	11,99248	-9,85513	8,27112	-6,72826	4,94335	-3,10450	1,60004	-0,61090	0,12900
9	4,97162	-4,27705	4,01380	-3,38500	2,78457	-2,45010	1,98057	-1,63020	1,18104	-0,80108	0,51528	-0,29138	0,12500
10	13,37598	-12,45509	10,72295	-8,97928	7,35893	-6,05438	4,88819	-3,86108	2,85164	-1,88876	1,11490	-0,52260	0,13800
11	16,48817	-15,22287	12,62125	-10,23900	8,46966	-7,23692	6,24648	-5,09752	3,77465	-2,46950	1,37404	-0,57453	0,13200
12	18,22041	-17,17540	15,09489	-12,55171	10,24976	-8,45903	6,71874	-5,19063	3,52021	-2,10167	1,08066	-0,41880	0,14300
13	1,32602	-0,31718	0,44277	-0,47070	0,32935	-0,41555	0,25775	-0,32079	0,13791	-0,23640	0,10103	-0,10136	0,11500
14	16,90640	-15,73723	13,30151	-10,82887	8,78690	-7,34521	6,21516	-5,11100	3,80281	-2,43990	1,33506	-0,55971	0,12900
15	21,73895	-20,42432	17,51117	14,44152	11,79131	-9,66735	7,90433	-6,19508	4,41275	-2,75545	1,46525	-0,59916	0,14000
16	21,04832	-19,72714	16,81197	-13,70183	11,07189	-9,12707	7,57941	-6,08064	4,40471	-2,74320	1,43897	-0,58079	0,13300

SERIES DE RECOMENDACIONES DEL UIT-T

Serie A	Organización del trabajo del UIT-T
Serie B	Medios de expresión: definiciones, símbolos, clasificación
Serie C	Estadísticas generales de telecomunicaciones
Serie D	Principios generales de tarificación
Serie E	Explotación general de la red, servicio telefónico, explotación del servicio y factores humanos
Serie F	Servicios de telecomunicación no telefónicos
Serie G	Sistemas y medios de transmisión, sistemas y redes digitales
Serie H	Sistemas audiovisuales y multimedios
Serie I	Red digital de servicios integrados
Serie J	Transmisiones de señales radiofónicas, de televisión y de otras señales multimedios
Serie K	Protección contra las interferencias
Serie L	Construcción, instalación y protección de los cables y otros elementos de planta exterior
Serie M	RGT y mantenimiento de redes: sistemas de transmisión, circuitos telefónicos, telegrafía, facsímil y circuitos arrendados internacionales
Serie N	Mantenimiento: circuitos internacionales para transmisiones radiofónicas y de televisión
Serie O	Especificaciones de los aparatos de medida
Serie P	Calidad de transmisión telefónica, instalaciones telefónicas y redes locales
Serie Q	Conmutación y señalización
Serie R	Transmisión telegráfica
Serie S	Equipos terminales para servicios de telegrafía
Serie T	Terminales para servicios de telemática
Serie U	Conmutación telegráfica
Serie V	Comunicación de datos por la red telefónica
Serie X	Redes de datos y comunicación entre sistemas abiertos
Serie Y	Infraestructura mundial de la información y aspectos protocolo Internet
Serie Z	Lenguajes y aspectos generales de soporte lógico para sistemas de telecomunicación